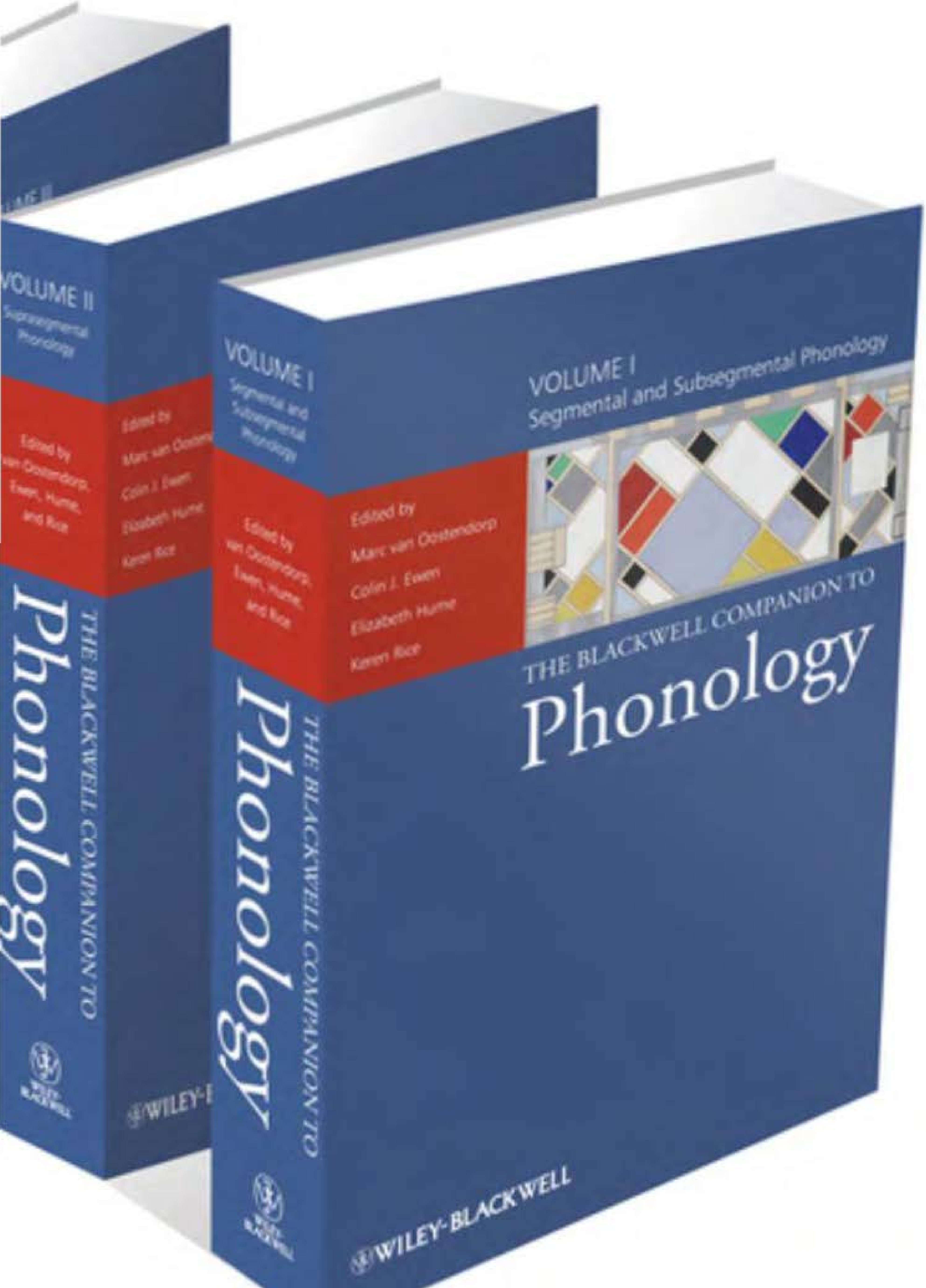

VOLUME I
Segmental and Subsegmental Phonology
Edited by
Marc van Oostendorp
Colin J. Ewen
Elizabeth Hume
Keren Rice
THE BLACKWELL COMPANION TO
Phonology
WILEY-BLACKWELL
VOLUME II
Suprasegmental Phonology

This edition first published 2011

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

The Blackwell companion to phonology / edited by Marc van Oostendorp . . . [et al.].
p. cm. — (Blackwell companions to linguistics series)
Includes bibliographical references and index.
ISBN 978-1-4051-8423-6 (hardcover : alk. paper)
1. Phonetics. 2. Grammar, Comparative and general—Phonology.
I. Oostendorp, Marc van, 1967–
P217.B53 2011
414—dc22 2010042206

A catalogue record for this book is available from the British Library.

Set in 10/12pt Palatino by Graphicraft Limited, Hong Kong

1 2011

Brief Contents

Full Table of Contents

Volume I

Volume II

Suprasegmental and Prosodic Phonology

Volume III

Phonological Processes

Volume IV

Phonological Interfaces

Volume V

Phonology across Languages

Contributors

Akinbiyi Akinlabi is Professor of Linguistics at Rutgers University, New Brunswick. He is the current President of the World Congress of African Linguistics, and serves on the councils of both the West African Languages Congress and the Annual Conference on African Linguistics (USA). His expertise lies primarily in the phonologies of the Benue–Congo languages of West Africa. He has published articles in leading journals of theoretical and descriptive linguistics, including *Linguistic Inquiry*, *Journal of Linguistics*, and *Journal of African Languages and Linguistics*. His forthcoming book is titled: *Yoruba: A phonological grammar*.

Adam Albright is an Associate Professor (Anshen-Chomsky Professorship in Language, and Mind Career Development Chair) at MIT. His research interests include phonology, morphology, and learnability, with an emphasis on using computational modeling and experimental techniques to investigate issues in phonological theory.

Stephen R. Anderson is Dorothy R. Diebold Professor of Linguistics at Yale University. After receiving his Ph.D. from MIT in 1969, he taught at Harvard, UCLA, Stanford, and Johns Hopkins Universities before coming to Yale in 1994. He is the author of numerous articles and six books, including *Aspects of the theory of clitics* (2005). He has done field research on several languages, most recently the Surmiran form of Rumantsch. In addition to phonology and morphology, his research interests include animal communication systems and the evolution of human language.

Diana Archangeli received her Ph.D. from MIT in 1984. She has been a faculty member at the University of Arizona since 1985 and was a Fellow at the Center for Advanced Study in the Behavioral Sciences at Stanford in 2007–8. She co-authored *Grounded phonology* (1994) with Douglas Pulleyblank, and co-edited *Optimality Theory: An over-view* (1997) with D. Terence Langendoen. She is Director of the Arizona Phonological Imaging Lab, used for ultrasound study of the articulation of language sounds.

Amalia Arvaniti is an Associate Professor in the Department of Linguistics at the University of California, San Diego. She received her Ph.D. from the University of Cambridge and has held research and teaching appointments at the Universities of Cambridge, Oxford, Edinburgh, and Cyprus. Her research focuses on the phonetics and phonology of prosody, with special emphasis on the experimental investigation and formal representation of rhythm and intonation.

Bao Zhiming is a linguist working in the Department of English Language and Literature, National University of Singapore. He has two main research interests: Chinese phonology and contact linguistics.

Outi Bat-El is Professor of Linguistics at Tel-Aviv University. She is engaged in the study of Semitic phonology and morphology. In her 1994 *Natural Language & Linguistic Theory* article, she initiated the Semitic root debate, arguing that there is no consonantal root in Semitic morphology. Her subsequent work within the framework of Optimality Theory (*Recherches Linguistiques de Vincennes* 2003) provided further support to this argument. She has also authored articles on Hebrew truncation (*Language* 2002) and reduplication (*Linguistic Inquiry* 2006), on blends (*Phonology* 1996) and hypocoristics (*Phonology* 2005), as well as on language acquisition (*Language Sciences* 2009).

Michael Becker received his Ph.D. from the University of Massachusetts Amherst in 2009, and is currently a lecturer at Harvard University. His work focuses on the grammatical principles that govern lexicon organization, especially as a way to discover Universal biases in the phonological grammar. He also works on the acquisition and learning of lexical patterns.

Ricardo Bermúdez-Otero is Senior Lecturer in Linguistics and English Language at the University of Manchester. He previously held a Postdoctoral Fellowship at the British Academy, followed by a Lectureship in Linguistics at the University of Newcastle upon Tyne. His research focuses on the morphosyntax–phonology and phonology–phonetics interfaces, with particular attention to diachronic issues. His publications include chapters in *Optimality Theory and language change* (2003), *The handbook of English linguistics* (2006), *The Cambridge handbook of phonology* (2007), *Deponency and morphological mismatches* (2007), *Optimality-theoretic studies in Spanish phonology* (2007), and *Morphology and its interfaces* (forthcoming).

Christina Y. Bethin is Professor of Linguistics at Stony Brook University, New York, who works on prosody and syllable structure. She has written numerous articles on the diachronic and synchronic phonology of the Slavic languages, including Polish, Ukrainian, Russian, Belarusian, Serbian and Croatian, Czech, and Slovene, and published two award-winning books, *Polish syllables: The role of prosody in phonology and morphology* (1992) and *Slavic prosody: Language change and phonological theory* (1998).

Geert Booij is Professor of Linguistics at the University of Leiden. He is the author of *The phonology of Dutch* (1995), *The morphology of Dutch* (2002), *The grammar of words* (2005), and *Construction morphology* (2010), and of a number of linguistic articles in a wide range of Dutch and international journals, with focus on the interaction of phonology and morphology, and theoretical issues in morphology.

Anna R. K. Bosch has taught linguistics at the University of Kentucky since 1990, where she is now Associate Dean for Undergraduate Programs and a faculty member of the Linguistics Program and English departments. She has written on

Scottish Gaelic phonology and dialectology, and is currently working on a new project on the history of phonetic transcription in European dialect studies.

Bert Botma is associate professor at the Leiden University Centre of Linguistics (LUCL) and a research fellow at the Netherlands Organization of Scientific Research (NWO). His main research interest is segmental phonology. He has published on such topics as nasal harmony, English syllable structure, and on the phonological contrast between obstruents and sonorants.

Diane Brentari is Professor of Linguistics and Director of the ASL Program at Purdue University. She has published widely in the area of sign language phonology and morphology. She is the author of *A prosodic model of sign language phonology* (1998) and editor of *Sign languages: A Cambridge language survey* (2010). Her current research involves the cross-linguistic analysis of sign languages.

Adam Buchwald is an Assistant Professor at New York University in Communicative Sciences and Disorders. He holds a Ph.D. in Cognitive Science from Johns Hopkins University and had a post doctoral fellowship in the Speech Research Lab at Indiana University. His work is interdisciplinary and spans topics in linguistics, psycholinguistics, and communication sciences.

Eugene Buckley is Associate Professor of Linguistics at the University of Pennsylvania. His research interests include metrical and syllable structure, and the proper locus of phonetic and functional explanation. Much of his work focuses on native languages of North America.

Luigi Burzio is Professor Emeritus, Department of Cognitive Science, Johns Hopkins University. He has also taught at Harvard University, Department of Romance Languages and Literatures. His research interests include syntax, phonology and morphology. He is the author of *Italian syntax* (1986) and *Principles of English stress* (1994).

Patrik Bye is currently a researcher with the Center for Advanced Study in Theoretical Linguistics at the University of Tromsø. He has published scholarly articles on a variety of topics including the syllable structure, quantity, and stress systems of the Finno-Ugric Saami languages, North Germanic accentology, and phonologically conditioned allomorphy. With Martin Krämer and Sylvia Blaho he edited *Freedom of analysis?* (2007).

Charles Cairns is Professor Emeritus of Linguistics at the City University of New York. He received his Ph.D. in linguistics from Columbia University in 1968. Author of a number of articles in phonology, his most recent work is a co-edited volume with Eric Raimy, *Contemporary views on architecture and representations in phonology* (2009).

Andrea Calabrese was born in Campi Salentina in the southeastern tip of Italy. He obtained his Ph.D. in linguistics at MIT in 1988 and is currently teaching at the University of Connecticut. His main interests are phonology, morphology, and historical linguistics. He has published more than 50 articles in books and journals such as *Linguistic Inquiry, Linguistic Review, Studies in Language, Brain and Language, Journal of Neuro-Linguistics, and Rivista di Linguistica.* His recent book, *Markedness and economy in a derivational model of phonology* (2005), proposes a theory integrating phonological rules and repairs triggered by markedness constraints into a derivational model of phonology.

Roderic F. Casali's interests include phonetics, phonological theory, and descriptive work on the phonology of African languages. His research has focused primarily on vowel phenomena, especially vowel hiatus resolution and ATR vowel harmony. He has done linguistic fieldwork in Ghana with SIL and currently teaches linguistics at the Canada Institute of Linguistics at Trinity Western University in Langley, British Columbia.

Young-mee Yu Cho is Associate Professor of East Asian Languages and Cultures at Rutgers University and author of *Parameters of consonantal assimilation* (1999), *Integrated Korean* (2000), and *Korean phonology and morphology* (forthcoming). She has written on Korean language and culture, theoretical linguistics, and Korean pedagogy.

Abigail C. Cohn is Professor of Linguistics at Cornell University, Ithaca, New York. Her research addresses the relationship between phonology and phonetics and is informed by laboratory phonology approaches. She also specializes in the description and analysis of a number of Austronesian languages of Indonesia. Her published work includes articles in *Phonology* and a number of edited volumes and she is co-editor of the forthcoming *Oxford handbook of laboratory phonology*.

Jennifer Cole is Professor of Linguistics and Computer Science at the University of Illinois at Urbana-Champaign, and is a member in the Cognitive Science group at the Beckman Institute for Advanced Science and Technology. She is the founding General Editor of *Laboratory Phonology*, and has served on the editorial boards of the journals *Language*, *Linguistic Inquiry*, and *Phonology*.

Bruce Connell is based at York University, Toronto. His research interests include the phonetics of African languages, the relationships between phonetics and phonology, historical linguistics, and language endangerment and documentation. He makes regular fieldtrips to Africa for research and is an authority on languages in the Nigeria–Cameroon borderland. He received his Ph.D. in Linguistics from the University of Edinburgh in 1991. His publications include "The perception of lexical tone in Mambila," *Language and Speech* 43, and "Tone languages and the universality of intrinsic F0: Evidence from Africa," *Journal of Phonetics* 30.

Marie-Hélène Côté is Associate Professor of Linguistics at the University of Ottawa. Much of her research in phonology relates to the role of perceptual factors in phonological processes, the status of the syllable, and the treatment of variation. She also specializes in French phonology, with a focus on Laurentian (Québec) French.

Onno Crasborn is a Senior Researcher at the Deparhnent of Linguistics of Radboud University Nijmegen, where he heads the sign language research theme of the Centre for Language Studies. After completing a dissertation on phonetic variation in Sign Language of the Netherlands (2001) at Leiden University, he has broadened his research interests beyond sign phonetics and phonology to sociolinguistics, discourse, and corpus linguistics. In 2008 he published the Corpus NGT, the first open access sign language corpus in the world.

Megan J. Crowhurst is an Associate Professor in the Department of Linguistics at the University of Texas at Austin. Her publications in theoretical phonology have concentrated on prosodic phenomena related to word stress and reduplication.

In collaboration with Mark Hewitt, she has also contributed to the literature on constraint conjunction. Her current research, conducted in an experimental paradigm, explores the ways in which humans' perception of rhythm might contribute to the form and frequency of stress patterns found in natural languages.

Stuart Davis is Professor of Linguistics at Indiana University, Bloomington. His primary area of research is in phonology and phonological theory with a secondary area of research in the early history of linguistics in the USA. His work in phonology has especially focused on issues related to syllable structure and word-level prosody. His work has appeared in a wide variety of edited volumes and journals including *Linguistic Inquiry*, *Phonology*, *Lingua*, *Linguistics*, and *American Speech*.

Kenneth de Jong is currently Professor of Linguistics, Cognitive Science and Second Language Studies at Indiana University. He has worked extensively on questions of how prosodic organization pervades the details of speech production, and more generally how speech production and perception interact with one another and how this relates to the phonological system. He is author of more than 50 articles on aspects of phonetic behavior and is currently Associate Editor of the *Journal of Phonetics*.

Paul de Lacy is an Associate Professor in the Department of Linguistics at Rutgers University, an Associate of the Rutgers Center for Cognitive Science, Co-director of the Rutgers Phonetics and Fieldwork Laboratory, and editor of the *Cambridge handbook of phonology*. He works on phonology and its interfaces with syntax, morphology, and phonetics.

Laura J. Downing is a research fellow at the ZAS, Berlin, leading projects on the phonology–focus–syntax interface in Bantu languages. She has published several articles on tone and depressor tone, morphologically-conditioned morphology, syllable structure and prosodic morphology in Bantu languages, and a book, *Canonical forms in prosodic morphology* (2006).

B. Elan Dresher is Professor of Linguistics at the University of Toronto. He has published on phonological theory, learnability, historical linguistics, and West Germanic and Biblical Hebrew phonology and prosody. He is the author of *Old English and the theory of phonology* (1985) and "Chomsky and Halle's revolution in phonology" in the *Cambridge companion to Chomsky* (2005). His recent books include *Formal approaches to poetry: Recent developments in generative metrics* (co-ed., with Nila Friedberg, 2006), *Contrast in phonology: Theory, perception, acquisition* (co-ed., with Peter Avery and Keren Rice, 2008), and *The contrastive hierarchy in phonology* (2009).

San Duanmu is Professor of Linguistics, University of Michigan, where he has taught since 1991. He obtained his Ph.D. in linguistics from MIT in 1990 and from 1981 to 1986 held a teaching post at Fudan University in Shanghai.

Amanda Dye is a graduate student in Linguistics at New York University. She holds a B.A. in Linguistics from Harvard University. Her past work has focused on theoretical and experimental study of the morphophonology and syntax of mutation in Welsh. Her research interests are in experimental phonology, segmental phonology, the study of Welsh, and the phonology–semantics interface.

Mirjam Ernestus is an Associate Professor at Radboud University in Nijmegen. She obtained her Ph.D. from the Free University Amsterdam in 2000 and then held several postdoctoral positions at the Max Planck Institute for Psycholinguistics in Nijmegen.

Kathryn Flack Potts received her Ph.D. in linguistics from the University of Massachusetts, Amherst in 2007. She is a Lecturer in Linguistics at Stanford University.

Stefan A. Frisch is an Associate Professor in the Department of Communication Sciences and Disorders at the University of South Florida. He uses the tools of laboratory phonology to examine frequency effects in the lexicon and in metalinguistic judgments of well-formedness. He is also interested in speech articulation and the use of experimentally elicited speech errors in the study of speech production processes.

Adamantios I. Gafos is an Associate Professor at New York University's Linguistics Department and a senior scientist at Haskins Laboratories. His interests lie at the intersection of phonology and cognitive science.

Randall Gess is Associate Professor of Linguistics, Cognitive Science, and French, and Director of the School of Linguistics and Language Studies at Carleton University. His research interests are in historical phonology, the phonetics–phonology interface, and French and Romance phonology.

Heather Goad is an Associate Professor in Linguistics at McGill University. She works principally on the acquisition of phonology. Her work has been published in *Linguistic Review, Lingua, Language Acquisition*, and *Second Language Research*. Her research is currently funded by the Social Sciences and Humanities Research Council of Canada and the Fonds québécois de la recherche sur la société et la culture. She has been an Associate Editor of *Language Acquisition* since 2004.

Matthew Gordon is Professor of Linguistics at the University of California, Santa Barbara. He is the author of *Syllable weight: Phonetics, phonology, typology* (2006) and co-editor of *Topic and focus: Cross-linguistic perspectives on meaning and intonation* (2008). His research focuses on prosody, including stress and intonation, and the phonetic description of endangered languages.

Janet Grijzenhout is Professor of English Linguistics at the University of Konstanz and director of the Baby Speech Lab there. Her research focuses on the phonological representation of voicing and stricture, the phonology–morphology interface, infant speech perception, and the acquisition of prosody and morphology.

Naomi Gurevich received a Ph.D. from the Linguistics Department at the University of Illinois at Urbana-Champaign in 2003. The dissertation was published in the Outstanding Dissertations in Linguistics Series in 2004 under the title *Lenition and contrast: Functional consequences of certain phonetically conditioned sound changes*. Recently Naomi's research interests have shifted from theoretical linguistics to neurologically-based language disorders. She is currently working on a clinical certification in Speech Language Pathology.

Carlos Gussenhoven is Professor of General and Experimental Phonology in the Department of Linguistics at Radboud University, Nijmegen and Professor of Linguistics in the School of Languages, Linguistics, and Film at Queen Mary University of London. One of his research topics is the prosodic structure of English, including word stress and sentence intonation. Other research has focused on stress and tone in a variety of languages, including Dutch, Japanese, Nubi, and a group of Franconian dialects with a lexical tone contrast. Among his publications are *The phonology of tone and intonation* (2004) and *Understanding phonology* (1998, 2005).

Gregory R. Guy (Ph.D., Pennsylvania) is Professor of Linguistics at New York University, and has been on the faculty at Sydney, Temple, Cornell, Stanford, and York. He has taught at five Linguistic Institutes of the Linguistic Society of America, and three Institutes of the Associação Brasileira de Lingüística. He specializes in sociolinguistics and phonological variation and change, and works on Portuguese, English, and Spanish. His current research interests include the treatment of variation in linguistic theory, the relationship between individual and community grammars, and the theoretical treatment of grammatical similarity. His books include *Towards a social science of language* (1996) and *Sociolingüística quantitativa* (2007).

Daniel Currie Hall received his Ph.D. from the University of Toronto in 2007 with a thesis entitled "The role and representation of contrast in phonological theory." He has taught on all three campuses of the University of Toronto and at Queen's University and worked as a researcher at the Meertens Instituut of the Royal Netherlands Academy of Arts and Sciences, and is currently an Assistant Professor at Saint Mary's University in Halifax, Nova Scotia.

Nancy Hall is an Assistant Professor at California State University, Long Beach. She received her Ph.D. from the University of Massachusetts, Amherst in 2003, and has taught at Rutgers University, the University of Haifa, Ben-Gurion University of the Negev, and Roehampton University.

T. A. Hall is Associate Professor of Germanic Studies and Adjunct Associate Professor of Linguistics at Indiana University. He has published widely on a number of topics in general phonological theory and Germanic phonology in journals such as *Linguistics*, *Lingua*, *Natural Language & Linguistic Theory*, *Journal of Germanic Linguistics*, *Phonology*, *Linguistic Review*, *Journal of Comparative Germanic Linguistics*, and *Morphology*.

Michael Hammond is Professor of Linguistics and Department Head at the University of Arizona. He received his Ph.D. in Linguistics from UCLA in 1984. He is the author of numerous books and articles. His research interests are broad, including the metrical theory of stress, Optimality Theory, poetic meter, language games, computational linguistics, psycholinguistics, mathematical linguistics, syllable structure, probabilistic phonotactics, English phonology generally, and Welsh.

S. J. Hannahs is Senior Lecturer in Linguistics at Newcastle University. Co-author of *Introducing phonetics and phonology* (2005), much of his research has focused on prosodic structure, particularly at the interface between phonology and morphology. His recent and ongoing work has concentrated on the phonology and morphophonology of modern Welsh.

John Harris, Professor of Linguistics at University College London, writes on various topics connected with phonology, including phonological theory, the interface with phonetics, language impairment, and variation and change in English. Among his publications are the books *Phonological variation and change* (1985) and *English sound structure* (1994).

Jeffrey Heinz is Assistant Professor in the Department of Linguistics and Cognitive Science at the University of Delaware and has held a joint appointment with the department of Computer and Information Sciences since 2009. He received his Ph.D. in Linguistics from UCLA in 2007 and is keen to help bridge divides between theoretical phonology, computational linguistics, theoretical computer science, and cognitive science.

Ben Hermans is a senior researcher at the Meertens Institute in Amsterdam. He was trained as a Slavist and Germanicist. In 1994 he defended his thesis "The composite nature of accent" at the Free University. He now focuses on the tonal accents of the Limburg dialects, for example, "The phonological structure of the Limburg tonal accents" (2009), and is particularly interested in their formal representation. He also publishes on the history of the phonology of Dutch and its dialects.

José Ignacio Hualde (Ph.D. in Linguistics, 1988, University of Southern California) is Professor in the Department of Spanish, Italian, and Portuguese and the Department of Linguistics at the University of Illinois at Urbana-Champaign. He is author of *Basque phonology* (1991) and *The sounds of Spanish* (2005), co-author of *The Basque dialect of Lekeitio* (1994) and co-editor of *Generative studies in Basque linguistics* (1993), *Towards a history of the Basque language* (1995), *A grammar of Basque* (2003) and *Laboratory phonology 9* (2007), among other books. He has also published a number of articles on synchronic and diachronic issues in Basque and Romance phonology.

Harry van der Hulst is Professor of Linguistics at the University of Connecticut. He has published four books, two textbooks, and over 130 articles, and has edited 20 books and six journal theme issues in areas including feature systems and segmental structure, syllable structure, word accent systems, vowel harmony, and sign language phonology. He has been Editor-in-Chief of the international linguistic journal *The Linguistic Review* since 1990 and he is co-editor of the series Studies in Generative Grammar.

Elizabeth Hume is Professor and Chair of the Department of Linguistics at Ohio State University. She holds a Ph.D. and M.A. in Linguistics from Cornell University, an M.A. in French and Social Psychology of Language from McMaster University (Canada), and a B.A. in French from Université Laval (Quebec). Her research interests lie in language sound systems, cognitive science, language variation, and language change. She has published widely on topics including consonant–vowel interaction, feature theory, geminates, markedness, metathesis, sound change, the interplay of speech perception and phonology, and Maltese phonetics and phonology.

Brett Hyde received his Ph.D. in Linguistics from Rutgers University in 2001 and currently has an appointment at Washington University in St. Louis. His primary research interests are in metrical stress theory and related areas.

Larry M. Hyman is Professor of Linguistics at the University of California, Berkeley. He received a Ph.D. in Linguistics from UCLA in 1972. He has published several books (e.g. *Phonology theory and analysis, A theory of phonological weight*) and numerous theoretical articles in such journals as *Language, Linguistic Inquiry, Natural Language & Linguistic Theory, Phonology, Studies in African Linguistics*, and *Journal of African Languages and Linguistics.*

Sharon Inkelas is Professor of Linguistics at the University of California, Berkeley. She specializes in the phonology–morphology interface and has branched out in recent years into child phonology. With co-author Cheryl Zoll she published *Reduplication: Doubling in morphology* in 2005.

Gregory K. Iverson is Professor of Linguistics at the University of Wisconsin-Milwaukee and Research Professor at the University of Maryland Center for Advanced Study of Language. His research interests span the fields of historical linguistics, especially Germanic, the phonetics and phonology of laryngeal systems, especially Korean, and the acquisition of second language phonological patterns.

Jongho Jun is Professor of Linguistics at Seoul National University. His research interests are phonetics in phonology, variation in phonology, and the formal properties of Optimality Theory.

Barış Kabak is currently an Assistant Professor of English and General Linguistics at the University of Konstanz.

Ellen M. Kaisse is Professor of Linguistics at the University of Washington. She has co-edited the journal *Phonology* since 1988. Her research concentrates on the interactions of phonology with morphology and syntax, on distinctive features, and on the phonology of Modern Greek, Turkish, and Spanish.

Yoonjung Kang is an Assistant Professor in the Department of Humanities at the University of Toronto, Scarborough and the Department of Linguistics at the University of Toronto. Her area of specialization is phonology and its interface with phonetics and morphology, with a special focus on Korean.

Shigeto Kawahara is an Assistant Professor in Linguistics and Rutgers Center for Cognitive Science (RuCCs) at Rutgers University. He was awarded his Ph.D. from the University of Massachusetts, Amherst in 2007. His research focuses on the phonetics–phonology interface, experimental investigations of phonological judgments, corpus-based studies of verbal art, and studies on intonation and accents.

Robert Kennedy is a lecturer in the Department of Linguistics at the University of California, Santa Barbara. His areas of expertise include reduplication, prosodic morphology, hypocoristics, vowel systems of English varieties, and articulatory phonology.

John Kingston is a Professor in the Linguistics Department, University of Massachusetts, Amherst. His publications include: "Phonetic knowledge," *Language* (with R. L. Diehl, 1994); "Lenition," *Proceedings of the Third Conference on Laboratory Approaches to Spanish Phonology* (2007); "The phonetics–phonology interface," *Cambridge handbook of phonology* (2007); "On the internal perceptual structure of distinctive features: The [voice] contrast," *Journal of Phonetics* (with R. L. Diehl, C. J. Kirk, and W. A. Castleman, 2008); "Contextual effects on the perception of duration," *Journal of Phonetics* (with S. Kawahara, D. Chambless, D. Mash, and E. Brenner-Alsop, 2009); "Auditory contrast versus compensation for coarticulation: Data from Japanese and English listeners," *Language and Speech* (with S. Kawahara, D. Mash, and D. Chambless, in press).

Charles W. Kisseberth has just retired from his position as Professor of Linguistics at Tel Aviv University (where he taught from 1996 to 2010) and is also Emeritus Professor at the University of Illinois (where he taught from 1969 to 1996). He is best known for his work in theoretical phonology, where his work on "conspiracies" laid the foundations for Optimality Theory, and his work on Chimwiini prosody helped to lead the way to current studies in the phonology–syntax interface. His work over the last 30 years has focused on Bantu tonal systems, and Optimal Domains Theory has evolved out of that work. He is co-author with Michael Kenstowicz of *Generative phonology* (1979), a standard introduction to classical generative phonology.

Alexei Kochetov is currently an Assistant Professor in the Linguistics Department at the University of Toronto. He received his Ph.D. at the University of Toronto with the thesis "Production, perception, and emergent patterns of palatalization" (published in 2002). His research deals with various issues in the phonetics–phonology interface, cross-language speech production, and Slavic phonology and phonetics.

Astrid Kraehenmann holds a Ph.D. in Theoretical Linguistics from Konstanz University, Germany. Her main research interests are phonology, phonetics, the phonology–phonetics interface, historical linguistics, and Germanic languages.

Haruo Kubozono is Professor and Director at the National Institute for Japanese Language and Linguistics in Tokyo. His main publications include *The organization of Japanese prosody* (1993), "Mora and syllable" (in *The handbook of Japanese linguistics*, 1999), and "Where does loanword prosody come from?" (*Lingua* 116, 2006).

Darlene LaCharité is Professor of Phonetics and Phonology at Laval University. Her areas of research include loanword phonology (in collaboration with Carole Paradis), the L2 acquisition of phonetics and phonology, and creole phonology and morphology (in collaboration with Silvia Kouwenberg). She has published in a variety of linguistics journals, including *Linguistic Inquiry*, *Phonology*, *Journal of Linguistics* and *Journal of Pidgin and Creole Languages*.

William R. Leben is Professor Emeritus of Linguistics at Stanford University. He has worked on the phonology of tone in languages of West Africa and has also co-authored pedagogical works on Hausa. He continues to work on phonology in Kwa languages of Côte d'Ivoire. The second edition of *English vocabulary elements*, a textbook he co-authored, was published in 2007.

Clara C. Levelt is Associate Professor at the Linguistics Department of Leiden University, and affiliated to the Leiden University Centre for Linguistics (LUCL) and the Leiden Institute for Brain and Cognition (LIBC). She received her Ph.D. in 1994, and has worked on child language phonology ever since. In 2007 she received a prestigious research grant from the Netherlands Organization for Scientific Research (NWO), which enabled her to set up a baby lab. Combining insights from perception and production data, phonology and phonetics, she tries to uncover the source of children's deviating productions.

Susannah V. Levi is an Assistant Professor in the Department of Communicative Sciences and Disorders at New York University. She received her Ph.D. in 2004 from the Department of Linguistics at the University of Washington. She then completed a three-year postdoctoral fellowship at Indiana University in the Department of Psychological and Brain Sciences. She has worked on glides, Turkish stress and intonation, and most recently on speech perception and language processing in children and adults.

Yen-Hwei Lin is Professor of Linguistics at Michigan State University, and taught at the 1997 and 2003 Linguistic Society of America Linguistic Institutes. Her research has focused on phonological representations and constraints in the theoretical context of non-linear phonology/morphology and Optimality Theory. She is author of *The sounds of Chinese* (2007) and editor of *Special Issue on Phonetics and Phonology* (*Language and Linguistics* 5.4, 2004).

Anna Łubowicz is Assistant Professor of Linguistics at the University of Southern California. Her research interests lie in the investigation of the role of contrast in phonology and morphology, lexical phonology, and the morphology–phonology interface. She specializes in Slavic languages and holds a Courtesy Appointment in the Department of Slavic Languages at the University of Southern California.

Andrew Martin received his Ph.D. in Linguistics from UCLA in 2007. His research is focused on understanding early phonological and lexical learning. He is also interested in how a language's lexicon changes over time, and how a word's phonological properties affect that word's ability to survive and spread in a speech community. He is currently a post-doctoral researcher at the Laboratory for Language Learning and Development in the RIKEN Brain Science Institute near Tokyo.

Joan Mascaró studied at the Universitat de Barcelona and at MIT where he got his Ph.D. in linguistics in 1976. He has taught at Cornell University and at the Universitat Autònoma de Barcelona, where he is currently full professor. His main research areas are linguistic theory, phonological theory, the phonology–morphology interface, and the phonological and morphological analysis of Romance languages.

April McMahon is Forbes Professor of English Language, and Vice Principal for Planning, Resources and Research Policy, at the University of Edinburgh. She co-edits the journal *English Language and Linguistics*. Her research focuses on the interaction of phonological theory and sound change, and methods for the comparison and classification of accents and languages.

Jacques Mehler is the director of the Language, Cognition, and Development laboratory at the International School for Advanced Studies, Trieste, Italy (SISSA). After obtaining a Ph.D. from Harvard University in 1964, he worked at the National Center for Scientific Research (CNRS) in Paris, France, from 1967 until 2001. In 1972, he founded *Cognition*, an international journal of cognitive science, and acted as Editor-in-Chief until 2007. He became Directeur de Recherche at CNRS in 1980 and was elected Directeur d'Études at the École des Hautes Études en Sciences Sociales in 1982. He has published influential experimental studies of language acquisition in the first year of life, and has also explored early bilingualism. He is a member of the American Academy of Arts and Sciences (2001), the American Philosophical Society (2008) and the Academic Europaea. He was awarded a Doctor Honoris Causa from Utrecht University (2009), and from Université Libre de Bruxelles (1995). His publications are available at: www.sissa.it/cns/lcd/publications.html.

Jeff Mielke is Assistant Professor of Linguistics at the University of Ottawa and Co-director of the Sound Patterns Laboratory. He completed his Ph.D. at Ohio State University in 2004 and undertook postdoctoral research at the Arizona Phonological Imaging Laboratory at the University of Arizona before moving to Ottawa in 2006. His work focuses on the way phonological patterns reflect influences such as physiology, cognition, and social factors. He is the author of *The emergence of distinctive features* (2008).

Amanda Miller is a Visiting Assistant Professor at Ohio State University. Her research focuses on the phonetics and phonology of African languages, particularly Khoesan languages. She has studied gutturals, complex segments (clicks and labial-velars), and contour segments (affricates and airstream contours). She has investigated acoustic voice quality cues and the role of acoustic similarity in a Guttural Obligatory Contour Principle constraint; as well as tongue root retraction in clicks and labial-velars, and its role in C-V co-occurrence patterns. She has also published papers on reduplication and the tonal phonology of Khoesan languages.

Brett Miller is a graduate student at the University of Cambridge. His interests include Indo-European phonology, especially stops; the phonetics–phonology interface; feature contrast, representations, and interaction in rules and constraints; and typology *vis-á-vis* the comparative method.

Bruce Morén-Duolljá is a Senior Researcher at the Center for Advanced Study in Theoretical Linguistics at the University of Tromsø. He has published on synchronic and diachronic phonology, including Slavic palatalization, Icelandic preaspiration and Thai tones. He is the author of *Distinctiveness, coercion and sonority: A unified theory of weight* (2001).

Marina Nespor is Professor of General Linguistics at the University of Milano-Bicocca. She has focused her research on how the phonological shape of an utterance conveys information about its syntactic structure, the so-called theory of prosodic phonology. She has also investigated how prosody is used in comprehension and during language acquisition. Her book *Prosodic phonology* (co-authored with I. Vogel, 1986) is a citation classic and she has numerous articles in peer-reviewed journals.

Andrew Nevins is a Reader in Linguistics at University College London. His research has primarily focused on phonological and morphological theory, and the relations between different modules of the grammar. He has worked on locality, markedness, contrast, reduplication, the nature of underlying representations, and the structure of the morphological component. He is the author of *Locality in vowel harmony* (2010) and the co-editor of two books, with Bert Vaux, *Rules, constraints, and phonological phenomena* (2008), and with Asaf Bachrach, *Inflectional identity* (2008).

David Odden is Professor of Linguistics at Ohio State University. His areas of research specialization include phonological theory and language description, especially the structure of African languages. He served as Editor of *Studies in African Linguistics* from 2003 to 2009 and was Associate Editor of *Phonology* from 1998 to 2008. Recent publications include *Introducing phonology* (2005), "Ordering," in Vaux and Nevins, *Rules, constraints, and phonological phenomena* (2008), and "Tachoni verbal tonology," *Language Sciences* (2009).

Jaye Padgett is Professor of Linguistics at the University of California, Santa Cruz. At the broadest level his research is on the interplay of phonology and phonetics, with special attention to the role that perception plays in shaping phonological patterns. Though his research bears on larger questions it often focuses particularly on Russian phonetics and phonology.

Carole Paradis is Professor of Phonology at Laval University. She proposed a theory of constraints in 1987, as a visiting scholar at MIT. She co-edited a book on the special status of the coronals with J.-F. Prunet in 1991. With R. Béland's collaboration, she has drawn a parallel between aphasic errors and loanword adaptations. This study, extended to other speech error types, has been published in journals such as *Aphasiology* and has led to the construction of more optimal speech pathology exercises.

Steve Parker is a Professor with the Graduate Institute of Applied Linguistics (GIAL) in Dallas. He graduated from the doctoral program in linguistics at the University of Massachusetts, Amherst in 2002. He has served as a teacher and consultant with the Summer Institute of Linguistics (SIL International) for 30 years. In that capacity he has carried out direct fieldwork and research on a number of indigenous languages of South America and Papua New Guinea, two of which are now extinct.

Sharon Peperkamp is a senior researcher in the Laboratoire de Sciences Cognitives et Psycholinguistique at the École Normale Supérieure in Paris. Her main research interests are in experimental and computational approaches to speech perception and phonological acquisition.

Marianne Pouplier obtained her Ph.D. from Yale Linguistics in 2003 and is now a principal investigator of an Emmy-Noether Research Group funded by the Deutsche Forschungsgemeinschaft at the Institute of Phonetics and Speech Processing, University of Munich. She has also been affiliated with Haskins Laboratories for a number of years. Her research focuses on the interaction of speech planning and motor control and the phonetics–phonology interface. In particular, she has worked on speech errors, phonetic correlates of syllable structure, and word boundary assimilation.

Pilar Prieto holds a Ph.D. in Romance Linguistics at the University of Illinois at Urbana-Champaign, and is currently working as the coordinator of the Grup d'Estudis de Prosòdia at the Universitat Pompeu Fabra in Barcelona. Her main research interests focus on how melody and prosody work in language and how they interact with other types of linguistic knowledge. She is also interested in how babies acquire sounds and melody together with grammar, and how these components integrate in the process of language acquisition.

Douglas Pulleyblank is Professor of Linguistics at the University of British Columbia. His research has focused on the phonology and morphology of Nigerian languages, particularly Yoruba. He has worked extensively in autosegmental and optimality-theoretic frameworks, examining phenomena such as tone and vowel harmony.

Eric Raimy is an Associate Professor at the University of Wisconsin-Madison. He received his Ph.D. from the University of Delaware in 1999. He is the author of *The phonology and morphology of reduplication* (2000) and co-edited *Contemporary* *views on architecture and representations in phonology* (2009) and *Handbook of the syllable* (in press) both with Charles Cairns.

Anthi Revithiadou is an Assistant Professor of Linguistics at the Aristotle University of Theory, with emphasis on the structure of phonological representations and the issue of parallel grammars, and has published in journals (*Lingua, Linguistic Review, Journal of Greek Linguistics*) and edited volumes.

Curt Rice is the Vice President (Prorektor) for Research and Development at the University of Tromsø in Tromsø, Norway. He was the founding Director of the Center for Advanced Study in Theoretical Linguistics: A Norwegian Center of Excellence (CASTL) from 2002 to 2008, at the same university.

Keren Rice is University Professor and Canada Research Chair in Linguistics and Aboriginal Studies at the University of Toronto. She holds an M.A. and Ph.D. in Linguistics from the University of Toronto and a B.A. in Linguistics from Cornell University. Her research interests lie in language sound systems, contrast and markedness, interfaces with phonology, language variation and language change, and Athabaskan languages as well as ethics and responsibilities of linguists in fieldwork. She has published on topics including feature theory, sonorants, markedness, language change, and Athabaskan phonology and morphology.

Anastasia K. Riehl is Director of the Strathy Language Unit at Queen's University in Kingston, Canada, where she also teaches in the Linguistics Program. She received a Ph.D. in Linguistics from Cornell University. Her research interests include the phonology–phonetics interface, endangered language documentation, Austronesian languages, and varieties of English.

Jason Riggle is Assistant Professor of Linguistics at the University of Chicago, and his main research areas are phonology, learnability, and computational linguistics. Much of his research focuses on the ways that specific models of grammar, learning, and communication interact to make predictions about linguistic typology with special emphasis on the frequencies with which patterns are observed within and across languages.

Sharon Rose is Associate Professor of Linguistics at the University of California, San Diego. She has published journal articles covering topics such as consonant harmony, syllable contact, Semitic root structure, reduplication, phonotactics, and the interaction between tone and syllable structure. She specializes in African languages spoken in Ethiopia, Eritrea, and Sudan; her current research is an investigation of the Sudanese language, Moro, funded by the National Science Foundation.

Jerzy Rubach holds an appointment as Professor of Linguistics at two universities: the University of Iowa in the United States and the University of Warsaw in Poland. His expertise is primarily in Germanic and Slavic languages. He has published six books and 75 articles. His work has appeared in numerous journals, including *Linguistic Inquiry, Language, Phonology,* and *Natural Language & Linguistic Theory.*

Joseph C. Salmons is Professor of German at the University of Wisconsin–Madison and Executive Editor of *Diachronica: International Journal for Historical Linguistics.* His research interests include language change, sound systems, language contact, and language shift, all often involving data from the Germanic languages.

Wendy Sandler has been investigating the phonology, morphology, and prosody of American Sign Language and Israeli Sign Language for many years, beginning with her graduate work at the University of Texas at Austin, where she earned her Ph.D. Sandler has authored or co-authored three books on sign language linguistics. Currently, with colleagues Mark Aronoff, Irit Meir, and Carol Padden, she is investigating a new sign language that arose in an insulated Bedouin community in the Negev desert of Israel.

Tobias Scheer is currently Directeur de Recherche at the Centre National de la Recherche Scientifique in France. He works at the laboratory Bases, Corpus, Langage (UMR 6039) in Nice, of which he is the director. Being a phonologist, his main interests lie in syllable structure thematically speaking, in the (Western) Slavic family as far as languages are concerned, and in diachronic study. In 2004, he published a book on a particular development of GP, so-called CVCV (or strict CV), and a book on the (history of the) interface of phonology and morpho-syntax.

Misun Seo is an Assistant Professor in the Department of English Language and Literature at Hannam University in Daejeon, Korea. She received her Ph.D. in Linguistics at Ohio State University; her research interests include phonology, its interface with phonetics, and L2 acquisition.

Kimary Shahin is a phonetician/phonologist who specializes in Arabic and Salish. She also investigates first language acquisition and contributes to the documentation and revitalization of indigenous languages in Canada.

Mohinish Shukla has a background in molecular genetics, and holds a Ph.D. in Cognitive Neuroscience from the Scuola Internazionale Superiore di Studi Avanzati (SISSA), Italy. He is currently doing postdoctoral work at the University of Rochester, New York. He is interested in human cognition, with a focus on infant cognitive development, particularly the development of linguistic abilities from phonology to syntax. He is also interested in the neural bases of cognitive behavior, which he primarily explores using near-infrared spectroscopy in infants and adults.

Daniel Silverman earned his degree at UCLA in 1995 under the tutelage of Donca Steriade and Peter Ladefoged. He has published widely on phonology and phonetics, including *A critical introduction to phonology: Of sound, mind, and body* (2006), and *Neutralization: Rhyme and reason in phonology* (forthcoming). He is currently on the faculty of San José State University.

Jennifer L. Smith teaches phonology, phonetics, and Japanese linguistics at the University of North Carolina, Chapel Hill. Her research explores the ways in which interfaces between phonology and other domains of linguistics affect phonological constraints and representations.

Péter Szigetvári gives courses on phonology, linguistics, information technology, and typography at Eötvös Loránd University, Budapest. His main research areas include phonotactics and consonant lenition.

Nina Topintzi has taught as a Teaching Fellow in England and Greece (2006–2010) and will soon commence her appointment as Assistant Professor in Phonology at the English Department, Aristotle University of Thessaloniki. She has published in journals including *Natural Language & Linguistic Theory* and the

Journal of Greek Linguistics. Beyond onsets, her research interests include stress, the syllable, and various weight-based phenomena from a typological perspective.

Miklós Törkenczy is a Professor at the Department of English Linguistics and the Theoretical Linguistics Department of Eötvös Loránd University Budapest and a senior researcher at the Research Institute for Linguistics of the Hungarian Academy of Sciences. He is co-author of *The phonology of Hungarian* (2000).

Jochen Trommer received his Ph.D. from the University of Potsdam, has worked as lecturer at the University of Osnabrück and is currently Lecturer in Phonology at the Department of Linguistics at the University of Leipzig. His work focuses on the theoretical and typological aspects of phonology and morphology, especially prosody, non-concatenative morphology, hierarchy effects in agreement morphology, and affix order. His publications include "Case suffixes, postpositions, and the phonological word in Hungarian," *Linguistics* 46(1) (2008); "Hierarchy-based competition and emergence of two-argument agreement in Dumi," *Linguistics* 44(5) (2006); and "The interaction of morphology and syntax in affix order," *Yearbook of Morphology 2002* (2003).

Suzanne Urbanczyk is an Associate Professor in the Linguistics Department at the University of Victoria. She received her Ph.D. from the University of Massachusetts, Amherst and has published on Salish reduplication and non-concatenative morphology in *Natural Language and Linguistic Theory* and *Linguistic Inquiry*. Her current research focuses on the role repetition plays in structuring language, and she is developing a model of morphology which is grounded in the mental lexicon, with an empirical focus on word-formation in Salish and Wakashan languages.

Adam Ussishkin is an Associate Professor in the Linguistics Department at the University of Arizona, where he also holds joint appointments in Cognitive Science and Near Eastern Studies. His areas of research include phonology, morphology, and lexical access, with an empirical focus on Semitic languages.

Bert Vaux is University Reader in Phonology and Morphology at the University of Cambridge and a Fellow of King's College, Cambridge. He is primarily interested in phenomena that shed light on the structure and origins of the phonological component of the grammar, especially in the realms of psychophonology, historical linguistics, and sociolinguistics. He also enjoys working with native speakers to document endangered languages, especially dialects of Armenian, Abkhaz, and English.

Rachel Walker is Associate Professor of Linguistics at the University of Southern California. She is the author of *Nasalization, neutral segments and opacity effects* (2000) and *Vowel patterns in language* (forthcoming). She has published widely on topics involving long-distance assimilation and copy, as seen in systems of harmony and reduplication. Her current research investigates the interaction of vowel patterns with positions of prominence in the word.

Natasha Warner has been a faculty member in Linguistics at the University of Arizona since 2001 and has worked at the Max Planck Institute for Psycholinguistics. Her interests are in the three-way interface of phonetics, experimental phonology, and psycholinguistics, with language interests in Dutch, Japanese, Korean, and English. She also works on revitalization of the Mutsun language.

Janet C. E. Watson is Professor of Arabic Linguistics at the University of Salford, UK. She has previously worked at the universities of Edinburgh and Durham, and held visiting posts at the universities of Heidelberg and Oslo. She has published extensively on Yemeni Arabic, and since 2006 has been working on the documentation of the Modern South Arabian language, Mehri. She is currently preparing a syntax of Mehri for publication. Her publications include *Phonology and morphology of Arabic* (2002) and *A syntax of San'ani Arabic* (1993).

Andrew Wedel is on the faculty of the Department of Linguistics and the Cognitive Science Program at the University of Arizona. His primary interests lie in exploring the causes for, and the interaction of, two opposing tendencies evident in language change: a tendency toward pattern-coherence, and a tendency to preserve semantically relevant contrasts. Relevant work in this domain includes "Feedback and regularity in the lexicon," *Phonology* 24, and with Juliette Blevins, "Inhibited sound change: An evolutionary approach to lexical competition," *Diachronica* 26 (2009).

Jeroen van de Weijer is Full Professor of English Linguistics at Shanghai International Studies University (College of English Language and Literature) within the Chinese "211-Project." He is a Distinguished University Professor of Shanghai (Oriental Scholar), an award bestowed by the Shanghai Municipal Education Commission. He has published widely on segmental structure, Optimality Theory, English phonology, and East Asian languages. His current research is focused on combining phonological and psycholinguistic theories.

Richard Wiese has been a Professor of German Linguistics at the Philipps-Universität, Marburg, Germany, since 1996. His work concentrates on theoretical linguistics, phonology, and psycholinguistics. He has written a monograph, *The phonology of German* (2000) and numerous articles on issues of (German) phonology, morphology, psycholinguistics, and orthography, and is Co-editor of the book series Linguistische Arbeiten.

Ronnie Wilbur, Professor and Director of Linguistics and Professor of Speech, Language, and Hearing Sciences, received her Ph.D. in Linguistics at the University of Illinois, Urbana-Champaign. She has taught at the University of Southern California, Boston University, University of Amsterdam, University of Graz, Austria, and University of Zagreb. Her 1973 dissertation "The phonology of reduplication" was reissued as a "Classic in Linguistics" in 1997. She was Founding Editor of the journal *Sign Language & Linguistics* and Editor from 1998 to 2006.

Matthew Wolf received his Ph.D. in 2008 from the University of Massachusetts, Amherst. After having been a Visiting Assistant Professor at Georgetown University and a Visiting Lecturer at Yale University, he is currently a Postdoctoral Associate at Yale. His research has focused primarily on aspects of the phonology–morphology interface, including process morphology, allomorph selection, and paradigm gaps, as well as on opacity and serial versions of Optimality Theory.

Moira Yip is Professor of Linguistics (Emerita) at University College London. Previously she was at the University of California, Irvine, and Brandeis University. She received her Ph.D. from MIT in 1980, studying under Morris Halle. She has published extensively on tone, reduplication, distinctive feature theory, and loanword phonology. Much of her work has been on Chinese languages. Recently she has become interested in comparisons between birdsong and human phonology, and is now publishing in this area.

Alan C. L. Yu is Associate Professor of Linguistics at the University of Chicago. He is the author of *A natural history of infixation* (2007) and co-editor (with John Goldsmith and Jason Riggle) of the *Handbook of phonological theory*, 2nd edition (forthcoming). His work on phonetics and phonology has appeared in *Language*, *Phonology*, *Journal of Phonetics*, and *PLoS One*.

Draga Zec is Professor of Linguistics at Cornell University. She has worked in several areas of phonology and its interfaces: the moraic theory of syllable structure, the representation of pitch accent, and both the phonology-morphology and the phonology–syntax interfaces.

Elizabeth C. Zsiga is an Associate Professor in the Department of Linguistics at Georgetown University. Her research primarily addresses the phonetic and phonological patterns that occur at and across word boundaries in connected speech. Her publications have examined a range of processes in different languages, including final consonant deletion and palatalization in English, vowel assimilation in Igbo, palatalization in Russian, tone simplification in Thai, voicing in Setswana, pitch accent in Serbian, and most recently nasalization and voicing in Korean and Korean-accented English.

Preface

Progress in phonology, like other disciplines, grows out of debate. Every journal article, every conference paper, every book chapter, every book on phonology, can be seen as a contribution to one or more discussions on some theoretical topic. New data are sought to sharpen theoretical claims and new theories are proposed to accommodate previously undescribed data. Already familiar data paradigms are frequently appealed to in arguing that some new theory fares better than its competitors in the description of various phenomena.

Arguably, the synchronic study of phonology is currently celebrating its first centenary: Ferdinand de Saussure was teaching his *Cours de linguistique générale* up to his death in 1913 (it was published posthumously in 1916). One hundred years of debates have yielded many insights into the sound structure of human language. We now know much more about a whole range of specific phonological phenomena, including topics such as vowel harmony, the typology of word stress systems, and the structure of affricate sounds. We know more about how sound systems interact with morphological and syntactic systems, and about the importance of taking factors like variation and frequency into account in the study of phonology. Phonologists have not developed a theory which completely captures each of those phenomena (let alone one which captures every aspect of all of them in a uniform way). But at least we have a much better picture of some of the properties which a successful theory of phonology should have.

Alongside external and universal challenges, such as university administrators who do not see the importance of something as esoteric as the study of sound systems, we can identify at least two internal dangers which challenge our field.

The first is that many debates are abandoned at some point, and then forgotten, with the issues involved sometimes being rediscovered much later, without the earlier research necessarily being known to the new generation of researchers. The reasons for this are often perfectly understandable, and there is probably no way to avoid this state of affairs completely. After lengthy discussions in the literature on, say, the relationship between continuancy and place of articulation, apparently involving arguments for and against almost every logically possible view on the issue, the topic may seem to be intractable, and scholars may tire of it and move on to new topics of debate. Thus we may avoid unfruitful attempts to solve problems for which we just do not have the right tools at that particular moment.

However, these discussions will frequently have led to at least partial agreement on important properties of the phenomenon at hand, even if only that the phenomenon is extremely complicated: just positing, for instance, some feature-geometric structure in which continuancy and place of articulation are in an unambiguous universal relation to each other will encounter many problems that help us better understand the phenomenon, even if they do not lead to a generally accepted solution. However, by abandoning the debate on the topic, we run the risk of losing that knowledge. The danger is that new generations will have to rediscover the subtleties of the phenomenon when the topic is taken up again.

The second danger is over-specialization. Saussure was an all-round linguist, with a deep understanding of much of (Western) linguistics as it was known at the time. Today there are very few, if any, people who can make such a claim. Scholars with a thorough understanding of phonology may have some knowledge of neighboring disciplines such as morphology or phonetics, but usually not even of both of those. There are many more phonology talks given every year at specialized phonology conferences than there are phonology talks at general linguistics conferences. And even within the field itself there are fairly well-established dividing lines: an expert on intonation is unlikely to be completely up to date on the literature on coronals; somebody who studies the Iambic–Trochaic Law may skip the talks on sign language phonology in the local phonology workshop; somebody who works on stress may have little knowledge of segmental phonology. This means that crucial insights within one sub-discipline of phonology are becoming less and less accessible to phonologists working within other sub-disciplines.

The Blackwell Companion to Phonology does not have the ambition to offer solutions to these problems; indeed, they are probably unavoidable. Yet a tool such as this allows us to document at least some of the many insights into human language that phonologists have gathered in the past decades, and also to give an overview of what at present seem to be the major issues that those interested in sound structure are thinking and arguing about.

The *Companion* is in essence an encyclopedia of case studies. Each chapter addresses some topic which has been debated in one way or another in the history of our field. Authors were invited to concentrate primarily on the empirical arguments that have been put forward by the various sides in such debates. Because of this concentration on case studies, there are many topics we have ignored. For instance, there is a chapter on coronals, but not one on labials, simply because there has been much more discussion in the phonological literature on the internal structure and the behavior of coronal sounds than on their labial counterparts. Similarly, there is a chapter on palatalization, but not one on labialization, again because the former has been discussed broadly in the phonology literature while the latter has not.

Some chapters have turned out more like the case studies we originally had in mind than others. Inevitably, some chapters had to be organized differently, for instance those concentrating more on a specific theoretical device (such as constraint conjunction or rule ordering) than on some empirical phenomenon. However, even the authors of these chapters were asked to provide some discussion of the data which led scholars to develop such theoretical concepts in the first place.

We are, of course, conscious of the fact that the reader will find that many possible topics are missing from the *Companion*, including some hotly debated ones

– the 124 case studies included could easily have been doubled in number. There are various reasons for this, perhaps most importantly the need to keep an already substantial enterprise within reasonable bounds, so that many topics which had a good claim to be included had to be discarded. Sometimes, too, chapters have taken a rather different direction from that originally anticipated – in such cases the editors generally felt that including such chapters was preferable to imposing an over-rigid framework into which the authors would have to fit their material. It may, however, prove possible to add further topics to the online version of the *Companion* (www.companiontophonology.com), which will provide features not available in the print version, and which is anticipated to grow in the next few years.

In spite of this, a great deal of ground is covered in this first print edition. Phonology as a discipline is theoretically and methodologically in a period of great change, making it a very exciting time to be a phonologist. We find scholars constructing formal theories working alongside others who apply statistical methods adopted from the social and computational sciences, while those who prefer to test ideas with the use of experimental techniques may be working along-side those who collect their data in the field.

This diversity is reflected in the group of authors. Altogether, almost 150 phonologists have been involved in the production of this reference work – a significant sample of the population of phonologists. The insights of many major frameworks of the beginning of the twenty-first century, such as Articulatory Phonology, Evolutionary Phonology, Government Phonology, Laboratory Phonology, and Optimality Theory, are presented in various chapters, and in some cases compared explicitly. Where appropriate, authors were also invited to end their chapters by presenting their own views on the topic under discussion, and sometimes their own analyses of particular phenomena, and many have done so. As well as being a collection of case studies, then, the *Companion* can also be read as an overview of theoretical positions within phonology at the beginning of the twenty-first century.

As already noted, the *Companion to Phonology* comes in two different shapes. One is a traditional print edition. The other is an electronic edition, published online, and convenient for searching and downloading individual papers. The chapters in the print edition are organized thematically into five volumes of more or less equal size.

The first volume opens with a number of chapters on general phonological topics, for example markedness and learnability, which will be of wide general interest to those working in the field. The bulk of this volume is devoted to chapters on topics in segmental phonology, and includes studies of individual classes of sounds, such as the chapters on clicks and fricatives, as well as more general discussions on the organization and structure of speech sounds (for example the chapter on distinctive features).

The chapters in Volume II contain discussion of topics in suprasegmental phonology as well as prosody (which, at least on some views, amount to the same thing). The division between Volumes I and II – and indeed the organization of the volumes in general – has not always been straightforward, as there are so many intersecting parameters involved; the chapter on geminates has been included in the volume dealing with suprasegmental phonology, although, as noted in the chapter, consonantal length is considered by some to be a segmental property.

Volume III deals with phonological processes. Again, many of these chapters could have found a home in Volume I or II, as the processes considered in them typically involve segmental or suprasegmental phonology.

The chapters in Volume IV are concerned with interfaces with other branches of linguistics, for example morphosyntax or phonetics, and also with topics which some might consider to be at least partly extralinguistic (such as frequency or variability).

Volume V contains language-specific issues, concentrating in particular on the discussion of processes and phenomena in individual languages or language families which have played a major role in general phonological discussion and the development of theories and analyses with a wider scope than the particular phenomena dealt with here. Chapters have been included on topics such as vowel harmony in Hungarian and Turkish and flapping in English, but in this volume, in particular, there were many more potential topics than could in the end be included.

The division into volumes is thus to some extent arbitrary. However, we do not anticipate that readers will make use of the material in the *Companion* on a volume-by-volume basis, but rather as a complete reference work. To this end, every chapter contains a substantial number of cross-references to other chapters in the five volumes dealing with related topics.

We have been working on *The Blackwell Companion to Phonology* for roughly five years. During this period we have been heartened by the enthusiasm with which many of our colleagues have reacted to the plans for the project, both those who have contributed chapters and those who were not able to take part in the project for one reason or another. It appears that the phonological community at large believes that a reference work of this size will be valuable for progress in the field. We believe that the *Companion* shows how healthy and lively the field of phonology currently is. A great deal of knowledge has been accumulated, and many insights have been gained. Vast numbers of questions are still unanswered; the debate goes on.

That an enterprise of this magnitude has come to fruition in a comparatively short period of time is primarily due to the help of a great number of people, primarily, of course, the authors of the 124 chapters, nearly all of whom also acted as reviewers of other contributions to the *Companion*. We thank them all for their willingness to participate in the project, even when confronted with impossible deadlines for submitting chapters, writing reviews, and responding to editorial queries. We are very happy that April McMahon is one of those authors; it was April who, as prospective editor-in-chief, first developed the plans for this *Companion* in conjunction with the publishers, and set up the rest of the editorial team – unfortunately her many administrative obligations meant that she had to withdraw from editing the project at a fairly early stage.

We are extremely grateful to Tessa Obbens, who took on the task of editing and making uniform the large number of references (over 10,000) in the 124 chapters.

We also would like to thank Martin Everaert and Henk van Riemsdijk, who are general editors of the series of Blackwell Companions to Linguistics, of which *The Blackwell Companion to Phonology* is the second, and who also edited *The Blackwell Companion to Syntax*. We have benefited from their helpful comments.

We are grateful to everyone at Wiley-Blackwell for their support and patience, in particular the commissioning editor, Danielle Descoteaux, who took over the

project from her predecessor, Ada Brunstein, before the editorial team was complete, and has been an invaluable source of enthusiastic support ever since. Barbara Duke, production editor in Oxford, has managed to keep us on the straight and narrow even when we fell so far behind schedule that final deadlines threatened to be missed. The project managers, initially Louise Spencely and thereafter Janey Fisher, together with their large team of copyeditors and proofreaders, have been of invaluable help in preparing these volumes, and in guiding us through some of the more esoteric conventions of the publishing world. We thank them for their unfailing good humor and efficiency in the face of our persistent questions, and hope that they will be pleased with the final result.

REFERENCE

Saussure, Ferdinand de. 1916. *Cours de linguistique générale.* Lausanne & Paris: Payot.

1 Underlying Representations

JENNIFER COLE
JOSÉ IGNACIO HUALDE

Language is acquired and experienced primarily through the medium of speech or the manually signed signal. A primary goal of phonology, restricted here to the context of spoken language, is to discover the elements that serve as the building blocks of speech. Considering that languages differ in their spoken forms, two further questions for an understanding of phonology concern the relations between the sound elements that give shape to the phonological system of an individual language, and the constraints that determine how these sound elements may pattern in the formation of words and phrases in that language.

Over many centuries of scholarship and across continents, linguists have pursued answers to these questions for the practical purpose of providing a straightforward orthography for particular languages (see Pike 1947), explicating a method for describing the phonological component of individual languages, or for the scientific purpose of identifying the mental encoding of phonological form in the minds of the native speaker/hearer. Differences in the relative priority accorded to practical and scientific purposes have resulted in differences in the principles and methods of competing schools of phonology. But all approaches, from the work of the Sanskrit scholar Patanjali in the second century BCE to the theories that emerged during the heyday of European and American phonology in the twentieth century, presume that the basic elements of spoken language are at some level of abstraction from the physical form of speech as experienced by the speaker/hearer. The representation of words in terms of abstract elements is posited as a *basic* or *underlying representation* (UR) in nearly every phonological theory to the present day. Theories differ in the status of the UR (as an artifact of descriptive analysis, or part of the cognitive system of language), its relation to morphological form and phonetics, and whether it may encode morphosyntactic context, reflecting differences among theories in the kinds of data considered as primary evidence for phonological form. Different proposals for UR also reflect differences in the scope of the proposed theory, e.g. in modeling diachronic or synchronic phenomena, dialectal or style-dependent variation, corpus data, speaker intuitions, child productions, or instances of the intentional, creative manipulation of phonology in poetry or language games.

1 Underlying representations in phonemic theories

Phonological theories of the late nineteenth and early twentieth century take the *phoneme* as the basic element of phonological analysis (see CHAPTER 11: THE PHONEME). Jan Niecislaw Baudouin de Courtenay and his student Mikołaj Kruszewski of the Kazan school (established in the mid-1870s) introduced the phoneme as a mental construct encoding the "image" of a sound as it is perceived and recognized, and as the abstract units with which phonological alternations may be characterized (Baudouin de Courtenay 1871). The notion of the phoneme as an abstraction from the acoustic and articulatory manifestation of speech was also expressed in the contemporaneous work of Ferdinand de Saussure, published posthumously in 1916, and recognized as the origin of structuralist linguistic analysis. De Saussure's "sound images," corresponding to what other scholars would term "phonemes" (Anderson 1985: 38–40), were characterized in terms of the properties that distinguish between the abstract sound units. And while Baudouin de Courtenay's view evolved to assign psychological reality to the phoneme as a unit of representation, de Saussure did not share this attribution, emphasizing instead the importance of the rules that relate sound representations (Anderson 1985: 53, 68). Despite de Saussure's rejection of the phoneme as constituting a distinct level of representation – an underlying form – his work profoundly influenced a later generation of scholars who focused intensely on the question of phonemes as units of representation, notably in the work of Trubetzkoy (1939) and Jakobson (1949) of the Prague School and of American structuralist linguists such as Bloomfield (1933) and Harris (1944, 1960).

Both the Prague School and American structuralism adopted de Saussure's view of phonemes as being characterized in terms of a system of contrast (see CHAPTER 2: CONTRAST). The Prague School notion was that phonemes are elements that are related to one another in a system of oppositions that define lexical contrast. Similarly, Jones (1967: 10) defines the phoneme as "a family of sounds in a given language which are related in character and are used in such a way that no one member ever occurs in a word in the same phonetic context as any other member," and explains that what phonemes do "is to distinguish words from one another" (1967: 265). The American structuralists held a similar notion, and focused on the method for determining the phonemic representation of words based on observations of phonetic form.

In a first sense, any representation of the utterances of a language in terms of contrasting phonemes can be construed as providing an *underlying* representation of those utterances. Thus, the form **'fonim**, which we find in Kenyon and Knott's (1953) *Pronouncing dictionary of American English* for what is conventionally written *phoneme*, is to be taken as the representation that underlies the infinitely diverse actual and potential productions of this word by native speakers of American English. This UR is in terms of the contrasting segment-sized units of the language.

An important claim behind phonemic theories, by and large borne out by everyday experience, is that, given an adequate phonemic representation, a native speaker of the language will know how to pronounce a previously unknown word accurately, in all phonetic detail. That is, a native speaker of American English who encounters, say, the word *phoneme* for the first time in an English text will know how to pronounce it accurately upon consulting Kenyon and Knott's dictionary. Words containing the same sequence of phonemes cannot differ in any detail of their pronunciation. If they do, that would indicate that they have been incorrectly transcribed as having identical underlying phonemic representations. In the case of a language whose conventional orthography follows the phonemic principle to a greater extent than English, such as Spanish, it is not unusual for very small children to convincingly read the newspaper aloud even though a great percentage of the words that they are reading may be unknown to them (so that, in fact, they may not understand much of what they are reading).

A first hypothesis of the theory of phonemic transcription is thus that all utterances in a language can be analyzed as combinations of a small set of phonemes (consonants, vowels, and prosodic phonemes). Often there is an important additional hypothesis that there is a universal set of sounds among which each actual language chooses its set of contrasting phonemes. The International Phonetic Alphabet (IPA) represents an explicit proposal about the nature of this universal set. As stated in the *Handbook of the International Phonetic Association* (IPA 1999), "[t]he IPA is intended to be a set of symbols to represent all the possible sounds of the world's languages [. . .]. The sounds that are represented by the symbols are primarily those that serve to distinguish one word from another in a language" (IPA 1999: 159).

Everyday experience shows that, on the other hand, there is no universal phonetics. To give a trivial example, one of the authors of this chapter is a native speaker of American English and the other one is a native speaker of Spanish who learned English in adulthood. Both authors have a good understanding of what sounds the symbols of the IPA are intended to represent. Chances are that both authors' renditions of a given word in American English, say, **'fonim**, would be identified as the same sequence of phonemes, that is, as the word that is normally written *phoneme*. One of them, however, would be perceived as having been produced with a foreign accent (i.e. with non-native phonetics).

The implicit hypothesis of phonemic transcription, e.g. as reflected in Kenyon and Knott, is, then, that speakers' knowledge of the sounds of their language can be characterized as (a) knowledge of the phonemes and sequences of phonemes of their language (drawn form a larger potential set of contrastive sounds, as expressed in the IPA), and (b) knowledge of how to articulate those phonemes in the different phonological environments in which they can be found. Importantly, phonetic detail can be abstracted away from individual lexical entries. Given a UR consisting of a string of phonemes, a native speaker will know how to pronounce it in all contexts.

2 Indeterminacy in phonemic representations

Experience has shown that establishing the phonemic inventory of a language is for the most part a straightforward matter, but also that in any language there usually remain a few cases of unclear or ambiguous phonemicization (cf. for instance, Hualde 2004). Difficulties often arise in situations where the mapping between allophones and phonemes is not one-to-one (i.e. the *bi-uniqueness condition* of Harris 1944, 1951 breaks down). Some of the commonly attested types of problems for phonemicization are discussed in this section.

2.1 English flaps as "fuzzy" phonemes

Indeterminacy in phonemic analysis arises when a single surface segment can be analyzed as deriving from a sequence of two phonemes and when segmentation as one or two phonemes is unclear. There are well-known examples of this sort in English, such as the case of the rhotacized vowel that occurs in words like *bird* – is it an independent phoneme or a sequence of vowel followed by /r/? A similar question occurs for the velar nasal – is it an independent phoneme with defective distribution (banned from syllable-initial position) or the phoneme sequence /ng/? Another notorious problem for segmentation is posed by the tense, diphthongizing vowels, variously transcribed with one or two symbols by different authors. A distinct kind of problem for phonemicization concerns the treatment of schwa – should it be analyzed as an allophone of /ʌ/ (a phonetically similar vowel) that occurs in unstressed syllables? Or, in cases where there is a morphologically related word in which stress occurs on a different syllable (e.g. *'tel*[ə]*phone, te'l*[ɛ]*phony*), should the phonemicization of schwa depend on the value of the corresponding vowel in the related word? We will come back to this topic.

A different issue for phonemic analysis is that of the status of the flap [ɾ] in American English (see also CHAPTER 113: FLAPPING IN AMERICAN ENGLISH). Replacing [ɾ] with [tʰ] in *better, but again,* or *positive* does not result in a difference in lexical meaning, so by an analysis based on the test of lexical contrast, [ɾ] should be an allophone of /t/. But under Harris's (1951) criterion of the native speaker's judgment, the flap may qualify as a phoneme, since native speakers are aware that these are two different sounds (as reflected, for instance, in informal spellings such as *geddout of here, forgeddabouddit,* etc.).[1] The perceived difference may be associated with formality or personal choice (in *better, positive*), or with phrasing (in *but again*). If we consider the phoneme as a sound category, then the flap in American English appears to be an example of a "fuzzy" or quasi-phoneme that shares some but not all of the properties of more robust phonemes (see Janda 1999 and Hall 2009 for related discussion). This view treats phonemicization as being akin to other categorization phenomena (Taylor 2006), and may allow for more complexity in the relationships among linguistic sounds than that implied in any of the twentieth-century phonemic theories.

2.2 Neutralization

2.2.1 English obstruent sequences

As noted, phonemic theory invokes lexical contrast as a primary criterion for establishing the phonemic status of a sound relative to other sounds in the language. Problems for this approach arise when contrast relations between two or more sounds are not consistent throughout the language. For instance, in many languages, two or more sounds that contrast in some positions in a syllable or word fail to contrast in others. This phenomenon is known as the *neutralization* of contrast, and its resolution in phonemic analysis has led to increased abstraction in URs in several theories.

Consider the case of obstruent voicing in English. In English, coda sequences of obstruents always agree in voicing. Thus we observe obstruent voicing agreement

[1] Flapping causes neutralization of /t/ and /d/. Speakers also seem to be aware that the flap is different from /d/, although we don't have the same kind of evidence.

in words like *act* and *tasks*, while the corresponding disagreeing tautosyllabic sequences are unattested, */-kd/, */-zks/, etc. In a simple phonemic analysis (i.e. one that expresses only phonemes and allophones and a direct mapping between the two), the absence of clusters with disagreeing voicing results in a pattern of defective distribution of obstruents: only voiced obstruents occur adjacent to a tautosyllabic voiced obstruent, with a parallel restriction for voiceless obstruents. The defective distribution does not, in this simple phonemic analysis, have any implications for URs, nor is it explicitly treated in the phonemic analysis.

Prague School phonology, on the other hand, offers an explicit model of neutralization by positing an archiphoneme in the phonological representation (the UR) in contexts of neutralization. An archiphoneme is a unit that represents the common features of phonemes whose contrastive property is neutralized in specific contexts. The archiphoneme appears in only those contexts of neutralization, substituting as it were for any one of the specific phonemes it covers. In the English example under discussion here, in a sequence of obstruents in the syllable coda, archiphonemes unspecified for voicing (represented by capital letters) replace any occurrence of an obstruent phoneme after another obstruent. In English we would thus have representations such as *desks* /dɛsGZ/, *texts* /tɛkZDZ/, *adze*, *ads*, *adds* /ædZ/, etc., where the surface voice properties of the archiphoneme are predictable from the preceding context. The inclusion of archiphonemes renders URs somewhat more abstract than a simple phonemic representation, and anticipates future developments advocating abstractness of URs. But before leaving this example, notice that since the neutralization of obstruent voicing may affect consonants across morpheme boundaries in coda clusters, as in *texts*, *ads*, and *adds*, it leads to alternations in the shape of suffixes including the regular plural nominal suffix and the 3rd person verbal agreement, a topic to which we will return in §3.

2.2.2 *Japanese sibilants*

The treatment of neutralization in phonemic theory has further implications for the abstractness of URs, illustrated here in an example from Japanese. In Japanese [s] and [ʃ] appear to be in phonemic contrast in all contexts except before /i/, where only [ʃ] is found, and before /e/, where only /s/ is found, excluding recent borrowings. Thus Japanese presents another case of the defective distribution of phonemes due to the neutralization of contrast in specific contexts. In a Prague School analysis the archiphoneme /S/ would replace the two phonemes /s/ and /ʃ/ before a front vowel, where the contrast is neutralized.

There is another possible solution to phonemicization in cases of defective distribution such as the Japanese example, which does not involve archiphonemes. The solution allows the specification in UR of abstract phonemes that fail to map to surface allophones.[2] We refer to this here as the Abstract Phonemic analysis. For the Japanese case, an Abstract Phonemic analysis posits the phoneme /s/, relegating [ʃ] to the status of an allophone: /s/ maps onto the allophone [ʃ] in surface realization when it precedes phonemic /i/ and also before the glide /j/, a kind of "ghost" phoneme that serves to condition the palatal sibilant and is simultaneously absorbed into that consonant (see Table 1.1). In fact, there is a romanization of Japanese that assumes this second phonemicization, and this is essentially the representation that we also find in the native kana orthography.

[2] Goldsmith (2008) presents an insightful discussion of the historic precedent for this type of analysis in the work of Harris (1951).

Table 1.1 Phonemicization of Japanese surface allophones [s ʃ] in three phonemic analyses. Representations in parentheses are excluded from the set of possible URs

Surface allophones		*Simple phonemic*		*Prague School*		*Abstract phonemic*	
		phonemes: /s ʃ/		*phonemes:* /s ʃ/ *archiphoneme:* /S/		*phonemes:* /s/	
[sa]	[ʃa]	/sa/	/ʃa/	/sa/	/ʃa/	/sa/	/sja/
[se]	—	/se/	(*/ʃe/)	/Se/	(*/se ʃi/)	/se/	(*/sje/)
—	[ʃi]	(*/si/)	/ʃi/	/Si/	(*/si ʃi/)	/si/	(*/sji/)
[so]	[ʃo]	/so/	/ʃo/	/so/	/ʃo/	/so/	/sjo/
[sɯ]	[ʃɯ]	/sɯ/	/ʃɯ/	/sɯ/	/ʃɯ/	/sɯ/	/sjɯ/

2.3 *Basque palatal sonorants and the question of the "free ride"*

A similar situation arises in Basque. In some Basque dialects /l/ and /n/ historically became [ʎ] and [ɲ], respectively, when preceded by /i/, syllabic or non-syllabic, and followed by another vowel; e.g. [mutila] > [mutiʎa] 'the boy', [mina] > [miɲa] 'the pain'. When the trigger was a glide, it was absorbed: [sajna] > [saɲa] 'the vein'. Since, in the relevant varieties, /l/ and /n/ were not palatalized in the coda, this has resulted in numerous alternations in morpheme-final position: [mutil] 'boy', [mutiʎa] 'the boy'; [min] 'pain', [miɲa] 'the pain'; [sajl] 'difficult', [saʎa] 'the difficult one'; [sajn] 'vein', [saɲa] 'the vein' (in other dialects we find palatalization also in the coda). In a phonemic analysis with ordered rules, this mapping between phonemic and allophonic representation could be handled by the following ordered rules (glides are allophones of the high vowels and another rule would account for their distribution; see CHAPTER 15: GLIDES):[3]

(1) *Basque palatalization*

Palatalization: /l n/ → [ʎ ɲ] in contexts following /i j/ and preceding a vowel
Glide absorption: /j/ deletes in contexts preceding (intermediate) [ʎ ɲ]
e.g. /mina/ → [miɲa]
/saila/ → sajla → sajʎa → [saʎa]

Once we have these rules, we may let them apply also in the morpheme-internal context, where palatals do not participate in any alternations. Thus, [iʎe] 'hair', [oʎo] 'chicken', [iɲor] 'anybody', and [baɲatu] 'bathe' can be analyzed as /ile/, /oilo/, /inor/, and /bainatu/. The rules in (1) will successfully derive palatal sonorants from all positions except perhaps word-initially (where the context for the added rule of glide formation would not obtain); since word-initial palatal sonorants are found only in a very small number of words (mostly borrowings),

[3] Phonemic analyses with ordered rules mapping phonemes to surface allophonic representations are found in Bloomfield (1939), and, as highlighted in Goldsmith's recent work (2008), are again taken up by Wells (1949) in work that presages the major development in Generative Phonology a decade later.

this dynamic phonemic analysis may allow us to dispense with two phonemes, /ʎ ɲ/, from the underlying phoneme inventory for the language. The question for phonemic theory is whether this analysis should be allowed, where morpheme-internal palatal sonorants get a "free ride" on the analysis motivated for cross-morpheme contexts (Oñederra 1991). In this particular case, we have some evidence in favor of the abstract analysis that allows "free ride" derivations, in the form of some subsequent developments. In a couple of regional dialects palatal sonorants have undergone depalatalization, and this has affected both morpheme-final and morpheme-internal palatals. Indeed, palatals which did not have their historical origin in the palatalization process have also been depalatalized, generating a preceding glide when not following /i/: e.g. *teila* 'tile' < Romance *teʎa; *ladrilu* 'brick' < Spanish *ladrillo*; *dainu* 'damage' < Spanish *daño* (Zuazo 2010: 61-62). Although the explanation for this second sound change may be found in a hypercorrection process, it is consistent with the abstract URs of the "free ride" analysis.

2.4 *Summary*

The examples discussed above illustrate the challenge in determining the correct UR for a given word or phrase in a phonemic analysis. While there has been widespread support for the notion that the basic elements of phonology are units, such as phonemes, that are abstractions over detailed phonetic forms, there are still many questions remaining about the degree of abstraction that is appropriate in UR. A frequent problem arises when two sounds that contrast in some contexts do not contrast in other contexts, as in the Japanese example that we have considered. Further issues arise from the possibility of reducing the size of the phoneme inventory at the expense of greater abstractness in underlying phonemic representation, as in the case of Basque palatal sonorants, or the several problematic cases mentioned from English. Yet another challenge arises from cases where different criteria for phonemicization result in conflicting phonemicization, as in the case of the English flap as a fuzzy phoneme. Yet other challenges arise when the contrast between lexical items involves overlapping segments, and cannot be reduced to an analysis in terms of one-to-one correspondence between phones and phonemes (for further discussion see Lass 1984: ch. 2).

3 Underlying representations in morphophonemic theories

A different approach to phonemic analysis in cases of neutralization can be found in the work of American structuralist phonologists who tackle the problem of determining the underlying segments in cases of neutralization by taking into consideration the phonological form of inflectionally or derivationally related words. A classic demonstration of this approach is in the analysis of the underlying voicing of word-final plosives in German, based on their realization in inflected forms of the same paradigm. For instance, the final voiceless consonant of *Bund* [bʊnt] 'association' may be analyzed as the realization of an underlying voiced phoneme /d/, because the genitive *Bundes* [bʊndəs] appears with the voiced phoneme, as does the plural *Bunde* [bʊndə] (see CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). Similarly, American English *atom* [ærəm] may be represented as /ætam/ because the underlying nature of the neutralized segments is revealed in *atomic* [ətʰamɪk]. This view, which makes use of morphophonological considerations to determine underlying forms, was already present in the work of Baudouin de Courtenay (see Anderson 1985: 67–68), but is explicitly rejected by Jones (1967: 104–107) and other authors who maintain that phonemic representations of words should be established using purely phonological information.

In American structuralist approaches, a phonemic representation is based on observations of the distribution of sounds in phonetic form, and is distinguished from a separate morphophonological representation, where relations between words containing the same morpheme are considered. Thus, German *Bund* [bʊnt] would have the phonemic representation /bʊnt/ and the morphophonemic representation //bʊnd//. The admittance of a morphophonological level of representation raises the question of whether this representation should be considered as the underlying representation of words, and accorded status as psychologically real. A phonemic theory with no morphophonological level must resort to an explicit listing of the allomorphs as multiple URs for alternating morphemes, while in a theory with distinct levels of morphophonological and phonemic representations, allomorphs can be defined by the mapping between the two. The morphophonemic analysis is illustrated here with the English regular plural suffix. This morpheme can be said to possess three allomorphs in complementary distribution: /-z/, /-s/, and /-əz/. (In a Prague School analysis, it would have the allomorph /-Z/ after an obstruent, where there is no contrast between /s/ and /z/, as in *cats, dogs*, and the allomorph /-z/ after a sonorant, as in *boys, hens*, where it contrasts with /s/, cf. *voice, hence*). Since the distribution of the allomorphs is phonologically conditioned (see CHAPTER 99: PHONOLOGICALLY CONDITIONED ALLOMORPH SELECTION), and furthermore, essentially the same alternation is found with other suffixes such as the genitive and the regular past tense, one possibility is to choose a single underlying morphophonemic representation for each suffix, from which (phonemic and) surface forms could be derived by the application of general rules. The morphophonemic analysis is summarized in (2), in contrast to a phonemic analysis with a listing of allomorphs. Note that the analyses shown here are offered as concrete examples of the phonemic and morphophonemic approaches, and exist alongside other possible analyses of the specification of phonemic or morphophonemic form.

(2) *The English plural suffix in "simple" phonemic, Prague School phonemic, and morphophonemic analyses*

	"simple" phonemic	Prague School phonemic	morpho-phonemic
morphophonemic level:	—	—	//-z// (=UR) *hens, cats, dogs, kisses*
phonemic level:	/-z/ *hens, dogs* /-s/ *cats* /-az/ *kisses* (=UR)	/-z/ *hens* /-Z/ *cats, dogs* /-az/ *kisses* (=UR)	/-z/ *hens, dogs* /-s/ *cats* /-əz/ *kisses*

Relevant to our focus here on URs, the critical distinction between the phonemic and morphophonemic analyses illustrated in (2) is whether there is a unique representation specifying the phonological form of all surface realizations of the morpheme (the morphophonemic analysis), or whether each allomorph has an independent phonological representation (the phonemic analyses). The morphophonemic solution is also adopted in Generative Phonology, the theory that supplanted structuralism as the dominant school of American phonology, but with the important difference that the Generative Phonology model of grammar bypasses the "classical" phonemic level.

4 Underlying representations in Generative Phonology

In modern practice the term "underlying representation" (UR) has become associated with the underlying phonological representations of Chomsky and Halle's Generative Phonology, the major development in phonological theory following Bloomfield and his successors in American structuralism. As Chomsky and Halle (1968: 11) explain, their phonological representations are essentially equivalent to the morphophonemic representations of American structuralist phonology. They further make clear that they, however, prefer not to use the term *morphophonemic representation*, because this term seems to imply the existence of a different, *phonemic* level, which they do not believe to be necessary or useful as a level or representation.[4]

Chomsky and Halle's adoption of the morphophonemic level as input for the operation of phonological rules is mostly justified in terms of Chomsky's overall conception of grammar, where the phonology operates on the output of syntactic structures. Since the morphemes that compose a word may appear under different syntactic nodes, morphemes, not words, must be the units of lexical encoding. To use their example, the syntax provides sequences such as $[[sing]_V\ past]_V$ and $[[mend]\ past]_V$, which, after the operation of readjustment rules, become, respectively, the underlying phonological representations $[s^*ng]_V$ and $[[mend]_V\ d]_V$ (where * represents the addition to the feature specification of *i* of a new feature "indicating that it is subject to a later phonological rule which, among other things, happens to convert *i* to æ"; 1968: 11).

In Chomsky and Halle's framework the units in URs contain segments which are further decomposed into phonological distinctive features, including morphological and syntactic juncture features, and in some instances, such as the examples discussed above, specific diacritic features. URs are mapped onto surface forms through the application of phonological (transformational) rules. These rules apply in a linear order, and the output of a rule yields an *intermediate* form that is the input for subsequent rules, until the final ordered rule applies to yield the surface form.

[4] In denying the status of a distinct level of phonemic representation, Chomsky and Halle were essentially in agreement with Bloomfield (1933), as noted by Koerner (2003). Chomsky and Halle's rejection of structuralism, and phonemic analysis in particular, is directed at the taxonomic phonemic analysis of Twaddell, Bloch, and other post-Bloomfield structuralists (Odden 2005: ch. 3, supplement).

4.1 The criterion of maximizing grammatical generalization

The URs of Generative Phonology, like the phonemic representations of structuralist theories, abstract away from the detail of phonetic form. There is no explicit limit on the degree to which the UR diverges from the phonetic form, and the UR of a given morpheme is not constrained to be identical or even similar to the surface form of any of its allomorphs. For example, Kenstowicz and Kisseberth (1979: 204) propose an analysis of Russian vowel alternations in which the noun "head" is assigned the UR /golov/ 'head', with two full vowels. These vowels never occur simultaneously in the surface form of any word containing this root morpheme, but each occurs in stressed position in different words: ['goləvu] (ACC SG) and [gʌ'lof] (GEN PL).[5] The full vowels in the UR surface intact only in the presence of stress, which is assigned by morphophonological rules, and are otherwise transformed by rule into the reduced vowels [ə ʌ] in unstressed syllables.

URs specify lexically contrastive features (see CHAPTER 17: DISTINCTIVE FEATURES), and leave out any feature that is predictable from the phonological content (including juncture features), but the criterion of contrast is not the sole basis for determining URs in Generative Phonology. Another important criterion is maximizing grammatical generalization. The UR is the form that provides an optimal mapping to *all* the observed surface forms of the morpheme, maximizing the function of phonological rules in specifying predictable information, and in expressing regularities in the distribution of sounds in the language overall.

For example, consider the representation of nasal consonants in a language like Catalan (Herrick 2002; see also Bonet and Lloret 1998: 127–155; Wheeler 2005: 166–219). In certain phrasal contexts, the alveolar nasal /n/ assimilates in place of articulation to a following consonant, as in (3a). The rule of Nasal Assimilation (4), formulated using the notation of Chomsky and Halle (1968), operates on word-final /n/ to change the place of articulation feature in the appropriate contexts. There is a similar pattern of homorganicity in NC clusters that can be observed within words, shown in (3b), i.e. [n] is never found in heterorganic clusters morpheme-internally. These word-internal clusters do not participate in any morpho-phonological alternations involving nasal place of articulation, but allowing the rule of Nasal Assimilation to apply word-internally to /nC/ sequences offers the maximal generalization, permitting URs like /kanp-ɛt/ 'little field' to be mapped onto surface representations with homorganic NC clusters like [kamp-ɛt]. Under this account, the underlying structure /. . . np . . ./ may be posited even in the absence of any direct evidence for that structure from alternations in surface form, e.g. when the rule system operates to transform the underlying structure in every surface instance.

[5] This also applies to the vowels in the English example *atom, atomic*, mentioned in §3.

(3) *Catalan nasal assimilation with /n/*

a.	*son*	[n]	'they are'
	son amics	[n]	'they are friends'
	son pocs	[mp]	'they are few'
	son feliços	[ɱf]	'they are happy'
	son dos	[n̪d̪]	'there are two'
	son rics	[ɳr]	'they are rich'
	son germans	[nʲʒ]	'they are brothers'
	son lliures	[nʲʎ]	'they are free'
	son grans	[ŋg]	'they are big'
b.	*campet*	[mp]	'field (DIM)'
	tombet	[mb]	'walk, stroll (DIM)'
	puntet	[n̪t̪]	'point (DIM)'
	banquet	[ŋk]	'bank (DIM)'

(4) *Catalan nasal place assimilation*

$$\begin{bmatrix} \text{+nasal} \\ \text{+coronal} \\ \text{+anterior} \end{bmatrix} \rightarrow [\alpha\text{place}] \ / \ \_\_ \ \# \ [-\text{syllabic}, \alpha\text{place}]$$

4.2 *Underspecification in underlying representation*

An alternative analysis of the Catalan data that avoids positing /n/ as the UR in monomorphemic NC clusters is to allow the nasal consonant to be underspecified for place features in UR. Underspecification in UR was proposed by Kiparsky in an unpublished (1981) manuscript on vowel harmony, and further developed in Kiparsky (1982), Archangeli (1984), Steriade (1987), and Pulleyblank (1988), among others (see Steriade 1995 and CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION for an overview). The proposal is an elaboration of a basic tenet of Generative Phonology as put forth by Chomsky and Halle (1968), namely that URs are devoid of all predictable phonological information (which as noted above is also a core principle of phonemic representation in most phonemic theories). For Chomsky and Halle, segments are specified as bundles of distinctive features, and thus any non-contrastive feature, such as aspiration on voiceless plosives in American English, is omitted from UR. Taking this idea one step further, features that do not function to distinguish contrastive sounds may also be omitted from those contexts in UR, in what is termed *Contrastive Underspecification* (Steriade 1987). Notice that the solutions adopted for underlying representations in generative analysis with underspecification can be very similar or identical to Prague School representations incorporating archiphonemes. This will be the case when features are left underspecified only in contexts of neutralization and the features that are left unspecified are those that in other contexts serve to distinguish two or more segments, as in Catalan *campet* /kaNpɛt/ 'little field', *puntet* /puNtɛt/ 'little point'. A more extreme version of underspecification theory, termed *Radical Underspecification*, holds that for every binary distinctive feature, only one value (the *marked* value) is specified in UR, while the opposite value (the *unmarked* value) is filled in during the course of derivation by either context-sensitive or default phonological rules (Kiparsky 1985; Archangeli 1988).

Applied to Catalan, the principle of contrastive specification in UR means that the place of articulation feature will *not* be specified for nasals in NC clusters, where it is predictable from the following C even though in other contexts, where place features cannot be predicted, they are obligatorily included in UR. This analysis would be identical to a Prague School analysis. In Radical Underspecification analysis, on the other hand, one of the nasals may be left unspecified for place even in contexts where place distinctions are not neutralized, such as word-finally before pause or a vowel. Thus, *són* 'they are' would be represented as /soN/ in Radical Underspecification models even though in this context there is a contrast with the bilabial nasal of *som* /som/ 'we are'. An advantage of the Radical Underspecification approach, in which /n/ is systematically unspecified for place, as /N/, is that its representation accounts for why it is only /n/ that undergoes (major) place assimilation. Nasals with marked place features can occur in heterorganic clusters, e.g. *som dos* 'we are two', *a*[ɲ] *feliç* 'happy year'.

Needless to say, the adoption of underspecification of any sort renders URs more abstract. At the extreme, a segment may lack all distinctive feature content, being defined in UR with no more than a bare syllable position. For example, featureless vowels have been proposed by Choi (1995) for the analysis of Marshallese, and for the analysis of schwa (e.g. Anderson 1982; see also CHAPTER 26: SCHWA).[6]

With this development of underspecification in Generative Phonology in the 1980s, we have reached a zenith with respect to phonological theories with abstract and minimally specified URs. In §7 and §8 we return to consider subsequent developments in phonological theory, which pull URs in the opposite direction, away from abstractness and toward full specification.

4.3 *URs and novel word formation*

Some evidence in support of a theory that posits URs as a means to maximize grammatical generalization comes from observations about novel word formation (Kenstowicz and Kisseberth 1979: 26ff.). Consider for example the analysis of the English plural in terms of URs (as in (2) above). By positing the underlying form of the plural suffix as /-z/, with phonological rules mapping this UR to its surface reflexes in [-z], [-s], and [-əz], we have a ready account of the behavior of native speakers in forming novel plural words. As shown by Berko (1958), even young children show a preference for novel plurals that conform to familiar lexical patterns (e.g. the plural of *wug* is given with [-z]), which is consistent with the application of a general phonological rule to a common UR for the plural suffix.

The productivity of phonological patterns to novel words may be handled in a theory without URs by explicitly listing each allomorph in the lexicon along with its conditioning environment. Mechanisms of analogical extension can then select the correct allomorph for a novel word form based on its similarity to an existing word form. But, as discussed by Kenstowicz and Kisseberth (1979: 29ff.), the lexical listing alternative is not available for productive phonological rules that

[6] The featureless vowel lacks phonological place features, acquiring place specification only in phonetic implementation. Manner features are typically non-contrastive for vowels, and the major class features that distinguish vowels from consonants can be predicted on the basis of a minimal syllable structure that encodes the vowel as a syllable nucleus. Alternatively, syllable structure itself can be omitted from UR if the vowel is specified for the major class features [−consonantal, +syllabic].

are conditioned by phrasal context, as in the case of Chimwiini vowel length. More recently, Albright and Hayes (2003) present experimental evidence against an account of novel word formation that draws only on analogical extension of existing lexical patterns, based on data from novel word formation in English. They propose that English speakers' ratings of novel past tense forms reflect the operation of rules learned by induction over lexical patterns, but only if rules are constrained to encode phonological structural similarities between lexical items. However, we note that although Albright and Hayes argue for phonological rules as the mechanism for expressing phonological patterns over word forms, their analysis does not require abstract URs and does not give absolute priority to maximizing phonological generalization. Rather, they advocate a model of grammar that allows multiple rules governing morphophonological alternations that are formulated at varying degrees of specificity, reflecting "islands" of morphophonological regularity in the lexicon.

5 Indeterminacy in morphophonemic representations

In a framework with morphophonemic URs, including Generative Phonology in addition to some earlier American structuralist approaches, the problem of determining the most appropriate or optimal UR is even greater than in a simple phonemic theory that lacks a morphophonemic representation. Some issues that arise relate to: (a) the choice of UR when a morpheme has different allomorphs, (b) constraints regarding how abstract URs may be, and (c) determining which words are related.

5.1 *Indeterminacy in UR selection*

When we have distinct allomorphs of a morpheme, the choice of UR is sometimes less than obvious. Even in the relatively simple case of allomorphy in the English plural and other inflectional suffixes, there is a surprising variety of possible analyses, many of which have been explicitly proposed (see Zwicky 1975; Kenstowicz and Kisseberth 1979: 181; CHAPTER 99: PHONOLOGICALLY CONDITIONED ALLOMORPH SELECTION). For instance, different generative phonologists have proposed analyses where the UR of the plural suffix in English is /-z/, /s/, or /-ɪz/. In Spanish, the plural is generally formed adding /-s/ to stems ending in a vowel, as in *casa*, *casas* 'house, houses', and adding /-es/ to stems ending in a consonant, as in *amor*, *amores* 'love, loves'. Whereas generative phonologists appear to agree in taking /-s/ as the UR of the suffix, there has been much debate on the relative merits of an epenthesis analysis, where *amores* would be derived from /amor+s/ by a rule of vowel insertion (Saltarelli 1970), and a deletion analysis, where all consonant-final stems are provided with a final vowel in their UR, which is deleted in word-final position by rule (Foley 1967; Harris 1969), so that *amor* is /amore/ and *amores* is /amore+s/. In principle, nothing would rule out a third analysis where the UR of the plural suffix is /-es/, with deletion of the suffix-initial vowel in *casas* /kasa+es/.

There are few explicitly stated principles governing the analysis of URs. Deciding on a UR can require careful phonological argumentation, taking many kinds of facts into account and, as we see, different phonologists may come up with different solutions. It remains unclear what principles of Universal Grammar guide the language learner to a unique correct analysis in indeterminate cases such as these. Note that the issue of choosing the correct UR is especially acute in Generative Phonology, where, barring suppletion, all allomorphs of a given morpheme *must* derive from a unique underlying phonological representation. A theory that does not treat morphophonemic URs as mental constructs has the option of handling alternations of this type simply by lexical listing of each alternant.

5.2 *Abstractness in underlying representations*

The problems in the selection of UR are complicated by the possibility of having indeterminately abstract URs. URs assume a certain degree of abstraction just for adopting phonemic (i.e. phonetically underspecified) representations. When we identify a phoneme /t/ in English which is realized as aspirated [tʰ], unaspirated [t], glottalized [tˀ], flap [ɾ], or glottal stop [ʔ] in different contexts or instances, we are proposing an invariant abstract phonological unit underlying quite different phonetic realizations. But the issue of abstractness and its limits are even more vivid in Generative Phonology precisely because this theory takes the strong position that the relevant units of lexical encoding are morphemes. Since morphemes may appear in quite different shapes in different words, the URs of Generative Phonology can be considerably more abstract than the phonemic representations of words. Again using Chomsky and Halle's example for this point, the underlying representation of *telegraph* must be one from which the surface phonetic representation of *telegraph, telegraphic,* and *telegraphy* can be derived. They thus choose +tele+græf+. Elsewhere in the same work, they propose URs that differ quite radically from the surface form of words. Some of the early generative work by other authors also includes very abstract representations. We will consider a couple of examples below, in connection with the issue of word-relatedness.

5.2.1 *Constraining abstractness: The Alternation Condition*

A reaction to the abstractness of URs in Generative Phonology is found in the work of Kiparsky (1968), whose Alternation Condition is nevertheless too restrictive for some scholars (e.g. Kenstowicz and Kisseberth 1979; Kenstowicz 1994; Odden 2005). Hooper's (1976) True Generalization Condition amounts to a wholesale rejection of the theory behind morphophonemic URs, since the condition essentially limits the scope of phonological rules to phonotactics. Whereas in more recent times there has been a tendency to disfavor very abstract morphophonological URs, the fact is that the issue has not been explicitly resolved so much as sidestepped in contemporary work in Generative Phonology.

Because of its historiographic importance, we will briefly review Kiparsky's (1968) proposal here. In formulating the Alternation Condition, Kiparsky's focus is on analyses within the framework of Generative Phonology that posit underlying forms that contain elements that never surface as such, but which serve to condition the application of a phonological rule whose output could not otherwise be predicted on the basis of the surface forms that actually do appear. The Alternation Condition prohibits analyses in which all phonological derivations of an underlying form (a morpheme) result in the neutralization of a contrastive element, termed an "absolute" neutralization. The offending analyses posit different underlying representations for what appears in surface form as the same segment, in order to account for differences in phonological behavior conditioned by that segment, in different words. Generally these are cases where two historically distinct phonemes have merged. We will briefly consider one of the examples treated by Kiparsky.

A synchronic statement about Sanskrit is that velars palatalize before /i/ and before some, but not all, instances of /a/. The historical explanation for this state of affairs is that palatalization took place before the front vowels /i e/, but subsequently in the diachronic development of the language, all non-high /e a o/ vowels merged in /a/ (Hock 1991: 149; see also CHAPTER 71: PALATALIZATION).

(5) *Sanskrit velar palatalization*

*gegome	>	/ʤagaːma/	'went'
*giwo	>	/ʤiːva/	'alive'
*penke	>	/paɲʧa/	'five'

A possible synchronic analysis in a Generative Phonology approach would postulate underlying /e/ as distinct from /a/ and formulate the rule as palatalization of /k g/ before front vowels. This would be followed by another rule converting all instances of /e/ into /a/: /ke/ → /ʧe/ → [ʧa]. This derivation involves absolute neutralization, since underlying /e/ never surfaces as such in the morphemes that condition palatalization. In every instance it is neutralized with /a/ after the application of the palatalization rule. The /e/ vowel is posited in the underlying representations only to make the palatalization rule appear to be regular. This exemplifies the diacritic use of phonological content that Kiparsky's Alternation Condition is intended to disallow.

As mentioned, some generative phonologists argued that Kiparsky's constraint is too restrictive. For instance, Kenstowicz (1994: 113), following Chomsky and Halle (1968), claims that the alternation between [aj] and [ɪ] in words like *divine* and *divinity* derives from a common source in "underlying long [iː]." He points out that the putative underlying vowel does not surface as such in any surface realization of the root morpheme. Rather, the underlying vowel /iː/ is either diphthongized as in *divine* or undergoes shortening as in *divinity*. Kenstowicz reasons that the merits of the phonological analysis of Vowel Shift in these examples (and extending to certain other alternations between long and short vowels) favor rejecting the Alternation Condition. For Kenstowicz, the critical criterion for judging the validity of an abstract UR is whether positing such a form results in a simpler grammar (i.e. one with fewer and less complex rules), and achieves broader generalization in characterizing the sound patterns across the lexicon. These criteria require an evaluation method for measuring complexity and generalization, which is in itself problematic, but do not require any constraints on abstractness in UR per se, or methods for measuring the degree of abstractness in UR.

5.2.2 *Abstract URs and opacity*

A sound pattern that arises due to a phonotactic constraint or through morphophonological alternation is said to be *opaque* if its conditioning environment is not present in surface form, but can be identified in a UR. Kenstowicz and Kisseberth (1979) show that certain opaque patterns can be successfully and succinctly characterized in Generative Phonological analyses that involve abstract URs set up to contain appropriate triggering conditions for the opaque sound pattern, only to have the triggering elements subsequently modified or eliminated by rules that apply later in the derivation. An example is the analysis of Palestinian Arabic word stress (Kenstowicz and Kisseberth 1979: 229–231; see also CHAPTER 124: WORD STRESS IN ARABIC), which is described by the following rule:

(6) *Stress in Palestinian Arabic*

a. Stress the final syllable if it contains a long vowel or ends in a consonant cluster: [ka'maːn] 'also', [dara'stiː] 'you (FEM) studied it', [da'rast] 'I studied'; or else:
b. Stress the penultimate if heavy: [da'rasti] 'you (FEM) studied', [ba'tˤaːtˤa] 'potato'; or else:
c. Stress the antepenult: [da'rasatu] 'he studied it'.

There are two sets of surface exceptions to the pattern defined by these rules. In one group of words stress is antepenultimate even though the penultimate is heavy. A second group of exceptions have final stress even though the last syllable does not have a long vowel and does not end in a consonant cluster:

(7) *Surface exceptions to the stress rules*

a.	['btudursi]	'you (FEM) study'
	['simismu]	'his sesame seeds'
	['zuʔurtu]	'his bees'
b.	[bji'trin]	'string (3MASC)'
	[bji'trudʒ]	'shake (3MASC)'
	[bji'tam]	'persist (3MASC)'

Kenstowicz and Kisserberth argue that all these exceptions can be explained if the stress rules take morphophonemic URs into account. The set of words in (7a) have roots whose segments appear in a different order in contexts when they are not followed by a vowel-initial suffix; e.g. ['btudrus] 'you (MASC) study'. The stress assignment in these words would be regular if stress were assigned to the URs before a systematic rule of metathesis: /b-tudrus-i/ stress assignment → /'btudrusi/ metathesis → /'btudursi/. As for the examples in (7b), other forms in the paradigms of these words show that the UR of the stem ends in a geminate, e.g. [bi'trinni] 'ring (2SG FEM)'. The surface forms in (7b) would be derived by a totally general rule that simplifies geminates at the end of a word, applying after stress assignment.

In a theory that eschewed abstract URs in favor of representations that are transparent to surface phonetic form, the facts in (7) would be treated as true exceptions to the otherwise systematic, syllable-dependent distribution of stress in Palestinian Arabic. The possibility of the systematic analysis of opaque systems, as above, makes a compelling case for allowing URs to contain elements that don't survive in surface forms. But the question arises whether the problem justifies the solution. Are opaque sound patterns sufficiently robust and productive to warrant an analysis in terms of regular grammatical rules or constraints? Or do speakers of the language treat such patterns as localized exceptions, in which case an analysis in terms of lexical exceptions to a regular pattern would be more appropriate? Productivity may be implemented as analogical patterns without the need for abstract, morphophonemic URs (see Cole and Hualde 1998 for discussion).

5.3 *Indeterminacy in word relatedness*

In order to provide consistent underlying representations at the morpheme level, a phonologist (and a language learner) should be able to determine in some principled way which words contain the same morpheme. It should be obvious, however, that, except for inflectional paradigms – and even there we may have suppletion – deciding which sets of words are related in terms of underlying phonological representations becomes very much a subjective decision of the analyst in many cases. Phonological theory has yet to offer a principled way to decide these issues.

Two examples suffice to illustrate the problem in determining morphophonological relatedness. As an example of early work in the Generative Phonology framework, Harris (1969: 169) considers that the Spanish noun *eje* /'exe/ 'axle, axis' and the adjective *axial* /ak'sial/ 'axial' are related – as they surely are from a historical point of view – and proposes an underlying form /akse/ for ['exe]. Similarly, he analyzes *leche* ['letʃe] 'milk' as /'lakte/ to capture its relationship with the adjective *láctico* 'lactic'. More than a decade later, Lightner (1983: 205), after arguing that the root of English *long* and *length* should be given a single UR in synchronic analysis, suggests that, since the adjective *dolichocephalic* 'long-headed' is surely also related to these other words, a better UR for the root morpheme may be /dl-/, followed by a suffix in /dl-nkʰ/ *long*. An exceptionless phonological rule of English would simplify the initial group /dl/. The problem for this method lies in deciding how much derivation is appropriate in a synchronic grammar – are there any practical limits that constrain the language learner in establishing a shared component of UR for a pair of words? Adding the possibility of diacritic features and abstract URs only further broadens the range of possible analysis. We are faced with many plausible or possible analyses, and few if any criteria for deciding which one is correct. Much seems to depend on which sets of words the analyst is willing to consider as containing the same morpheme. Odden (2005: 273) explicitly addresses this concern, concluding that "[t]he question of how to judge formal word-relatedness remains controversial to this day, and with it, many issues pertaining to phonological abstractness."

An independent but related problem, given claims of psychological realism, is that the theory must allow for constant updating of underlying representations as new words are learned. Chomsky and Halle (1968: 233) propose that in order to account for both the lack of vowel laxing and the presence of the affricate [tʃ] instead of [ʃ] in *righteous* (from *right*), the UR of the root should be /rixt/. That is, the UR of *right* is altered after the learner encounters the word *righteous*. Likewise, the Spanish-speaking child may need to wait until her school years, when she may learn the word *láctico*, to determine the ultimate underlying representation of the word *leche* 'milk' and may have to wait until late adulthood to learn the word *axial*, which would trigger a change in UR from /exe/ to /akse/ for the word *eje* 'axle' that she learned in childhood (see also Janda 2003: 419). In the analytic framework of Generative Phonology, the consequences of even small changes in the UR of established words could have very large ramifications for the grammar as a whole, with ripple effects possibly extending throughout the rule system. We are not aware of any work that explores this prediction, testing for effects of large-scale grammatical restructuring in late stages of language acquisition or in adulthood.

On the positive side, morphophonological URs can be a useful mechanism for capturing speakers' intuitions regarding word-relatedness. For instance, speakers may have the intuition that two words that are phonetically identical (homophones) are different if they show different behavior under suffixation. This knowledge can be represented in morphophonological URs.

5.4 *Summary*

In this section we have seen that the validity of a phonological theory that posits morphophonological forms as URs depends on a successful and constrained method for determining URs, and that such a process will necessarily involve the determination of word relatedness. Indeterminacy about the level of abstractness in URs, together with indeterminacy in establishing which words are related through a common morpheme in UR, can render the analysis opaque, which leaves us to wonder how the phonologist can arrive at the correct analysis, or beyond that, how language learners converge on a common, correct analysis of the URs of their target language. Despite serious efforts to resolve some of these issues in the years since the publication of Chomsky and Halle's seminal work (1968), notably in Kiparsky's (1968) work on constraining abstractness, and his later work on Lexical Phonology (1985), the problem of the indeterminacy of URs remains largely unresolved today.

6 Underlying representations in Optimality Theory

In Generative Phonology, as proposed by Chomsky and Halle (1968), the phonological rules that map URs to surface forms in successive steps are "input-oriented"; they apply only if the necessary conditioning environments are present in the representation that is the input to the rule (i.e. the underlying or intermediate form), and are not sensitive to properties of the output form. Optimality Theory (Prince and Smolensky 1993) is a development from Generative Phonology in which input-oriented rules are eliminated in favor of constraints on surface form. Optimality Theory maintains the morphophonological URs of Generative Phonology, but in place of a stepwise derivation that maps URs onto surface forms through the application of ordered rules, Optimality Theory invokes static constraints that evaluate surface forms for their adherence to phonotactic constraints and for the "faithful" correspondence between the UR and a candidate surface realization of that form.

A principle of Optimality Theory is the claim that URs are entirely unconstrained ("Richness of the Base"): any structure that can be defined through the legal combination of phonological elements is a potential UR in any language. Like its predecessors in Generative Phonology, Optimality Theory maintains the claim of a unique UR for each morpheme, and many analyses employ the same kinds of morphophonological URs as in rule-based Generative Phonology. Optimality Theory inherits many of the concerns discussed above related to abstractness of representations and indeterminacy in identifying morphological relatedness. To constrain the process of selecting a UR from the vast set of possible forms allowed under Richness of the Base, Optimality Theory introduces a principle of Lexicon Optimization, which aids in the specification of a UR by forcing the selection of the underlying representation that gives the most harmonic mapping between UR and surface form, which is calculated by comparing the number of constraint violations for equally ranked constraints (see McCarthy 2002: 23, 77).

The architecture of Optimality Theory, with constraints that evaluate the identity between two phonological representations, allows for the possibility of eliminating URs altogether, in a purely surface-oriented grammar, as noted by authors such as Burzio (1996). Making no reference to URs, surface phonological patterns can be modeled through constraints that evaluate the identity between distinct surface forms of words under specific morphological conditions (e.g. when two surface forms share the same morpheme, or in the presence of a reduplicating morpheme).

The emphasis on surface constraints as the source of explanation in Optimality Theory has also led to analyses with URs that are phonetically specified, and to the formulation of constraints that refer to non-contrastive phonetic detail. The "surface-oriented" approach of Optimality Theory has invited a greater focus on the phonetic factors that shape phonological systems (as illustrated by many of the papers in Hayes *et al.* 2004), a trend that extends also to Exemplar Phonology and Articulatory Phonology, to which we now turn.

7 Phonetic detail in lexical representations: Exemplar Phonology

The preceding sections document the long history of the notion that the building blocks of speech, i.e. the basic elements of phonological form, are abstractions over detailed phonetic form, but in the period of scholarship that predates Generative Phonology there was substantial disagreement between scholars about the psychological reality of abstract (phonemic or morphophonemic) representations. Thus, while Baudouin de Courtenay and Kruszewski of the Kazan School emphasized the status of phonemes as mental entities, Bloomfield asserted a behaviorist view of the independence of linguistic analysis from any psychological assumptions about the status of linguistic constructs, a position that goes back to Saussure, and was shared by Twaddell (1935), among others (see Anderson 1985 for further discussion of mentalism in the works of these and other phonologists).

The strongest claim for the psychological reality of phonological representations is made in Generative Phonology, where abstract morphophonemic representations are the basis of lexical encoding. URs, which are composed of discrete distinctive feature specifications and, as we have seen, are often highly abstract relative to phonetic form, comprise the representations of spoken language that are stored in long-term memory, and thus they are the units that serve the physical processes of speech production and perception.

This view, which went largely unchallenged for several decades after the seminal papers in Generative Phonology (including Halle 1959 and Chomsky and Halle 1968), has been revisited in recent years. A rapid expansion of research using methods from experimental and computational sciences and corpus linguistics provides converging evidence that phonetic detail is part of the information that is stored in the long-term memory of words, influencing processes of speech production and perception, and ultimately shaping patterns of sound change (Pierrehumbert 2002). Evidence that phonetic detail influences lexical representation is offered by Bybee (2000, 2001), based on her findings that the incidence of lenition or deletion of word-final /t d/ in English is related to the frequency of occurrence of individual words in everyday language use. High-frequency words are more likely to exhibit lenition or deletion than low-frequency words (see also Bell *et al.* 2003). Bybee argues that small changes in the phonetic realization of a phoneme, however they may be conditioned, are reflected in the phonetically detailed lexical representation, which may be construed as a cluster of exemplars or a specification of the distribution of continuous-valued features in phonetic space. High-frequency words are more frequently "updated"; any contextually driven lenition affecting the most frequent forms will yield an incremental process of phonetic reduction which, over time, and in the appropriate sociolinguistic context, can result in sound change. Even phonetic detail that is not related to linguistic form, such as the phonetic detail that distinguishes one speaker's voice from another, can influence the long-term memory representation of a specific word spoken by that speaker, as shown in work by Goldinger, Pisoni, and their collaborators, among others (e.g. Palmeri *et al.* 1998; Goldinger 2000).

These are only some examples from a growing variety of studies that raise questions about the traditional division between phonetics and phonology (Pierrehumbert *et al.* 2000; see also CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY and CHAPTER 90: FREQUENCY EFFECTS). The findings are at odds with the assumption that phonetic detail is removed from phonological representation, and are incompatible with theories in which phonetic detail plays no role in phonological representation or in the functioning of rules and constraints of phonological grammar.

The presence of phonetic effects on phonology can be better modeled in exemplar theory, originating in psychological theories of categorization. Whereas in other approaches to phonology, and Generative Phonology in particular, the phonetic detail that arises in speech production derives from an abstract lexical representation, in Exemplar Phonology it is the abstract elements that are formed on the basis of statistical patterning of phonetic detail as experienced by the speaker/hearer (e.g. Pierrehumbert 2001; Johnson 2007; Cole 2009). It follows then that higher-level phonological structures (features, phonemes, syllables, etc.) may differ from word to word, and from speaker to hearer. In Exemplar Phonology there is no single, discrete UR that identifies the sound representation for each word in the language; rather, the mental encoding (i.e. lexical form) consists of a patchwork representation that links together information at different levels of granularity, from abstract category-level information (e.g. specifying the syllable structure of a word) to fine detail (e.g. specifying the range of VOT values of a plosive occurring in the word). And even though exemplar models do not explicitly recognize distinct levels of representation, relationships between words that share morphemes (e.g. *telegraph, telegraphy*) can be modeled in Exemplar Phonology without recourse to an explicit, abstract morphophonological form.

The status of abstract elements in phonological representation is still very much a matter of debate in phonology, as researchers continue to investigate the evidence for the role of phonetic detail in shaping phonological systems and influencing speech behavior on one hand, and the evidence for the priority of abstract phonological structures on the other.

8 A non-segment-based theory of UR: Articulatory Phonology

Many theories of phonology refer to phonetic properties as the basis of phonemic (or lexical) contrasts between sounds. Jakobson's distinctive features (Jakobson *et al.* 1952) incorporated both acoustic and articulatory features, while subsequent work in Generative Phonology emphasized the articulatory basis of phonological features, assigning features to hierarchically grouped classes (Clements and Hume 1995). But despite the phonetic attributes associated with phonological features, they are not equated with the actual articulatory or acoustic parameters that specify phonetic form.

As one of the first among contemporary works that integrate phonetic and phonological analysis, Browman and Goldstein (1986) introduced a model of phonology in which the atoms of phonological encoding are articulatory gestures (see also CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATIONS). In their theory of Articulatory Phonology, gestures are the low-dimensional features that encode the dynamic actions of the speech articulators in the lexical form of words. Gestures represent the actions of the lips, tongue, and jaw in the formation of constrictions along the length of the vocal tract, and are coordinated in "ensembles." Segments have no direct representation in this model, but may be emergent from stable and recurring gesture ensembles.

The mapping from abstract gestures to their implementation in physical actions of the articulators is achieved with a mathematically explicit mechanical model rather than a formal symbolic grammar. Articulatory gestures differ from the segments and distinctive features of earlier theories in that they have inherent temporal and size dimensions. The phonological and phonetic content of words is represented using a set of gestures whose relative timing is coordinated in a limited number of patterns (e.g. in-phase or anti-phase) (Goldstein *et al.* 2006). These timing patterns result in sequences of gestures that may overlap in time and reduce in magnitude. Familiar phonological phenomena such as assimilation and lenition are some of the phonological effects that are modeled through patterns of gesture overlap and reduction in this approach.

Articulatory Phonology, like Exemplar Phonology, does not recognize explicit, distinct levels of phonological representation, and does not attempt to model morphophonemic alternation beyond cases that have a transparent basis in articulation, such as assimilation to an adjacent speech gesture. At the same time, Articulatory Phonology is distinguished from Exemplar Phonology in its strong claim that phonological encoding is articulatory and not acoustic, and by the characterization of phonological form as a distinct and singular representation, not a cluster of individual instances of spoken words.

9 Conclusion

A recurrent theme throughout the history of phonological theory is that in each language there is a representation of the spoken form of a word that specifies the essential contrastive elements that distinguish that word in its spoken form from all other non-homophonous words in the language. In the preceding pages we have traced the development of this notion through the European and American theories of phonology over approximately the last century, where we observe an historical progression toward representations that are increasingly abstract relative to the physically experienced spoken word. Not all theories attribute psychological reality to these abstract phonological forms, but since the introduction of Generative Phonology in the 1960s, the focus of phonological theory has been precisely on the matter of representations and grammar as components of the uniquely human cognitive system.

The trend toward increasingly abstract representations has reversed in much of the work in phonology since around 1990, and continuing to the present day. In theories as divergent as Optimality Theory, Exemplar Theory, and Articulatory Phonology, there is an increasing acceptance of the notion that phonetic detail of the sort typically relegated to a phonetic component plays a role in defining the properties of individual phonological systems, and, by extension, partly determines properties of phonological typology across languages. Contemporary theories differ in whether phonetic factors play a role in synchronic grammar, e.g. in some work in Optimality Theory, or only in diachrony as the basis for sound change, as claimed in Evolutionary Phonology (Blevins 2004; see also Hale and Reiss 2000). But both views require a theory where phonetic detail is available to phonological generalization, and a rejection of the strict separation of phonetic and phonological levels.

We observe two factors that have driven the move to abstraction in URs. First is the problem of determining the identity of the phonological units (phonemes) in contexts of neutralization, where there is not a one-to-one mapping between phonetic and phonological units. This concern marked the development of the Prague School phonemic theory with archiphonemes, and was also seen as one motivation for the distinction between morphophonemic and phonemic levels in American structuralist theory. A concern for the mapping between phonetic and phonological form is a factor in contemporary theories, and is a primary motivation for the adoption in Articulatory Phonology of gestural features, which are abstractions over the phonetic variability of different instances of the same word.

A second factor behind the adoption of abstract URs was the treatment of morphological alternations, and the perceived need to provide a common phonological representation for (non-suppletive) allomorphs of the same morpheme. To unify the phonological representation of systematically related allomorphs, structuralist theories and Generative Phonology alike rely on abstract morphophonological representations (though as noted earlier, the current focus in Generative Phonology has shifted away from questions of morphophonological representation and toward the question of the link between phonetics and phonology). The adoption of abstract morphophonological URs in Generative Phonology is necessitated by the adherence to a principle of compactness of phonological grammars. The overriding goal of phonological analysis in classical, rule-based Generative Phonology is to arrive at a set of URs and a set of grammatical rules that maximally express generalizations about phonotactics and alternations. The optimal analysis will be compact, with fewer URs and fewer rules, which are specified with minimum phonological structure, necessitating abstract URs.

Theories that lack morphophonological representations must resort to specifying a distinct phonological form for each allomorph of any given morpheme. This is the case for simple phonemic theory (without a morphophonological level), and also for some contemporary theories. For instance, in Articulatory Phonology the phonological representation specifies gestures, which are directly mapped onto articulatory actions. Any two words that comprise different gestures must have different phonological representations, including many instances of morphologically related words that contain different allomorphs of the same morpheme, e.g. *cats* and *dogs* in English, which contain different allomorphs of the plural suffix. Articulatory Phonology does not address how in the general case the phonological relationship between allomorphs should be modeled in the mind of the speaker/hearer.

A solution to the problem of how to model the phonological relatedness of morphologically related words while allowing phonetically detailed mental representations is offered in Exemplar Theory. Beckman and Pierrehumbert (2003) argue that words are related to other words through two different kinds of connections: those based on shared meaning (e.g. due to shared morphological content) and those based on shared sound structure (due to shared phonological or phonetic content). The two sets of connections don't have to converge on a common representation; phonological relations are formed over phonetic units, while morphological relations are formed over units that encode structural and semantic information related to morphemes. The mental representation of a word consists, then, of a family of interconnected forms coding different linguistic properties of the word, which Beckman and Pierrehumbert describe in terms of a connectionist network. This model falls within the family of exemplar models in that words are represented in the mind of the speaker/hearer in terms of units of phonetic experience, preserving predictable and idiosyncratic phonetic detail alike. Abstract units such as phonemes are viewed as categories formed over phonetic units (and other kinds of units), and are considered as formal syntactic objects in the overall language system.

The association between the physical experience of spoken language and its mental representation will continue to be the focus of research in phonology, as many questions remain to be answered. What is clear from the treatment of URs in phonological theory over the last century is that a complete account of phonology must model both the phonetic and the morphological relationships between words, based on evidence from a rich variety of languages, and on observations about human behavior related to spoken language.

ACKNOWLEDGMENTS

We are grateful to the anonymous reviewers and to the editors for very useful comments.

REFERENCES

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119–161.

Anderson, Stephen R. 1982. The analysis of French schwa: Or, how to get something from nothing. *Language* 58. 534–573.

Anderson, Stephen R. 1985. *Phonology in the twentieth century: Theories of rules and theories of representations.* Chicago: University of Chicago Press.

Archangeli, Diana. 1984. Underspecification in Yawelmani phonology and morphology. Ph.D. dissertation, MIT. Published 1988, New York: Garland.

Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5. 183–207.

Baudouin de Courtenay, Jan. 1871. Some general remarks on linguistics and language. Translated 1972 from Russian in Edward Stankiewicz (ed.) *A Baudouin de Courtenay anthology: The beginnings of structural linguistics*, 49–80. Bloomington: Indiana University Press.

Beckman, Mary E. & Janet B. Pierrehumbert. 2003. Interpreting "phonetic interpretation" over the lexicon. In John Local, Richard Ogden & Rosalind Temple (eds.) *Phonetic interpretation: Papers in laboratory phonology* VI, 13–38. Cambridge: Cambridge University Press.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113. 1001–1024.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns.* Cambridge: Cambridge University Press.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Bloomfield, Leonard. 1939. Menomini morphophonemics. *Travaux du Cercle Linguistique de Prague* 8. 105–115.

Bonet, Eulàlia & Maria-Rosa Lloret. 1998. *Fonologia catalana*. Barcelona: Ariel.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Burton-Roberts, Noel, Philip Carr & Gerard J. Docherty (eds.) 2000. *Phonological knowledge: Conceptual and empirical issues*. Oxford: Oxford University Press.

Burzio, Luigi. 1996. Surface constraints versus underlying representations. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, vol. 1, 123–141. Salford: ESRI.

Bybee, Joan. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In Michael Barlow & Suzanne Kemmer (eds.) *Usage-based models of language*, 65–85. Stanford: CSLI.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Choi, John D. 1995. An acoustic-phonetic underspecification account of Marshallese vowel allophony. *Journal of Phonetics* 23. 323–347.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

Cole, Jennifer. 2009. Emergent feature structures: Harmony systems in exemplar models of phonology. *Language Sciences* 31. 144–160.

Cole, Jennifer & José Ignacio Hualde. 1998. The object of lexical acquisition: A UR-free model. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 34. 447–458.

Foley, James. 1967. Spanish plural formation. *Language* 43. 486–493.

Goldinger, Stephen D. 2000. The role of perceptual episodes in lexical processing. In Anne Cutler, James M. McQueen & Rian Zondervan (eds.) *Proceedings of SWAP (Spoken Word Access Processes)*, 155–159. Nijmegen: Max Planck Institute for Psycholinguistics.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.

Goldsmith, John A. 2008. Generative phonology in the late 1940s. *Phonology* 25. 37–59.

Goldstein, Louis, Dani Byrd & Elliot L. Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael A. Arbib (ed.) *Action to language via the mirror neuron system*, 215–249. Cambridge: Cambridge University Press.

Hale, Mark & Charles Reiss. 2000. Phonology as cognition. In Burton-Roberts *et al.* (2000), 161–184.

Hall, Kathleen Currie. 2009. A probabilistic model of phonological relationships from contrast to allophony. Ph.D. dissertation, Ohio State University.

Halle, Morris. 1959. *The sound pattern of Russian: A linguistic and acoustical investigation.* The Hague: Mouton.

Harris, James W. 1969. *Spanish phonology.* Cambridge, MA: MIT Press.

Harris, Zellig S. 1944. Simultaneous components in phonology. *Language* 20. 181–205.

Harris, Zellig S. 1951. *Methods in structural linguistics.* Chicago: University of Chicago Press.

Harris, Zellig S. 1960. *Structural linguistics.* Chicago: University of Chicago Press.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology.* Cambridge: Cambridge University Press.

Herrick, Dylan. 2002. Catalan phonology: Cluster simplification and nasal place assimilation. In Caroline Wiltshire & Joaquim Camps (eds.) *Romance phonology and variation,* 69–83. Amsterdam & Philadelphia: John Benjamins.

Hock, Hans Henrich. 1991. *Principles of historical linguistics.* 2nd edn. Berlin & New York: Mouton de Gruyter.

Hooper, Joan B. 1976. *An introduction to natural generative phonology.* New York: Academic Press.

Hualde, José Ignacio. 2004. Quasi-phonemic contrasts in Spanish. *Proceedings of the West Coast Conference on Formal Linguistics* 23. 374–398.

IPA. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet.* Cambridge: Cambridge University Press.

Jakobson, Roman. 1949. On the identification of phonemic entities. *Travaux du Cercle Linguistique de Copenhague* 5. 205–213.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge, MA: MIT Press.

Janda, Richard. 1999. Accounts of phonemic split have been greatly exaggerated – but not enough. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee Bailey (eds.) *Proceedings of the 14th International Congress of Phonetic Sciences,* 329–333. Berkeley: Department of Linguistics, University of California, Berkeley.

Janda, Richard. 2003. "Phonologization" as the start of dephonetization – or, on sound change and its aftermath: Of extension, generalization, lexicalization, and morphologization. In Brian Joseph & Richard Janda (eds.) *The handbook of historical linguistics,* 401–422. Malden, MA: Blackwell.

Johnson, Keith. 2007. Decisions and mechanisms in exemplar-based phonology. In Maria Josep Solé, Patrice Speeter Beddor & Manjari Ohala (eds.) *Experimental approaches to phonology,* 25–40. Oxford: Oxford University Press.

Jones, Daniel. 1967. *The phoneme: Its nature and use.* Cambridge: Cambridge University Press. 1st edn, 1950. Heffer & Son.

Kenstowicz, Michael. 1994. *Phonology in generative grammar.* Cambridge, MA & Oxford: Blackwell.

Kenstowicz, Michael & Charles W. Kisseberth. 1979. *Generative phonology: Description and theory.* New York: Academic Press.

Kenyon, John S. & Thomas A. Knott. 1953. *A pronouncing dictionary of American English.* 4th edn. Springfield, MA: Merriam.

Kiparsky, Paul. 1968. How abstract is phonology? Distributed by Indiana University Linguistics Club. Reprinted in Paul Kiparsky (1982), *Explanation in phonology,* 119–163. Dordrecht: Foris.

Kiparsky, Paul. 1981. Vowel harmony. Unpublished ms., MIT.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm,* 3–91. Seoul: Hanshin.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Koerner, E. F. K. 2003. Remarks on the origins of morphophonemics in American structuralist linguistics. *Language and Communication* 23. 1–43.

Lass, Roger. 1984. *Phonology: An introduction to basic concepts.* Cambridge: Cambridge University Press.

Lightner, Theodore M. 1983. *Introduction to English derivational morphology.* Amsterdam: John Benjamins.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory.* Cambridge: Cambridge University Press.

Odden, David. 2005. *Introducing phonology.* Cambridge: Cambridge University Press.

Oñederra, Miren Lourdes. 1991. Free ride edo *tella* = *teila*. In Joseba A. Lakarra & Iñigo Ruiz Arzalluz (eds.) *Memoriae L. Mitxelena Magistri Sacrum*, 625–634. Donostia-San Sebastián: Diputación de Gipuzkoa.

Palmeri, Thomas J., Stephen D. Goldinger & David B. Pisoni. 1998. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition* 19. 309–328.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam & Philadelphia: John Benjamins.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology 7*, 101–139. Berlin & New York: Mouton de Gruyter.

Pierrehumbert, Janet B., Mary E. Beckman & D. Robert Ladd. 2000. Conceptual foundations of phonology as a laboratory science. In Burton-Roberts *et al.* (2000), 273–303.

Pike, Kenneth L. 1947. *Phonemics: A technique for reducing languages to writing.* Ann Arbor: University of Michigan.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Pulleyblank, Douglas. 1988. Underspecification, the feature hierarchy and Tiv vowels. *Phonology* 5. 299–326.

Saltarelli, Mario. 1970. Spanish plural formation: Apocope or epenthesis? *Language* 46. 89–96.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale.* Lausanne & Paris: Payot.

Steriade, Donca. 1987. Redundant values. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 339–362.

Steriade, Donca. 1995. Underspecification and markedness. In Goldsmith (1995), 114–174.

Taylor, John R. 2006. Where do phonemes come from? A view from the bottom. *International Journal of English Studies* 6. 19–54.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie.* Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology.* Berkeley & Los Angeles: University of California Press.

Twaddell, W. Freeman. 1935. *On defining the phoneme.* Baltimore: Waverly Press. Reprinted in Martin Joos (ed.) 1957. *Readings in linguistics*, 55–80. Washington, DC: American Council of Learned Societies.

Wells, Rulon. 1949. Automatic alternations. *Language* 25. 99–116.

Wheeler, Max W. 2005. *The phonology of Catalan.* Oxford: Oxford University Press.

Zuazo, Koldo. 2010. *Burundako Sakanako euskara: ezaugarriak.* Pamplona & Iruñea: Nafarroako Gobernua.

Zwicky, Arnold M. 1975. Settling on an underlying form: The English inflectional endings. In David Cohen & Jessica R. Wirth (eds.) *Testing linguistic hypotheses*, 598–604. Washington, DC: Hemisphere.

2 Contrast

Daniel Currie Hall

1 Twixt Ptydepe and Chorukor

Václav Havel's play *The memorandum* (Havel 1966) imagines a dysfunctional bureaucracy that suffers from, among other things, the imposition of two quite different but equally perverse artificial languages. The first of these, called Ptydepe, is designed to maximize redundancy, and thus distinctiveness. Every word of Ptydepe must differ from every other word of the same length by at least 60 percent. To mitigate the sesquipedalian consequences of this restriction, forms are allocated to meanings according to frequency; the shortest word of Ptydepe is *gh*, which means 'whatever' (Czech *cokoliv*). Ptydepe is eventually deposed as the official medium of communication, and replaced by Chorukor, which represents something like the opposite extreme. Words in Chorukor are minimally distinct, and semantically related words cluster particularly closely together; for example, the names of the days of the week differ from one another by only a single vowel. If a typographical error should cause a meeting scheduled for Friday to happen on Tuesday instead, then, the rationalization goes, that is only all the more efficient.

No natural human language is quite like either Ptydepe or Chorukor, but contrast and redundancy have been the subjects of considerable attention, and controversy, in the study of natural language phonology. In 1985, for example, Stephen Anderson's influential history of the discipline, *Phonology in the twentieth century*, suggested that the field as a whole had lately been sailing dangerously close to Chorukor in assuming that redundant features are excluded from underlying representations. This assumption was, in Anderson's view, a poorly motivated one, driven by an uncritical acceptance of Saussure's (1916: 166) assertion that "dans la langue, il n'y a que des différences," and by a misguided appeal to efficiency of encoding. Cherry *et al.* (1953), for example, had presented a theory of minimal contrastive specifications as a means of quantifying the information content of Russian phonemes; this exercise, Anderson points out, does not in itself give us any reason to believe that only unpredictable features are in fact stored in the Russian lexicon, particularly in light of subsequent revelations about "the sheer bulk and internal redundancy of mental storage of information" (Anderson 1985: 13).

The current state of the field – approximately a quarter of a century after Anderson's history – is rather different. One of the major changes in the intervening years has been the development of Optimality Theory (OT; Prince and Smolensky 1993) as a prominent and influential approach to phonology. The optimality-theoretic device of Lexicon Optimization, which maximizes the faithfulness of underlying representations to their corresponding surface forms, effectively bans underspecification in most contexts in which it is not necessary to account for surface alternations. (See Inkelas 1995, 1996, McCarthy 2005 and Hall 2007a for more detailed discussion.) The principle of Richness of the Base, meanwhile, excludes the possibility of language-specific restrictions on input forms, and thus preemptively rules out a great many of the ways in which contrastive underspecification might be put to explanatory use. These principles are important to Optimality Theory: Richness of the Base sets a rigorous standard under which the constraint hierarchy must take responsibility for static phonotactic patterns and dynamic alternations alike, and Lexicon Optimization is a key element of optimality-theoretic accounts of acquisition. A side-effect of these principles, however, is that input forms are now widely assumed to contain quite a lot of non-contrastive information and, because this is a side-effect, it has not been the subject of much direct scrutiny. The prevailing view of underlying representations has swung back toward the richly redundant style of Ptydepe, but not as the result of the kind of re-evaluation Anderson felt that contrast was due for (see also CHAPTER 1: UNDERLYING REPRESENTATIONS). The purpose of this chapter is not to provide such a thorough evaluation of the role of contrast as is still needed (an undertaking of rather greater scope, for which the reader is referred to Hall 2007b and Dresher 2009), but rather to provide a sense of some of the ways in which contrast and redundancy are relevant to current phonological theory, and vice versa. §2 provides a brief overview of phonemic and phonetic contrast, §3 takes a more detailed look at phonemic contrast and the notion of a contrastive hierarchy, §4 discusses the treatment of (primarily phonetic) contrast in Optimality Theory and §5 offers some concluding remarks.

2 Phonemic and phonetic contrast

Contrast in the sound systems of natural languages may be said to exist at two levels: the phonemic and the phonetic. At the phonemic level, more or less abstract contrasts exist among segments of the underlying inventory. These contrasts have traditionally been held to be categorical (although cf. Goldsmith 1995 and Hall 2009 for discussion of the notion of a "cline" of phonemic contrastiveness), and they are classically defined as having the potential to distinguish one word or morpheme from another. Contrasts at the phonetic level are concrete articulatory, acoustic, and auditory differences between sounds, and are clearly gradient, although the question of how best to quantify their degree is far from trivial.

In phonological theory, these two levels are connected through the use of distinctive features, which encode contrasts among phonemes by dividing them into natural classes that are typically defined in terms of phonetic properties. However, as Keating (1984: 289) argues, "while we want the phonological features to have some phonetic basis, we also want to distinguish possible [phonological] contrasts from possible [phonetic] differences." Moreover, many of the phenomena with which phonology concerns itself involve some kind of mismatch between phonemic and phonetic contrast.

Phonetic contrast without phonemic contrast exists most obviously in cases of allophony, in which a single underlying segment has phonetically distinct surface realizations. For example, Russian makes no phonemic contrast between /ʧ/ and /ʤ/, having only the former in its underlying inventory, but a surface contrast arises between [ʧ] and [ʤ] through regressive voicing assimilation, as in [ʒeʧʲ lʲi] 'should one burn?' *vs.* [ʒeʤ bɨ] 'were one to burn' (Halle 1959: 22).[1]

Not all such phonetic differences are allophonic in precisely the traditional sense. For example, Trubetzkoy (1939: 32) observes that tone is not phonemically contrastive in German, but that phonetic differences in pitch can be used "for the function of appeal" (*Appell*, on which see Bühler 1934). Into this category we might also put sociolinguistically or dialectally conditioned variants of phonemes, which could be seen as allophones from the perspective of the language as a whole, but which do not necessarily exist as such in the mental grammar of any one speaker. Phonetic contrasts also enable listeners to distinguish one speaker's voice from another's. Phonetic contrasts can thus be not only perceptible but even informative without necessarily corresponding to any phonemic contrast. In this connection, Trubetzkoy (1939: 51–52) speaks of multiple "sieves" (*Siebe*) that extract meaningful differences from the phonetic signal, of which the phonological sieve is only one. In recent exemplar-based models, on the other hand, phonological and metalinguistic information are extracted from the signal by mechanisms that are not necessarily distinct or separate; see e.g. Johnson 2006 for an exemplar-based model of the perception of gender in speech.

Phonemic contrast without phonetic contrast exists in cases of neutralization. In Classical Manchu, for example, the underlying contrast between /u/ and /ʊ/ is neutralized to [u] at the surface everywhere except after dorsal consonants (Dresher and Zhang 2005). The extreme case of phonemic contrast without phonetic contrast is, of course, absolute neutralization, in which an underlying contrast is eliminated in all surface environments. Kiparsky's (1968) Alternation Condition holds that absolute neutralization should not occur at all (or, in a weaker formulation, that absolute neutralization has a high cost in complexity), but arguments for such processes have been advanced (and disputed) with some frequency in the phonological literature.

The combination of allophony and neutralization can produce phonetic displacement of phonemic contrast, as in the opaque interaction of raising and flapping in the pair *writer* [ɹʌjɾɚ] – *rider* [ɹajɾɚ] in Canadian English (Joos 1942; Chambers 1973). Here the underlying voicing contrast between /t/ and /d/ conditions an allophonic contrast in the immediately preceding diphthong, but is itself neutralized;

[1] For the sake of consistency, I have used the symbols of the International Phonetic Alphabet (IPA) in all transcriptions in this chapter, converting from various other systems where necessary in material drawn from other sources. Here, for example, I have replaced the traditional Slavicist's symbols [č ǯ ž] used by Halle (1959) with their IPA counterparts [ʧ ʤ ʒ].

a phonemic contrast on one segment surfaces as a phonetic contrast on another. Because such patterns often involve some form of phonological opacity, they are of particular interest in investigations of the mechanisms through which opacity arises. It has been pointed out (see e.g. Kaye 1974; Gussmann 1976; Kisseberth 1976) that although such displacement makes phonological rules opaque (Kiparsky 1973), it also makes underlying representations more easily recoverable than they would be in cases of simple neutralization. This idea has been taken up by some recent work in Optimality Theory (e.g. Łubowicz 2003, 2007; Tessier 2004; Barrie 2006), in which the functional utility of recoverability plays a direct role in selecting optimal output forms; see CHAPTER 73: CHAIN SHIFTS for further discussion.

Cases in which absolute neutralization has been argued for necessarily involve some sort of displacement; in the absence of effects on surrounding material, there would be no reason to posit the existence of an underlying contrast that never appears at the surface.[2] For example, Hyman (1970) suggests that Nupe is best analyzed as having underlying /ɛ/ and /ɔ/ that are neutralized to [a] in all environments (in addition to an underlying /a/, which also surfaces as [a]). The primary motivation for positing this underlying contrast is the effect of /ɛ/ and /ɔ/ on immediately preceding consonants: /ɛ/ triggers palatalization, and /ɔ/ triggers labialization, while underlying /a/ triggers neither. The absolute neutralization analysis, which is further supported by evidence from loanword adaptation, offers a way of avoiding the inference that Nupe has phonemically contrastive series of palatalized and labialized consonants that contrast phonetically only before [a]. A small number of abstract contrasts thus obviates the need for a large number of concrete but marginal contrasts.[3]

It should be clear from this brief summary that phonemic and phonetic contrast and the relations between them have been at the root of some of the most basic – and some of the most controversial – questions in phonology. In current phonological theory, there are two major lines of inquiry that explicitly focus on contrast. Phonetic contrast features crucially in work on phonetically driven approaches to phonology, particularly as they apply to neutralization and to typological predictions about the shapes of phonological inventories. Phonemic contrast, on the other hand, is central to investigations into the specification and underspecification of features. These two areas of research form the foci of the following sections of this chapter. As in the foregoing discussion, the main phenomena considered in both of the following sections involve contrast at the segmental and featural levels of structure; however, it is worth bearing in mind that similar kinds of issues arise in suprasegmental structure as well.

[2] In Optimality Theory, Richness of the Base (Prince and Smolensky 1993: §9.3) implies that, in principle, every language must neutralize some input contrasts absolutely. However, Lexicon Optimization (in its original formulation by Prince and Smolensky 1993: §9.3 and in the refined version proposed by Inkelas 1995 as Alternant Optimization) ensures that an input segment that never surfaces faithfully and that is not recoverable from any effects it has on other aspects of the surface form will never be listed in any actual lexical entry, and thus will never be phonemically contrastive in the relevant sense.

[3] See Harms (1973) for counterproposals in which /ɛ/ and /ɔ/ are analyzed as /ɪa/ and /ua/ or /ja/ and /wa/, and Hyman (1973) for a rebuttal.

3 Phonemic contrast and the contrastive hierarchy

3.1 *Challenges to contrastive specification*

Over the course of modern phonological theory, there have been various proposals for restricting the content of lexical representations to exclude redundant information. (Some of these are outlined in this section, but see CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION and Dresher 2009 for further discussion and additional references.) One challenge that Anderson (1985: 11) identifies for such theories is the existence of "reciprocally dependent properties." If a feature value [αF] always entails the presence of some other feature value [βG], and the presence of [βG] always entails the presence of [αF], does this mean that both [αF] and [βG] are redundant, and that both should be omitted from the lexicon?

For example, consider the vowel inventory of Arapaho, which Goddard (1974: 115) transcribes as /i e o u/. Full specifications of these vowels for the binary features [±high], [±back], and [±round] are shown in (1).

(1) *Full specifications for Arapaho vowels*

	i	e	o	u
[±high]	+	−	−	+
[±back]	−	−	+	+
[±round]	−	−	+	+

In this system, any vowel that has the feature value [α back] also has the value [α round], and vice versa.[4] This could be taken to mean that neither [±back] nor [±round] is contrastive, but eliminating both of these supposedly redundant features leads to a set of representations that fails to distinguish /i/ from /u/ or /e/ from /o/:

(2) *Inadequate "contrastive" specifications for Arapaho vowels*

	i	e	o	u
[±high]	+	−	−	+
[±back]				
[±round]				

Archangeli (1988) presents a similar case in arguing against the theory of Contrastive Specification advanced by Steriade (1987) and Clements (1987). As an explicit means of identifying and removing redundant feature values, Archangeli (1988: 192) presents the algorithm in (3) (dubbed the Pairwise Algorithm by Dresher 2003, 2009).

(3) *The Pairwise Algorithm* (Archangeli 1988: 192)

a. Fully specify all segments.
b. Isolate all pairs of segments.
c. Determine which segment pairs differ by a single feature specification.

[4] Indeed, Arapaho is one of the languages that Rice (1995, 2002) cites in arguing that vowel place in general should be encoded phonologically with just two privative features: Peripheral (which encompasses both Dorsal and Labial) and Coronal.

d. Designate such feature specifications as "contrastive" on the members of that pair.
e. Once all pairs have been examined and appropriate feature specifications have been marked "contrastive," delete all unmarked feature specifications on each segment.

The Pairwise Algorithm fails to produce adequate feature specifications in cases of reciprocally dependent feature values, as in Arapaho; applied to the inventory in (1), it yields the specifications in (2). It produces similarly inadequate specifications for other types of inventories as well, such as the Maranungku vowel inventory in (4).

(4) *Full specifications for Maranungku vowels* (Archangeli 1988: 201, citing Tryon 1970)

	i	ɑ	ə	æ	ʊ
[±high]	+	−	−	−	+
[±low]	−	+	−	+	−
[±back]	−	+	+	−	+

Given the specifications in (4), the algorithm in (3) will identify the following minimal pairs of segments: /ə/ and /ʊ/ differ only in [±high]; /ɑ/ and /ə/ differ only in [±low]; /i/ and /ʊ/ differ only in [±back]; /æ/ and /ɑ/ differ only in [±back]. Accordingly, the Pairwise Algorithm will identify the feature specifications in (5) as contrastive, and eliminate all the others.

(5) *Inadequate "contrastive" specifications for Maranungku vowels* (Archangeli 1988: 202)

	i	ɑ	ə	æ	ʊ
[±high]			−		+
[±low]		+	−		
[±back]	−	+		−	+

The system of specifications in (5) includes identical representations for /i/ and /æ/; each of these phonemes is specified only as [−back] because each enters into only one minimal pair, and each is the [−back] member of the pair in which it appears. This by itself makes the representations inadequate, but it is perhaps also worth noting that the specifications for /ɑ/ and /ʊ/ are not distinct from each other, nor is the representation of /ə/ distinct from the representation shared by /i/ and /æ/, in the sense defined by Halle's (1959: 32) Distinctness Condition, quoted in (6). In other words, the feature system in (5) is essentially a ternary-valued one: /ɑ/ is different from /ʊ/ only if positive values are considered distinct from zero, while /ə/ is different from /i/ and /æ/ only if negative values are considered distinct from zero.

(6) *The Distinctness Condition* (Halle 1959: 32)

Segment-type {A} will be said to be different from segment-type {B}, if and only if at least one feature which is phonemic in both has a different value in {A} than in {B}; i.e. plus in the former and minus in the latter, or vice versa.

Furthermore, the putatively "full" specifications shown for Arapaho in (1) and for Maranungku in (4) in fact omit some potentially relevant features. If the feature value [+low] applies to the Arapaho vowels /e/ and /o/, as in (7a), then there will be no minimal pairs at all, and the Pairwise Algorithm will produce the totally unspecifiied representations in (7b).

(7) a. *Fuller feature specifications for Arapaho vowels*

	i	e	o	u
[±high]	+	−	−	+
[±low]	−	+	+	−
[±back]	−	−	+	+
[±round]	−	−	+	+

b. *"Contrastive" (total) underspecification of Arapaho vowels*

	i	e	o	u
[±high]				
[±low]				
[±back]				
[±round]				

Similarly, adding the feature [±round] to the specifications for Maranungku, as in (8a), will eliminate the two minimal pairs involving /ʊ/, leading to the "contrastive" specifications in (8b).

(8) a. *Fuller feature specifications for Maranungku vowels*

	i	ɑ	ɔ	æ	ʊ
[±high]	+	−	−	−	+
[±low]	−	+	−	+	−
[±back]	−	+	+	−	+
[±round]	−	−	−	−	+

b. *"Contrastive" specifications for Maranungku vowels, based on* (8a)

	i	ɑ	ə	æ	ʊ
[±high]					
[±low]		+	−		
[±back]		+		−	
[±round]					

The Pairwise Algorithm is clearly not a viable means of assigning phonological feature specifications. This is perhaps not surprising, in light of the fact that Archangeli (1988) formulated it specifically to serve as a straw man in her argument against Contrastive Specification. However, as Dresher (2009) demonstrates, this approach to identifying contrastive feature values has been put to serious use in influential works of phonological theory, although not always explicitly or consistently. The examples cited by Dresher include Trubetzkoy (1939) on French and German, Jakobson (1949) on Serbo-Croatian, Martinet (1964) on French, and, more recently, Nevins (2004).

3.2 *The contrastive hierarchy and its uses*

There is, however, another approach to identifying contrastive features that avoids the pitfalls of the Pairwise Algorithm. In this approach, features – and thus contrasts – are organized into a hierarchy, so that some contrasts take scope over others. Dresher *et al.* (1994: x) present a version of this approach that they term the Continuous Dichotomy Hypothesis, which is based on Jakobson and Halle's (1956) account of phonological acquisition:

(9) *The Continuous Dichotomy Hypothesis* (Dresher *et al.* 1994: x)

a. In the initial state, all sounds are assumed to be variants of a single phoneme.
b. An initial binary distinction (dichotomy) is made on the basis of one [member] of the universal set of distinctive features.
c. Keep applying the dichotomy to each remaining set until all distinctive sounds have been differentiated.

Dresher (2009) offers a more formally explicit procedure, which he terms the Successive Division Algorithm. This version is also more general, in that it does not assume that the features (and the divisions they make) will necessarily be binary:

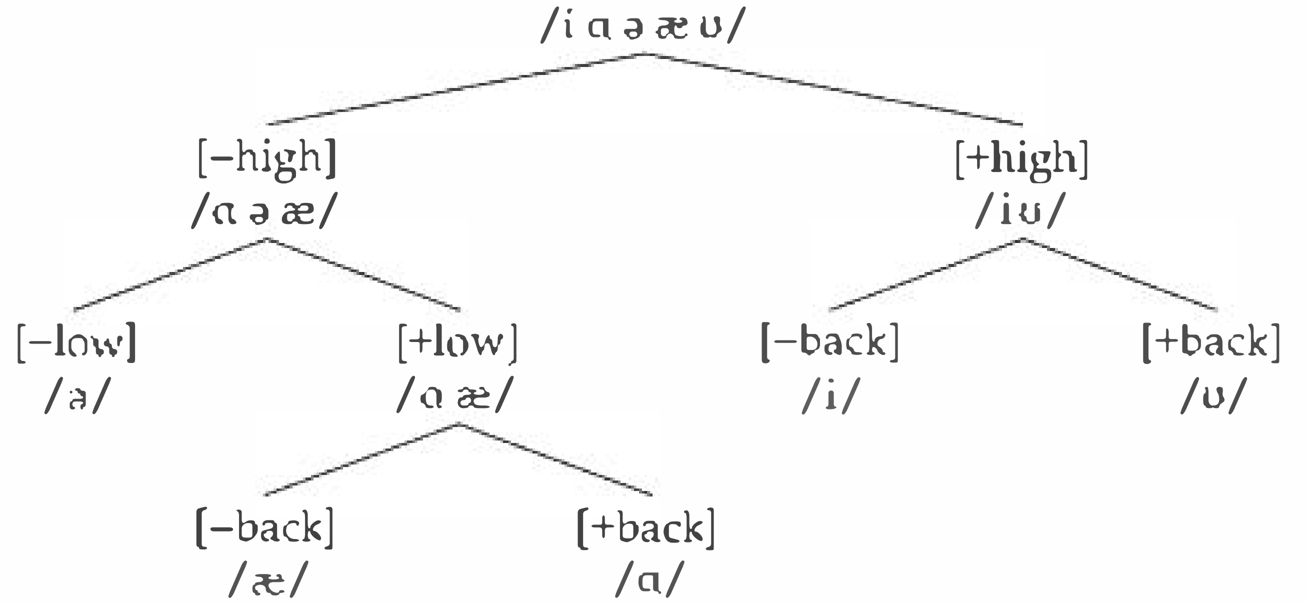
(10) *The Successive Division Algorithm* (Dresher 2009: 17)

a. In the initial state, all tokens in inventory I are assumed to be variants of a single member. Set I = S, the set of all members.
b. i. If S is found to have more than one member, proceed to (c).
   ii. Otherwise, stop. If a member, M, has not been designated contrastive with respect to a feature, G, then G is *redundant* for M.
c. Select a new *n*-ary feature, F, from the set of distinctive features. F splits members of the input set, S, into *n* sets, F_1–F_n, depending on what value of F is true of each member of S.
d. i. If all but one of F_1–F_n is empty, then loop back to (c). (That is, if all members of S have the same value of F, then F is not contrastive in this set.)
   ii. Otherwise, F is *contrastive* for all members of S.
e. For each set F_i, loop back to (b), replacing S by F_i.

Unlike the Pairwise Algorithm, the Successive Division Algorithm will necessarily produce a set of feature specifications that is sufficient to distinguish all segments in the input inventory, because it proceeds by assigning features to differentiate segments, and does not terminate until all segments are adequately specified, rather than starting with full specifications and deleting ones that appear to be redundant. By organizing features into scopal relations, the Successive Division Algorithm resolves the question raised by reciprocally dependent properties: if two feature values are mutually predictable, then the feature that takes higher scope will be assigned as a contrastive specification, and the other one will be considered redundant.

For example, suppose that the Maranungku vowel inventory in (8) is specified using the Successive Division Algorithm, with features taking scope in the following order: [±high] >> [±low] >> [±back] >> [±round]. The divisions made by the algorithm can be represented in a tree structure, as in (11a), yielding the feature specifications summarized in (11b), which suffice to distinguish the Maranungku vowels in accordance with Halle's (1959) Distinctness Condition.

(11) a. *Dividing the Maranungku vowel inventory*



b. *Feature specifications assigned by* (11a)

	i	ɑ	ə	æ	ʊ
[±high]	+	−	−	−	+
[±low]		+	−	+	
[±back]	−	+		−	+
[±round]					

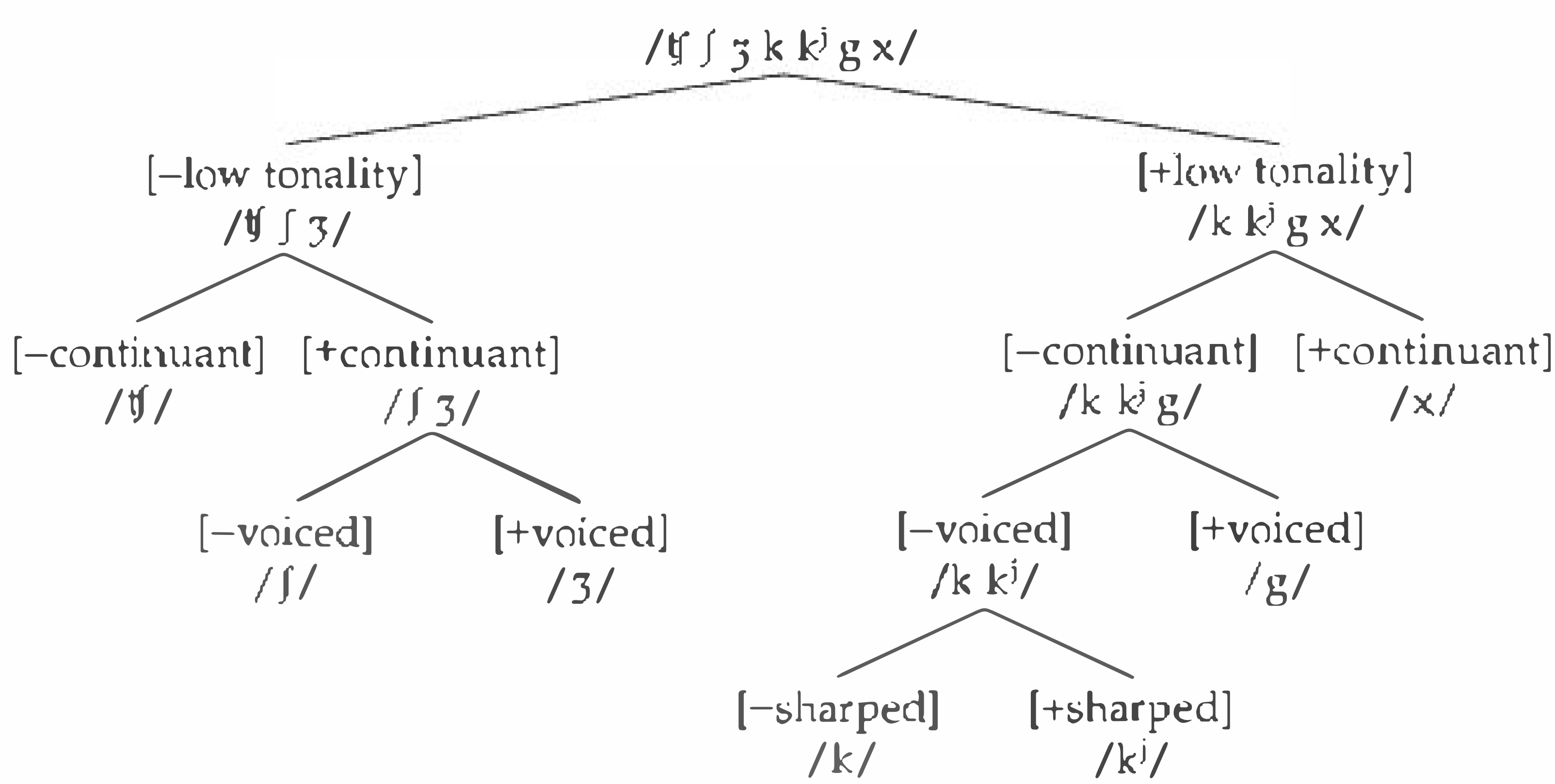
In the example in (11), all five vowels are specified for [±high], because this feature has scope over the entire inventory. The feature [±low] does not divide the [+high] vowels, so it is not contrastive for them, but it does divide the [−high] vowels, and so this feature is included in the specifications assigned to /ɑ ə æ/. Because the specification [+low] entails [−high], the value [−high] on /ɑ æ/ would be considered redundant by a post hoc approach to contrast such as the Pairwise Algorithm. Under the Successive Division Algorithm, however, the assignment of a lower-ranking feature (such as [±low] in this case) cannot erase any specifications that have already been assigned (such as [−high]), even if the higher-ranking features are logically predictable from the lower-ranking ones.

The Successive Division Algorithm requires that features be assigned in order according to a hierarchy, but it does not itself specify what orderings are possible. Unlike the Pairwise Algorithm, the Successive Division Algorithm can produce different sets of specifications for the same inventory, depending on the scope assigned to each feature. For example, the feature [±round] is not used at all in (11) because it is placed so low in the hierarchy that all the segments have been fully distinguished before [±round] is brought to bear on them. If [±round] were given scope over [±back], then [±round] would distinguish /i/ and /ʊ/, and [±back] would be redundant for those segments, although [±back] would still be contrastive elsewhere in the system, distinguishing /æ/ from /ɑ/.

Researchers working within this general approach to identifying contrastive features have proposed different ways of determining or restricting the order of features in the contrastive hierarchy. Jakobson and Halle (1956) proposed a universal partial ordering of features based on observations of the order in which children acquire the corresponding contrasts. Halle (1959), in using a contrastive hierarchy to establish underlying representations for Russian morphonemes, suggests that an ordering should be chosen in such a way as to minimize the number of specified values. More recent work, particularly in the framework referred to as Modified Contrastive Specification by Paradis and Prunet (1991), has focused on the role that specifications assigned in this manner can play in accounting for the phonological activity or inactivity of features. For example, Dyck (1995) offers evidence, based primarily on the phonological behavior of vowels in Spanish and Italian dialects, that the feature [low] consistently takes scope over [high] in Romance, and perhaps more generally as well. The papers collected in Hall (2003) offer additional examples of this approach, in which only contrastive features are predicted to be active in (at least some well-defined component of) the phonological computation.

Not all work that makes use of the contrastive hierarchy posits a correlation between phonological activity and contrastiveness. Consider, for example, the specifications assigned by Halle (1959: 45–46) to the palatal and velar obstruents of Russian:

(12) a. *Partial contrastive hierarchy for Russian* (Halle 1959: 46)



b. *Features assigned by* (12a) (Halle 1959: 45)

	ʧ	ʃ	ʒ	k	kʲ	g	x
[±low tonality]	−	−	−	+	+	+	+
[±continuant]	−	+	+	−	−	−	+
[±voiced]		−	+	−	−	+	
[±sharped]				−	+		

By giving [±continuant] scope over [±voiced], Halle (1959) ensures that the unpaired voiceless segments /ʧ/ and /x/ – as well as /ts/, in the dental series, not shown in (12) – will be unspecified for the latter feature. These segments do, however, pattern together with the contrastively voiceless obstruents of Russian with respect to regressive voicing assimilation; indeed, Halle (1959) famously used this fact to argue against the strict structuralist distinction between morphophonemic and phonemic rules, under which Russian voicing assimilation would have to be treated as two separate processes with suspiciously similar effects. In order to ensure that /ts ʧ x/ correctly trigger regressive assimilatory devoicing of other obstruents, Halle (1959) requires another rule, ordered before the assimilation rule, which fills in the predictable value [−voiced] on these segments. As Dresher (2009) points out, however, this redundancy rule would not be needed if the positions of [±continuant] and [±voiced] in the contrastive hierarchy were reversed. If [±voiced] were ranked above [±continuant] and below [±nasal], it would be possible to say that voicing is contrastive among Russian obstruents in general, and that all segments underlyingly specified for [±voiced] participate in the assimilation rule.

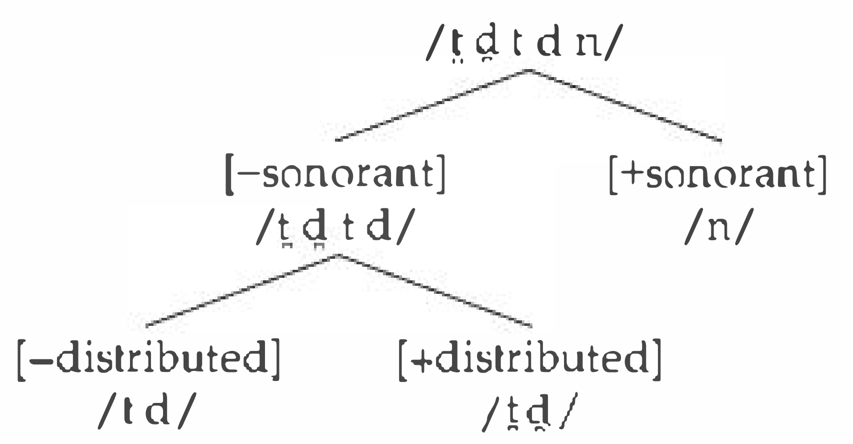
For Halle (1959), then, the contrastive hierarchy serves mainly to eliminate redundant features from the lexicon, a goal whose utility is disputed by Anderson (1985), as noted above. In the framework of Modified Contrastive Specification, however, the Successive Division Algorithm has been used to restrict the role of redundant features in phonology more generally. A particularly compelling example is offered by Mackenzie (2005, 2009) in her analysis of dental harmony in Luo and Anywa. Each of these two languages has a coronal stop inventory that includes the segments shown in (13).[5]

(13) *Dental and alveolar stops in Luo and Anywa* (Mackenzie 2009: 36)

	voiceless plosive	voiced plosive	nasal
alveolar	t	d	n
dental	t̪	d̪	

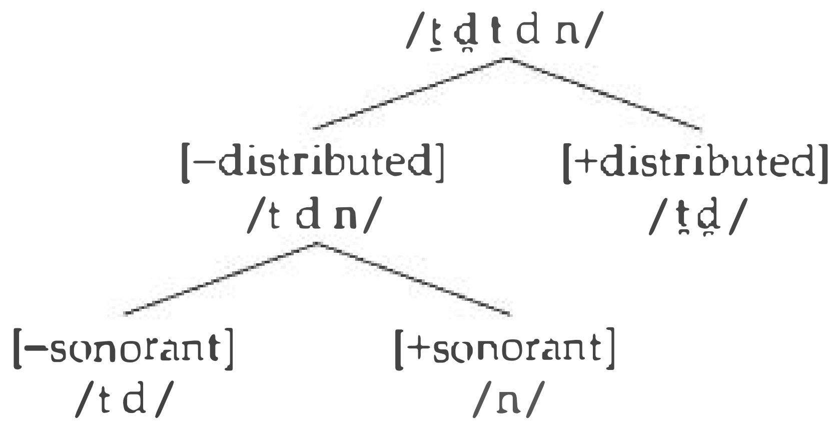
In Luo, the nasal /n/ is ignored by dental harmony, and occurs freely with dental and alveolar plosives (although these cannot co-occur with each other). In Anywa, on the other hand, /n/ is targeted by dental harmony, being realized allophonically as [n̪] in words containing dentals and as [n] elsewhere. Mackenzie argues that this difference can be attributed to different contrastive hierarchies in the two languages: in Luo, [±sonorant] takes scope over [±distributed], while in Anywa the reverse ordering applies.

(14) a. *Partial contrastive hierarchy for Luo* (Mackenzie 2009: 43)



[5] For the sake of simplicity, I set aside the Luo prenasalized stops /ⁿd ⁿd̪/, which participate in dental harmony as expected.

b. *Partial contrastive hierarchy for Anywa* (Mackenzie 2009: 42)



In both languages, dental harmony applies to segments that are specified for [±distributed]. In Anywa, this set includes /n/; in Luo it does not. Because the Successive Division Algorithm allows for cross-linguistic variation in feature scope, the shape of the inventory in (13) only partially determines the possible feature specifications. The unpaired alveolar /n/ is the only segment that can be unspecified for [±distributed], but if the feature is given wide enough scope, it will be contrastive on /n/ as well. Mackenzie's analysis accommodates the attested cross-linguistic variation while supporting the strong hypothesis that non-contrastive features are absent from the phonological computation.

4 Phonetic contrast and Optimality Theory

The rise of Optimality Theory in phonology has brought with it an interest in exploring the extent to which it is possible to characterize phonological patterns in terms of interactions between constraints that express potentially conflicting formal or functional ideals. These constraints are generally held to be surface oriented, in that the desiderata they encode have to do either with the form of the output itself (in the case of markedness constraints) or with the relation between the output and the input (in the case of faithfulness constraints).[6] Per the principle of Richness of the Base, inputs themselves are not directly constrained. A large part of the conceptual appeal of Optimality Theory is that it offers the prospect of explaining all mismatches between phonemic and phonetic contrast by means of the same markedness constraints that enforce static phonotactic generalizations.

In accordance with the generally surface-oriented character of Optimality Theory, much of the work on contrast in this framework focuses on phonetic contrast. In particular, there has been much interest in the role of phonetic distinctiveness in determining the environments in which underlying contrasts are preserved or neutralized, and in phonetic contrast as an explanation for the shapes of sound inventories.

4.1 *Phonetic contrast and neutralization*

The preservation and neutralization of underlying contrasts can be seen within Optimality Theory as arising from a basic conflict between constraints that favor

[6] Some work in Optimality Theory has proposed various additions to these two basic types of constraints; see in particular Archangeli and Suzuki (1997).

phonetically robust realization of contrasts and constraints that favor the elimination of contrasts. Input contrasts are preserved or enhanced to the extent that constraints of the former type dominate constraints of the latter type; contrasts are neutralized to the extent that constraints of the latter type predominate.

4.1.1 Formal and functional constraints

There are several different possible instantiations of each of these two types of constraint. For one thing, each type of constraint can be formulated in either formal or functional terms. A constraint that favors neutralization may be a markedness constraint in the purely formal sense of the term: e.g. the constraint *[*xxx*Place] (de Lacy 2002: 180) penalizes segments with a dorsal place of articulation (represented formally as three degrees of markedness in the Place dimension), and will have the effect of neutralizing dorsal–non-dorsal contrasts to the extent that it prevails over conflicting faithfulness constraints. On the other hand, a constraint such as *Gesture(body) (Boersma 1998: 230), which also penalizes dorsals, does so in terms of functional markedness by penalizing the articulatory effort involved in moving the tongue body to make a dorsal articulation.

Among constraints that preserve contrast, formal and functional versions also exist. On the formal side are constraints such as Ident[*xx*Place] (de Lacy 2002: 180), which says that all segments with at least two degrees of markedness in the Place dimension in the input (i.e. segments that are underlyingly labial or dorsal) must have their place features preserved in the output. As an example of an explicitly functionalist faithfulness constraint, we could take *Replace(|a|, /ɛ/, ≥30%) (Boersma 2000: 37), which relates faithful production to accurate perception by penalizing any realization of |a| that will be perceived at least 30 percent of the time as /ɛ/. While the functional utility of faithfulness in general is intuitively obvious, it is perhaps worth noting that in a framework that assumes Richness of the Base, there is no guarantee that the contrast whose preservation is mandated by any given faithfulness constraint actually carries any functional load. Rather, the connection between faithfulness constraints and the use of contrasts to mark lexical distinctions must run in the opposite direction: per Lexicon Optimization, a contrast will exist in the lexicon only if the constraints that mandate its preservation are ranked sufficiently high. (Furthermore, even the inclusion of different representations in the lexicon does not necessarily mean that the distinctions between them carry any functional load; Lexicon Optimization will ensure that all non-alternating aspects of form are stored, even if they are redundant.) Faithfulness constraints can be seen as functional in a more general information-theoretic sense, however, in that increasing the number of contrasts present in the output increases the rate at which the signal is capable of transmitting information. (See e.g. Shannon 1956 for the information-theoretic point, and Boersma 1998: 184–185 for some comments on its relevance to functionalist optimality-theoretic phonology; also chapter 63: markedness and faithfulness constraints for further discussion of the differences between markedness and faithfulness constraints.)

4.1.2 Positional licensing and positional faithfulness

Another point of difference among optimality-theoretic accounts of the preservation and neutralization of contrast is which set of constraints is held to be responsible for identifying the relevant environments. The Licensing by Cue approach presented by Steriade (1997), for example, captures the generalization that laryngeal contrasts are most likely to be neutralized in contexts in which they are least perceptible by means of the fixed ranking of markedness constraints in (15) (see also CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS):

(15) *Hierarchy of markedness constraints penalizing surface specifications for voicing* (Steriade 1997: 11, 35)

a. *[α voice] / [−son] __ [−son], [−son] __ #, # __ [−son]
b. *[α voice] / V __ [−son]
c. *[α voice] / V __ #
d. *[α voice] / V __ R[−son]
e. *[α voice] / V __ R#
f. *[α voice] / V __ R̥, V __ RV, V __ V

This is one instantiation of the notion of positional licensing, in which marked features are penalized unless they are licensed by virtue of occupying a particular position. In Steriade's (1997) approach, this positional licensing is phonetically grounded, in that the licensing environments are identified as those providing the best acoustic or auditory cues for perception of the contrast being licensed.

An alternative to positional licensing is positional faithfulness, as proposed by Beckman (1998). In this approach, the relevant environments are defined not by the markedness constraints, but by the faithfulness constraints, as in the example in (16):

(16) *Hierarchy of positional faithfulness constraints preserving underlying specifications for voicing* (Beckman 1998: 35)

a. IDENTONSET(voice)
b. IDENT(voice)

Other differences between Steriade's (1997) and Beckman's (1998) approaches are orthogonal to the choice between positional licensing and positional faithfulness, and to each other. The constraints in (16) are formulated so that (16a) applies to a subset of the forms for which (16b) is relevant; Beckman (1998: 35) hypothesizes that their ranking is fixed, so that the more specific constraint universally dominates the more general one. The constraints in (15), on the other hand, are not in a subset relation, and their ranking is fixed by alignment to the relative phonetic perceptibility of voicing in the environments to which they refer. Additionally, the environments taken to be relevant by the constraints in (15) are defined in terms of segmental adjacency, while (16) refers to syllable structure instead.

One potential drawback of the positional licensing approach is that although a hierarchy of constraints like the one in (15), with the faithfulness constraint PRESERVE(voice) (Steriade 1997: 12) ranked somewhere in its midst, can predict *where* voicing neutralization will occur, it does not necessarily determine *how* it will occur. In Lithuanian, for example, voicing on obstruents is neutralized by regressive assimilation in obstruent clusters, and by devoicing word-finally, while voicing contrasts are preserved before sonorants:

(17) *Voicing neutralization in Lithuanian* (Steriade 1997: 18)

a. *at+gal* [dg] 'back'
b. *dirb+ti* [pt] 'work (INF)'
c. *daug* [k#] 'much'
d. *silpnas* [pn] 'weak'
e. *liūdnas* [dn] 'sad'

These facts indicate that in Lithuanian, PRESERVE(voice) is ranked below *[α voice] / V __ [–son] and *[α voice] / V __ #, but above *[α voice] / V __ RV, which is consistent with the hierarchy in (15). However, the resulting grammar does not predict which instances of neutralization will result in assimilation and which in devoicing; the *[α voice] constraints indicate only that underlying voicing specifications are not to be respected. One possibility would be to say that the output of the phonological computation includes segments that are unspecified for voicing where dictated by *[α voice] constraints, and that the actual realization of these segments as voiced (as in (17a)) or voiceless (as in (17b) and (17c)) is a matter of phonetic implementation. This, however, would seem to run counter to the aim of incorporating phonetic explanations into the phonological constraint grammar.

The positional faithfulness account of similar voicing facts in Polish offered by Lombardi (1999) does not suffer from the same difficulty, because in this model the constraints that motivate unfaithfulness to underlying voicing specifications also determine the appropriate surface voicing values in each context.

Lombardi's (1999) account of final devoicing and regressive voicing assimilation in Polish is based on the constraints in (18):

(18) *Constraints relevant for voicing patterns in Polish* (Lombardi 1999: 270–272)

a. AGREE: Obstruent clusters should agree in voicing.
b. IDENTONSET(Laryngeal): Consonants in onset position should be faithful to their underlying laryngeal specifications.
c. *LAR: Segments should not have marked laryngeal features.[7]
d. IDENT(Laryngeal): Consonants should be faithful to their underlying laryngeal specifications.

Ranked in the order AGREE, IDENTONSET(Laryngeal) >> *LAR >> IDENT(Laryngeal), these constraints will generate final devoicing and regressive assimilation, and respect underlying voicing specifications of onset obstruents, as illustrated in (19):

(19) *Tableaux for Polish laryngeal neutralization* (Lombardi 1999: 282)

a. *klub* [p] 'club'

Input: /klub/	AGREE	IDONS(Lar)	*LAR	ID(Lar)
a. [klub]			*!	
☞ b. [klup]				*

[7] In the tableaux in (19), I assume that this constraint penalizes all and only voiced obstruents; Lombardi (1999: 271, fn. 2) suggests that the constraint can be taken to be more general than this, but only its ability to penalize [voice] on obstruents is relevant here.

b. *żabka* [pk] 'frog (DIM)'

Input: /ʐabka/	AGREE	IDONS(Lar)	*LAR	ID(Lar)
a. [ʐabka]	*!		**	
☞ b. [ʐapka]			*	*
c. [ʐabga]		*!	***	*
d. [ɕapka]		*!		**

c. *prosba* [zb] 'request (N)'

Input: /prosba/	AGREE	IDONS(Lar)	*LAR	ID(Lar)
a. [prosba]	*!		*	
☞ b. [prozba]			**	*
c. [prospa]		*!		*

d. *nigdy* [gd] 'never'

Input: /ɲigdi/	AGREE	IDONS(Lar)	*LAR	ID(Lar)
☞ a. [ɲigdi]			**	
b. [ɲikdi]	*!		*	*
c. [ɲikti]		*!		**

Under the positional faithfulness approach, then, faithfulness constraints determine where contrasts should be preserved, and markedness constraints are largely responsible for dictating what happens when they are not.

4.2 Phonetic contrast and inventories

Another significant way in which Optimality Theory has been brought to bear on questions of contrast is in accounting for cross-linguistic patterns in the shapes of phonological inventories. In much optimality-theoretic work, inventories are regarded as essentially epiphenomenal, having no theoretical status of their own: a surface inventory is simply the set of segments the grammar happens to permit in output forms, and an underlying inventory is simply the set of segments Lexicon Optimization leads the learner to posit in input forms. (See e.g. Kirchner 1995, 1997 on the derivability of contrastiveness in Optimality Theory.) Generalizations about inventories might therefore be expected to fall out automatically from a successful theory of the constraints that govern everything else in phonology. Dispersion Theory (Flemming 1995, 2002, among others), however, addresses inventories more directly, through the use of constraints that explicitly refer to the number or robustness of contrasts in the system as a whole, rather than only to individual forms.

4.2.1 *Dispersion Theory*

Flemming's Dispersion Theory offers an optimality-theoretic implementation of some of the insights of Lindblom's (1986, 1990) Theory of Adaptive Dispersion. Under this approach, the shapes of inventories emerge from competition among three types of constraints (Flemming 2002: 4):

- constraints that require the presence of particular numbers of surface contrasts, e.g. MAXIMIZECONTRASTS (Flemming 2002), NWORDS (Ní Chiosáin and Padgett 1997, 2001), *MERGE (Padgett 2003), CONTRAST (Bradley 2001), 𝒥-constraints (Sanders 2003);
- constraints that require surface contrasts to be phonetically robust, e.g. MINDIST (Flemming 2002), CONTRAST (Ni Chiosáin & Padgett 1997), SPACE (Ní Chiosáin and Padgett 2001; Padgett 2003), DISP (Kirchner 1998), 𝒟-constraints (Sanders 2003);
- constraints that penalize articulatorily effortful (or otherwise marked) surface forms, e.g. MINIMIZEEFFORT (Flemming 2002), LAZY (Kirchner 1997, 1998), ℳ-constraints (Sanders 2003).

Each of these three types potentially conflicts with the other two; for example, a contrast can be made more distinct either through the application of additional effort in the articulation of one or more of its members, or through the removal of one or more members from the set of contrasting forms.

If the constraints referring to contrasts are to be part of the evaluation, the grammar must consider sets of contrasting forms together; in principle, the evaluation may apply to an entire language of forms, although for expository purposes, tableaux are restricted to representative sets of forms illustrating how the grammar treats a particular type of contrast.

For example, consider the following dispersion-theoretic account of voice onset time (VOT) contrasts in English, adapted from a sketch presented by Flemming (2002: 47–48).[8] English stops exhibit a two-way contrast in VOT, but this contrast is realized differently in different positions. Foot-medially, the contrast is between voiced and voiceless unaspirated stops, as in *ogre* [ogɚ] *vs. ochre* [okɚ]. Word-initially, the voiced stops are partially devoiced, and the voiceless stops are aspirated, as in *goal* [g̊oɫ] *vs. coal* [kʰoɫ]. A dispersion-theoretic account of these facts requires the following constraints:

- MAXIMIZEVOT CONTRASTS gives preference to candidate systems with larger numbers of surface forms that are distinct in VOT. As Flemming (2002: 27) explains, MAXIMIZECONTRASTS constraints are positive scalar constraints; rather than assigning asterisks to candidates for violations, they assign check marks to candidates according to how many contrasting forms they include.

[8] I have departed from Flemming (2002) primarily in representing voiceless unaspirated [k] and partially devoiced [g̊] as (very slightly) phonetically distinct from each other, largely for the sake of providing an example in which the MINDIST constraints make a crucial contribution to the outcome. Flemming's (2002) constraint hierarchy also appears to predict that English should have a three-way VOT contrast, which is not the case in this adaptation.

- MinDist=VOT constraints require forms that contrast in VOT to do so by at least a certain degree. For this particular example, we can consider the range of potentially contrasting segments in (20), and say that the crucial MinDist constraint here, MinDist=VOT:2, requires contrasting forms to differ at least as much as do [k^h] and [g̊], or [k] and [g]. A more stringent constraint, MinDist=VOT:3, which would require forms to differ as much as [g] and [k^h], is ranked too low to have an effect.

(20) *Stops potentially contrasting in VOT*

g	g̊ k	k^h
negative VOT	VOT *circa* 0	positive VOT

- Markedness constraints penalize effortful articulations:
  *InitialVoicedStop penalizes prevoicing of word-initial plosives.
  *Aspiration penalizes aspirated stops.

Ranking MinDist=VOT:2 over MaximizeVOT Contrasts at the top of the hierarchy will mean that in both medial and initial positions the optimal system will have two contrasting forms. A system with more than two forms cannot satisfy MinDist=VOT:2, and a system with only one form will be ruled out by MaximizeVOT Contrasts.

As the tableau in (21) illustrates, the forms chosen in medial position will have [g] and [k]. Any other pair of forms that fares at least as well on the MinDist constraints will be ruled out by the constraint against aspirates. A pair with [g̊] and [k] is harmonically bounded by the winning candidate because it violates MinDist=VOT:2.

(21) *English VOT contrasts in medial position*

	MinDist =VOT:2	Max Contrasts	*IVS	*Asp	MinDist =VOT:3
☞ a. [ogɚ] [okɚ]		✓✓			*
b. [ogɚ] [ok^{h}ɚ]		✓✓		*!	
c. [og̊ɚ] [ok^{h}ɚ]		✓✓		*!	*
d. [ogɚ]		✓!			
e. [og̊ɚ] [okɚ]	*!	✓✓			*
f. [ogɚ] [og̊ɚ] [okɚ]	*!*	✓✓✓			***
g. [ogɚ] [og̊ɚ] [okɚ] [ok^{h}ɚ]	*!**	✓✓✓✓		*	*****

The corresponding tableau for word-initial position is shown in (22). Here, the winning candidate has forms with [g̊] and [kʰ]. Any other set of forms that is at least equally dispersed is eliminated by *INITIALVOICEDSTOP; any set of forms that is less dispersed is eliminated by MINDIST=VOT:2.

(22) *English VOT contrasts in initial position*

	MINDIST =VOT:2	MAX CONTRASTS	*IVS	*ASP	MINDIST =VOT:3
☞ a. [g̊oɫ] [kʰoɫ]		✓✓		*	*
b. [goɫ] [kol]		✓✓	*!		*
c. [goɫ] [kʰol]		✓✓	*!	*	
d. [koɫ]		✓!			
e. [g̊oɫ] [kol]	*!	✓✓			*
f. [koɫ] [kʰol]	*!	✓✓		*	*
g. [g̊oɫ] [koɫ] [kʰol]	*!*	✓✓✓		*	***
h. [goɫ] [g̊oɫ] [kol] [kʰoɫ]	*!**	✓✓✓✓	*	*	*****

As this example illustrates, Dispersion Theory deals not only with the overall shapes of inventories, but also with the ways in which syntagmatic patterns interact with the paradigmatic system of contrasts, and it does so in a way that is very much independent of underlying representations. In this treatment, it makes no difference whether the English VOT contrast is represented in the input as a voicing contrast (/g/ *vs.* /k/), or as an aspiration contrast (/k/ *vs.* /kʰ), or both, in part because faithfulness constraints do not play a crucial role in determining the surface forms. Because Dispersion Theory treats the system of contrasts as a whole, a markedness constraint that directly affects one form may indirectly affect another; for example, when a markedness constraint requires devoicing of the lenis member of the pair, the fortis member must be realized as aspirated in order to satisfy MINDIST=VOT:2.

4.2.2 *Dispersion and diachrony*

Padgett (2003) uses Dispersion Theory to account for a diachronic change that involves a system of contrasts, namely post-velar fronting of [ɨ] to [i] in Russian. This change, Padgett argues, was driven by constraints mandating robust phonetic contrast: the change from [kɨ] to [kʲi] resulted in a greater contrast with [ku], taking advantage of a gap that had been left by the First Velar Palatalization, in which [kʲi] had become [ʧʲi]. The tableau in (23) shows the essential components of Padgett's account of post-velar fronting.

(23) *Post-velar fronting in Russian* (Padgett 2003: 74)[9]

/pʲi/$_1$ /pɨ/$_2$ /pu/$_3$ /kɨ/$_5$ /ku/$_6$ /tʃʲi/$_4$	*MERGE	SPACE$_{COLOR}$≥½	IDENT(color)
☞ a. [pʲi]$_1$ [pɨ]$_2$ [pu]$_3$ [kʲi]$_5$ [ku]$_6$ [tʃʲi]$_4$		**	*
b. [pʲi]$_1$ [pɨ]$_2$ [pu]$_3$ [kɨ]$_5$ [ku]$_6$ [tʃʲi]$_4$		***!	
c. [pʲi]$_1$ [pɨ]$_2$ [pu]$_3$ [kʲi]$_5$ [kɨ]$_6$ [tʃʲi]$_4$		***!	**
d. [pʲi]$_{1,2}$ [pu]$_3$ [kʲi]$_5$ [ku]$_6$ [tʃʲi]$_4$	*!		**

Maintenance of contrasts here is driven by the faithfulness constraint *MERGE, which requires that separate forms in the input have distinct realizations in the output; input–output correspondence is indicated by subscript numerals. This constraint is violated by the last candidate, in which the input forms /pʲi/ and /pi/ are collapsed into [pʲi]. This is also the only candidate that fully satisfies SPACE$_{COLOR}$≥½, which requires that minimal pairs distinguished only by vowel color (backness and/or rounding) must each have at least half of the range of possible vowel colors available to them. This constraint thus imposes restrictions on both the number and the phonetic realization of contrasting vowels. If n vowels are evenly distributed in the available space, each vowel has $1/n$ of the space, as schematized in (24a). A system with more than two vowels cannot satisfy SPACE$_{COLOR}$≥½, but neither will a two-vowel system in which the vowels are unevenly spaced, such as the system in (24b).

(24) a. *Equal allocation of vowel color space* (Padgett 2003: 55)

i	ɨ̟	ɨ̠	u	Each vowel has ¼ of the available space.
i	ɨ	u		Each vowel has ⅓ of the available space.
i	u			Each vowel has ½ of the available space.
ɨ				Each vowel has all of the available space.

b. *Unequal allocation of vowel color space*

ɨ	u	/u/ has less than ½ of the available space.

[9] The tableau in (23) differs slightly from Padgett's (2003) original, in that I have omitted the constraint *au, which plays no role in this evaluation (although it was important at an earlier diachronic stage), and included the palatalization of [p] before [i].

The winning candidate in (23) improves upon the perfectly faithful candidate by realizing /ki/ as [kʲi], eliminating the violation of SPACE$_{COLOR}$≥½ in the [kV] series at the expense of the low-ranking faithfulness constraint IDENT(color). The two violations of SPACE$_{COLOR}$≥½ incurred by the [pV] series, however, cannot be eliminated without violating *MERGE, so the labial series is realized faithfully.

It is crucial to Padgett's account that SPACE$_{COLOR}$≥½ evaluates words, rather than segments; in order for the realization of /kɨ/ as [kʲi] to be useful, the [kV] series must be evaluated independently from the [pV] series. Hence Padgett's (2003: 56) definition of the SPACE constraint family explicitly refers to (potential) minimal pairs. This definition, however, allows for a rather counterintuitive repair strategy for forms with phonetically weak underlying contrasts. If some other faithfulness constraint (e.g. DEP) is ranked sufficiently low, then the forms could be made to satisfy SPACE vacuously by violating that constraint in such a way as to ensure that they are no longer minimal pairs (e.g. by inserting additional segments elsewhere in the form). The tableau in (25) suggests that if DEP were ranked below IDENT(color) in some language that otherwise resembles Russian at the stage at which post-velar fronting applied, a candidate in which several forms undergo epenthesis will be preferred over the winning candidate from (23):

(25) *Enhancing contrast through epenthesis*

/pʲi/$_1$ /pi/$_2$ /pu/$_3$ /kɨ/$_5$ /ku/$_6$ /tʃʲi/$_4$	*MERGE	SPACE$_{COLOR}$≥½	IDENT(color)	DEP
a. [pʲi]$_1$ [pi]$_2$ [pu]$_3$ [kʲi]$_5$ [ku]$_6$ [tʃʲi]$_4$		*!*	*	
☞ b. [spʲi]$_1$ [pɨ]$_2$ [plu]$_3$ [kɨn]$_5$ [ku]$_6$ [tʃʲi]$_4$				***

The obvious objection to this tableau is that, by restricting the input to these six forms, it conceals some violations of *MERGE. If [spʲi], [ki] and [plu] are phonotactically licit "words," then they should also be part of the input language, and the apparently successful candidate in (25) will be penalized for conflating them with /pʲi/, /kɨ/ and /pu/. However, if DEP is ranked sufficiently low, then the ideal way of avoiding these violations of *MERGE may be to epenthesize additional segments in the output correspondents of /spʲi/, /kin/ and /plu/ as well. Of course this, in turn, may collapse still other underlying distinctions, and so the truly optimal candidate would involve a massive chain shift of augmented words.[10]

In addition to being quite unlike any attested historical change, a shift of this sort would render intractable Padgett's (2003: §3.1) simplifying assumption that a language can be represented in a tableau as a small set of idealized forms. Because

[10] This scenario owes its inspiration to Cantor's (1892) diagonalization argument demonstrating the uncountability of the real numbers.

the forms in question represent phonotactically possible words rather than actual words, the full set of forms is, unlike an actual lexicon, at least potentially infinite. So long as the only changes contemplated in the tableau are substitutions, it is feasible to focus on small clusters of similar forms. However, if each form under consideration must be compared to one that is one segment longer, the candidate languages may need to be evaluated in their entirety.

This example may seem far-fetched, perhaps in part because it involves an unusually low ranking of a major faithfulness constraint. However, a system that includes constraints designed to evaluate the phonetic distinctiveness of entire words does open up the possibility that marginal contrasts may be "enhanced" by the introduction of apparently irrelevant changes. If such constraints exist and are freely rankable with respect to faithfulness constraints, then something very much like Ptydepe becomes a typological possibility.[11]

5 Concluding remarks

As the range of work touched on in the preceding sections suggests, there is considerable support among phonologists for the proposition that contrast is important, but there is rather less agreement about *how* it is important, or even about what, precisely, contrast is. Phonological theories that may be broadly described as based on the notion of contrast include formal algorithms for deriving minimally specified underlying representations (e.g. the Successive Division Algorithm of Dresher 2009), constraints that evaluate surface forms in minute phonetic detail (e.g. Dispersion Theory), and many things in between. Faced with the diversity not only of these theories themselves, but of their goals, it can be difficult to imagine a way of synthesizing them, or even of drawing meaningful comparisons among them.

There are, however, at least some areas in which these disparate theories may be brought to bear on similar sorts of questions. One of these is the shape of phonological inventories, discussed above in §4.2.1 in connection with Dispersion Theory. While the notion of a contrastive hierarchy, as described in §3.2, applies specifically to contrasts at the phonemic level, this approach to phonological feature specifications can also make significant predictions about the phonetic shape of inventories, particularly if it is used in combination with some theory of phonetic enhancement (on which see e.g. Keyser and Stevens 2001). Consider, for example, the highly implausible three-vowel inventory in (26), which represents a segmental analogue to the word-level indistinctness of Chorukor.

(26) *An unlikely three-vowel inventory*

ɨ	ʉ
ɘ	

The inventory in (26) is unlikely because, as noted by some of the work discussed in §4.2.1, contrasting segments tend to distribute themselves in such a way as to

[11] See also Dresher (2009: §8.3) for some other objections to Padgett's (2003) analysis of post-velar fronting and an alternative account based on the notion of a contrastive hierarchy.

take maximal advantage of the phonetic space available to them, as in the widely attested three-vowel inventory /i a u/. Dispersion Theory addresses this fact by positing constraints that evaluate the phonetic robustness of an entire system of contrasting segments or words.

The Successive Division Algorithm, as Hall (2007b: §4.3.2) points out, offers an alternative explanation. If the algorithm is used to assign features to the segments in (26), there will not be very many ways of doing so: the only features that can divide this inventory are [±high] and [±round]. If [±high] takes wider scope, the features will be assigned as in (27a); if [±round] is applied first, the features will be as in (27b).

(27) a. *Specifications for* (26) *with* [±high] >> [±round]

	ɨ	ɘ	ʉ
[±high]	+	−	+
[±round]	−		+

b. *Specifications for* (26) *with* [±round] >> [±high]

	ɨ	ɘ	ʉ
[±round]	−	−	+
[±high]	+	−	

What is noteworthy about the sets of specifications in (27) is that they are both entirely compatible with the inventory /i a u/. In a theory of full specification, the vowels in (26) would be identifiable as /ɨ ɘ ʉ/, because they would be specified for all the features they have in common as well as for the features that differentiate them; the Successive Division Algorithm can mark only how they differ from one another. If the Successive Division Algorithm is combined with some simple rules for robust phonetic implementation – e.g. a [−high] vowel is realized as low unless it is specified as [−low]; a [+round] vowel is realized as back unless it is specified to the contrary; etc. – then the inventory as specified in (27) will not only be phonologically indistinguishable from /i a u/, but it will even be pronounced as [i a u] or some other similarly dispersed set of surface forms. While the Successive Division Algorithm by itself does not guarantee phonetic dispersedness, it does guarantee that every specified feature serves to distinguish a segment from some other member or members of the inventory. If no redundant features are specified, then enhancement of specified features will contribute to surface dispersedness. Under this view, minimal representation of phonemic contrasts in underlying forms is a crucial part of a system that generates enhanced phonetic contrasts at the surface level, but that does so without making any explicit comparison between the surface forms themselves.

One theme that occurs repeatedly in the disparate treatments of phonological contrast is the relevance of information theory to phonology. As mentioned in §1, it was information theory that inspired Cherry *et al.* (1953) to apply a contrastive hierarchy to the phonemic inventory of Russian. For contemporary functionalist theories, the ability of the phonetic signal to communicate information is one of the major forces driving all phonological patterns. For Hall (2007b: §4.3.1) and Dresher (2009: §2.5.3) diagrams similar to the adjacency graphs of Shannon (1956) help to identify the type of inventory for which the Pairwise Algorithm (discussed above in §3.1) is inadequate. Broe (1993) uses information theory in constructing a theory of phonological similarity, the converse of phonological contrast. For Hume (2005), the information-theoretic concept of predictability provides new insight into phonological phenomena previously attributed to markedness (see also CHAPTER 4: MARKEDNESS). Hall (2009) uses entropy – Shannon and Weaver's (1949) formal measure of uncertainty – as a means of quantifying the degree to which two phones contrast with each other, thereby creating a non-categorical theory of phonemic contrast.

While there is, at this point, no overarching consensus on the role of contrast in phonology, or even on what the right questions to ask about it are, there are some interesting points of tangency, and sometimes agreement, among the contrasting views of contrast. There is reason, then, to think that the tensions among them – between phonemic and phonetic, formal and functional, abstract and concrete – are productive ones, and that, as Aristotle (1991: 240) suggested, "understanding is made greater by contrast." *Vivent les différences!*

ACKNOWLEDGMENTS

I am grateful to the editors and two anonymous reviewers for their helpful comments and suggestions. This chapter has also benefiited from discussions with Elan Dresher, Elizabeth Cowper, Marc van Oostendorp, Aleksei Nazarov, and Oliver Lenz. Any remaining errors, omissions or other infelicities are entirely my own fault.

REFERENCES

Anderson, Stephen R. 1985. *Phonology in the twentieth century: Theories of rules and theories of representations.* Chicago: University of Chicago Press.

Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5. 183–207.

Archangeli, Diana & Keiichiro Suzuki. 1997. The Yokuts challenge. In Iggy Roca (ed.) *Derivations and constraints in phonology*, 197–226. Oxford: Clarendon.

Aristotle. 1991. *The art of rhetoric.* London: Penguin. Translated and edited by H. C. Lawson-Tancred.

Barrie, Michael. 2006. Tone circles and contrast preservation. *Linguistic Inquiry* 37. 131–141.

Beckman, Jill N. 1998. Positional faithfulness. Ph.D. dissertation, University of Massachusetts, Amherst.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives.* The Hague: Holland Academic Graphics.

Boersma, Paul. 2000. The OCP in the perception grammar. Unpublished ms., University of Amsterdam (ROA-435).

Bradley, Travis G. 2001. The phonetics and phonology of rhotic duration contrast and neutralization. Ph.D. dissertation, Pennsylvania State University.

Broe, Michael B. 1993. Specification theory: The treatment of redundancy in generative phonology. Ph.D. dissertation, University of Edinburgh.

Bühler, Karl. 1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache.* Jena: Gustav Fischer.

Cantor, Georg. 1892. Über ein elementare Frage der Mannigfaltigkeitslehre. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 1. 75–78.

Chambers, J. K. 1973. Canadian Raising. *Canadian Journal of Linguistics* 18. 131–135.

Cherry, E. Colin, Morris Halle & Roman Jakobson. 1953. Toward the logical description of languages in their phonemic aspect. *Language* 29. 34–46.

Clements, G. N. 1987. Towards a substantive theory of feature specification. Paper presented at the UCLA Symposium on Segment Structure.
de Lacy, Paul. 2002. The formal expression of markedness. Ph.D. dissertation, University of Massachusetts, Amherst.
Dresher, B. Elan. 2003. The contrastive hierarchy in phonology. *Toronto Working Papers in Linguistics* 20. 47–62.
Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.
Dresher, B. Elan & Xi Zhang. 2005. Contrast and phonological activity in Manchu vowel systems. *Canadian Journal of Linguistics* 50. 45–82.
Dresher, B. Elan, Glyne L. Piggott & Keren Rice. 1994. Contrast in phonology: Overview. *Toronto Working Papers in Linguistics* 13: iii–xvii.
Dyck, Carrie. 1995. Constraining the phonology–phonetics interface, with exemplification from Spanish and Italian dialects. Ph.D. dissertation, University of Toronto.
Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles.
Flemming, Edward. 2002. *Auditory representations in phonology*. London & New York: Routledge. Revised version of Flemming (1995).
Goddard, Ives. 1974. An outline of the historical phonology of Arapaho and Atsina. *International Journal of American Linguistics* 40. 102–116.
Goldsmith, John A. 1995. Phonological theory. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 1–23. Cambridge, MA & Oxford: Blackwell.
Gussmann, Edmund. 1976. Recoverable derivations and phonological change. *Lingua* 40. 281–303.
Hall, Daniel Currie (ed.) 2003. Special issue on contrast in phonology. *Toronto Working Papers in Linguistics* 20.
Hall, Daniel Currie. 2007a. Laryngeal underspecification and Richness of the Base. In Sylvia Blaho, Patrik Bye & Martin Krämer (eds.) *Freedom of Analysis?*, 11–34. Berlin & New York: Mouton de Gruyter.
Hall, Daniel Currie. 2007b. The role and representation of contrast in phonological theory. Ph.D. dissertation, University of Toronto.
Hall, Kathleen Currie. 2009. A probabilistic model of phonological relationships from contrast to allophony. Ph.D. dissertation, Ohio State University.
Halle, Morris. 1959. *The sound pattern of Russian: A linguistic and acoustical investigation*. The Hague: Mouton.
Harms, Robert T. 1973. How abstract is Nupe? *Language* 49. 439–446.
Havel, Václav. [1966] 1991. *Vyrozumění*. [*The memorandum*.] In *Hry*, 47–104. Prague: Lidové Noviny.
Hume, Elizabeth. 2005. Deconstructing markedness: A predictability-based approach. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 13. 182–198.
Hyman, Larry M. 1970. How concrete is phonology? *Language* 46. 58–76.
Hyman, Larry M. 1973. Nupe three years later. *Language* 49. 447–452.
Inkelas, Sharon. 1995. The consequences of optimization for underspecification. *Papers from the Annual Meeting of the North East Linguistic Society* 25. 287–302.
Inkelas, Sharon. 1996. Archiphonemic underspecification: An optimization approach to the phonological description of morphemes. Unpublished ms., University of California, Berkeley.
Jakobson, Roman. 1949. On the identification of phonemic entities. *Travaux du Cercle Linguistique de Copenhague* 5. 205–213.
Jakobson, Roman & Morris Halle. 1956. *Fundamentals of language*. The Hague: Mouton.
Johnson, Keith. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34. 485–499.
Joos, Martin. 1942. A phonological dilemma in Canadian English. *Language* 18. 141–144.

Kaye, Jonathan. 1974. Opacity and recoverability in phonology. *Canadian Journal of Linguistics* 19. 134–149.
Keating, Patricia. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60. 286–319.
Keyser, Samuel J. & Kenneth N. Stevens. 2001. Enhancement revisited. In Michael Kenstowicz (ed.) *Ken Hale: A life in language*, 271–291. Cambridge, MA: MIT Press.
Kiparsky, Paul. 1968. How abstract is phonology? Distributed by Indiana University Linguistics Club. Reprinted in Paul Kiparsky (1982), *Explanation in phonology*, 119–163. Dordrecht: Foris.
Kiparsky, Paul. 1973. Phonological representations: Abstractness, opacity, and global rules. In Osamu Fujimura (ed.) *Three dimensions of linguistic theory*, 57–86. Tokyo: Taikusha.
Kirchner, Robert. 1995. Contrastiveness is an epiphenomenon of constraint ranking. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 21. 198–208.
Kirchner, Robert. 1997. Contrastiveness and faithfulness. *Phonology* 14. 83–111.
Kirchner, Robert. 1998. An effort-based approach to consonant lenition. Ph.D. dissertation, University of California, Los Angeles.
Kisseberth, Charles W. 1976. The interaction of phonological rules and the polarity of language. In Andreas Koutsoudas (ed.) *The application and ordering of phonological rules*, 41–54. The Hague: Mouton.
Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John J. Ohala & Jeri J. Jaeger (eds.) *Experimental phonology*, 13–44. Orlando: Academic Press.
Lindblom, Björn. 1990. Phonetic content in phonology. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm* 11. 101–118.
Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17. 267–302.
Łubowicz, Anna. 2003. Contrast preservation in phonological mappings. Ph.D. dissertation, University of Massachusetts, Amherst.
Łubowicz, Anna. 2007. Paradigmatic contrast in Polish. *Journal of Slavic Linguistics* 15. 229–262.
Mackenzie, Sara. 2005. Similarity and contrast in consonant harmony systems. *Toronto Working Papers in Linguistics* 24. 169–182.
Mackenzie, Sara. 2009. Contrast and similarity in consonant harmony processes. Ph.D. dissertation, University of Toronto.
Martinet, André, 1964. *Elements of general linguistics*. Chicago: University of Chicago Press.
McCarthy, John J. 2005. Taking a free ride in morphophonemic learning. *Catalan Journal of Linguistics* 4. 19–55.
Nevins, Andrew. 2004. Conditions on (dis)harmony. Ph.D. dissertation, MIT.
Ní Chiosáin, Máire & Jaye Padgett. 1997. Markedness, segment realization, and locality in spreading. Technical Report LRC-97-01, Linguistics Research Center, University of California, Santa Cruz.
Ní Chiosáin, Máire & Jaye Padgett. 2001. Markedness, segment realization, and locality in spreading. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 118–156. Cambridge: Cambridge University Press.
Padgett, Jaye. 2003. Contrast and post-velar fronting in Russian. *Natural Language and Linguistic Theory* 21. 39–87.
Paradis, Carole & Jean-François Prunet. 1991. Introduction: Asymmetry and visibility in consonant articulations. In Carole Paradis & Jean-François Prunet (eds.) *The special status of coronals: Internal and external evidence*, 1–28. San Diego: Academic Press.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.
Rice, Keren. 1995. On vowel place features. *Toronto Working Papers in Linguistics* 14. 73–116.

Rice, Keren. 2002. Vowel place contrasts. In Mengistu Amberber & Peter Collins (eds.) *Language universals and variation*, 239–270. Westport, CT: Praeger.

Sanders, Nathan. 2003. Opacity and sound change in the Polish lexicon. Ph.D. dissertation, University of California, Santa Cruz.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. 5th edn. 1962. Lausanne & Paris: Payot.

Shannon, Claude Elwood. 1956. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, IT-2. Reprinted 1993 in N. J. A. Sloane & Aaron D. Wyner (eds.) *Claude Elwood Shannon: Collected papers*, 221–238. Piscataway: IEEE Press.

Shannon, Claude Elwood & Warren Weaver. 1949. *The mathematical theory of communication*. Urbana-Champaign: University of Illinois Press.

Steriade, Donca. 1987. Redundant values. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 339–362.

Steriade, Donca. 1997. Phonetics in phonology: The case of laryngeal neutralization. Unpublished ms., University of California, Los Angeles.

Tessier, Anne-Michelle. 2004. Input "clusters" and contrast preservation in OT. *Proceedings of the West Coast Conference on Formal Linguistics* 23. 759–772.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.

Tryon, Darrell T. 1970. *An introduction to Maranungku (Northern Australia)*. (Pacific Linguistics B15.) Canberra: Australian National University.

3 Learnability

Jeffrey Heinz
Jason Riggle

1 Introduction

How can anything (such as a child or any other computing device) automatically acquire (any aspect of) a phonological grammar on the basis of its experience?

This is the fundamental (and unresolved) question of phonological learnability, and it is essentially independent of whether the grammar is taken to be the best description of the cognitive state of the language faculty, or whether the grammar is taken to be the best description of a language pattern, independent of its cognitive status. The problem of learning a general pattern from a finite set of observations has roots in the philosophical problem of inductive logic (Popper 1959; Sober 2008). Though language learning is simply a specific instance of this problem, it has played a perennially central role in the discussion. In fact, the modern formal study of learnability was inspired by the problem of language acquisition and much of the learnability literature is couched in formal language theory, whose early period was also influenced by the founding of generative linguistics. There have been many developments in formal learning theory and related disciplines, such as grammatical inference, computational learning theory, and machine learning. For the purposes of this chapter we will refer to all of these areas with the term *learning theory*.

All characterizations of learning – whether the domain is syntax, phonology, or gardening, and whether the models are connectionist, Bayesian, or symbolic – are subject to the results of learning theory. Even if they are not intended as such, answers to the question posed at the outset of this chapter constitute hypotheses about the broad characteristics of the computations that humans perform in learning the phonology of their language(s). Our goal in this chapter is to motivate the applicability of learning theory to the problem of learning phonological grammars. In this pursuit, we discuss but a fraction of the many grammatical formalisms and models of phonological learning that have been proposed (space does not permit a comprehensive survey, so we apologize in advance to those whose work is omitted in our brief discussion of the literature). Our main points are that learning theory: (i) makes clear *what* it is that is being learned; (ii) reveals little conceptual difference in the problems of learning gradient *vs.* categorical distinctions; (iii) makes a theory of universal grammar inevitable; (iv) can make clear which properties of phonological patterns are important for learnability; (v) emphasizes understanding the *general behavior* of learning models.

§2 reviews the foundations of formal language theory and its relevance to phonology. §3 covers the main contributions of a few of the theoretical learning frameworks to our understanding of the problem of learning phonological grammars. §4 examines the role of structure in generalization. §5 reviews several phonological learning models from the perspective of learning theory.

2 Formal language theory and phonology

How do we represent the patterns (i.e. languages) that a learner might attempt to learn? (See also CHAPTER 101: THE INTERPRETATION OF PHONOLOGICAL PATTERNS IN FIRST LANGUAGE ACQUISITION for a different perspective.) In formal language theory, languages are characterized as sets, relations, or, equivalently, functions (Harrison 1978; Hopcroft *et al.* 1979). This abstraction focuses attention on the patterns themselves rather than the particular grammars that describe the patterns. Though grammars typically reflect the generalizations that we are interested in, there can be many different ways to describe the same language.

2.1 *Phonotactic patterns*

For a concrete example, consider the set of all and only the words that obey a given phonotactic pattern. In this case, the phonotactic pattern makes a binary distinction between well-formed and ill-formed words (gradient distinctions are discussed later). For example, suppose that (1) designates all and only those words which obey the constraint that obstruents in codas do not disagree in voice.

(1) {fɪst, dæft, rabd, . . .}

Already the connection between the foundations of generative grammar and formal language theory are apparent. If the "three dots" in (1) are meant to include only actual English words, then clearly the set is finite. Generative phonologists reject such finite "list" representations, because the evidence is overwhelming that phonological competence goes beyond the finitely many words a speaker actually knows. In other words, the "three dots" are meant to include every conceivable word which includes many things that are not words of English, such as those in (2).

(2) {plɪst, θæft, wabd, . . . , peifɪst, . . .}

The fact that these sets can be infinite is what necessitates a generative grammar – that is, some finite device capable of generating all and only those *logically possible* words which obey the phonotactic pattern.

2.2 *Alternations*

(Morpho)phonological alternations can be represented as *relations* (i.e. sets consisting of all and only the pairs of words that obey the alternation). The example in (3) represents word-final obstruent devoicing found in languages like Dutch (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). Again, *all* logically possible pairs that obey the alternation are included (which, in the case of Dutch, includes pairs like (ag, ak) even though Dutch lacks /g/).

(3) {(ab, ap), (ad, at), (ag, ak), . . . , (bab, bap), (bad, bat), (bag, bak), . . .}

The pairs can be taken to mean that underlying /ab/ is realized as [ap], and so on. When an underlying form *u* is paired with a surface form s, we write ⟨*u*, *s*⟩. Again, much depends on how the "three dots" are interpreted. If they are interpreted as strictly as possible, then the alternation could be generated by the *SPE* rule:

(4) $\begin{bmatrix} \text{+voice} \\ \text{−son} \end{bmatrix} \rightarrow [\text{−voice}] \;/\; \_\_ \#$

In introductory phonology, it is often pointed out that the feature [+voice] in the target of the rule is unnecessary. In the interests of having shorter rules and rules which apply maximally without being falsified, the feature [+voice] is omitted in favor of the following.

(5) $[\text{−son}] \rightarrow [\text{−voice}] \;/\; \_\_ \#$

This would mean that the "three dots" are intended to include pairs like those in (6).

(6) {(ap, ap), (at, at), (ak, ak), . . .}

Thus the rule in (5) applies even to hypothetical form like /ap/, mapping it to [ap]. Koskenniemi (1983) takes this one step further, and considers the "three dots" to include pairs like those in (7).

(7) {(a, a), (as, as), (af, af), (ar, ar), (an, an), . . .}

In other words, all hypothetical underlying forms are included in the left-hand side of some pair. The corresponding *SPE* rule could be said to apply to all possible underlying forms, though in most cases its application is vacuous. The application only results in a change when the final consonant is a voiced obstruent. Mainstream phonology never adopted this perspective, for two reasons: it made the standard *SPE* rules more difficult to write (and sometimes more complex according to the *SPE* simplicity metric), and there were some discouraging complexity results (Barton 1986; Barton *et al.* 1987). But Koskenniemi observes a conceptual shift when thinking of the patterns in this way: rules can be thought of as constraints on alternations.

Note that all phonological knowledge can be deduced from the alternation pattern. In the case of phonotactic knowledge, this is straightforward: the set of forms in the right-hand side (the "surface forms") of the alternation pattern constitutes the infinite set of all forms that obey the surface phonotactic constraints of the language. Similarly, the set of forms in the left-hand side (the "underlying forms") is an infinite set which constitutes all forms that obey the Morpheme Structure Constraints (MSCs) of the language (CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS). If the notion of the rich base is adopted, there are no morpheme structure constraints, and the left-hand side becomes all logically possible underlying forms. Finally, the alternation pattern itself determines the contrasts. Generally, any two segments that are mapped to the same surface segment in all contexts (i.e. are neutralized) are not contrastive. Because knowledge of MSCs, phonotactics, and contrasts can all be deduced from alternations, the problem of devising learners for phonological alternations is one of the most important frontiers of phonological learnability.

2.3 *Gradience*

The above discussion only makes binary distinctions of well-formedness. Recently, however, many phonologists have argued for the importance of gradient distinctions (Coleman and Pierrehumbert 1997; Zuraw 2000; Albright and Hayes 2003; Coetzee 2008; Hayes and Wilson 2008). Gradient distinctions have been used to model the confidence of speakers in the face of lexical exceptions, variation in the productions of individual speakers, and variation across speakers in experimental settings.

The scope of formal language theory is not limited to binary distinctions. The sets and relations above can be thought of as functions whose domain is all logically possible words, or pairs of words, and whose co-domain is simply 0 and 1, for "ill-formed" and "well-formed," respectively; i.e. as *indicator functions*. Phonological patterns can also be thought of as functions whose co-domain is real-valued. Moreover, if these values sum to one, then the function is a probability distribution.[1]

From the perspective of formal language theory and learning theories, the differences between indicator functions and distributions are not particularly significant. Consider the Chomsky Hierarchy:

(8) finite $\subset$ regular $\subset$ context-free $\subset$ context-sensitive $\subset$ recursively enumerable

This inclusion hierarchy classifies patterns (e.g. sets of (pairs of) forms) in terms of the complexity of the kinds of formal devices (e.g. grammars) needed to generate them (see, e.g. Harrison 1978; Hopcroft *et al.* 1979; Thomas 1997). A remarkable range of ways to characterize complexity all converge on the distinctions in (8), which is why the hierarchy is considered to be so illuminating.

Crucially, the place of a function in the Chomsky Hierarchy is entirely independent of whether its co-domain is Boolean or real-valued. For learning theory – and the central problem of generalization – the co-domain matters little. Vapnik (1998: 8) writes:

[1] Formally, let Σ^* be the set of all logically possible words given a finite alphabet Σ. A pattern L is an indicator function if $L : \Sigma^* \rightarrow \{0, 1\}$. It is real-valued if $L : \Sigma^* \rightarrow \Re$ and it is a probability distribution iff $\sum_{w \in \Sigma^*} L(w) = 1$. If Δ is another alphabet, then $L : \Sigma^* \times \Delta^* \rightarrow \{0, 1\}$ is a Boolean alternation and $L : \Sigma^* \times \Delta^* \rightarrow \Re$ is a real-valued one.

> Generalizing the results obtained for estimating indicator functions (pattern recognition) to the problem of estimating real-valued functions (regressions, density functions, etc.) was a purely technical achievement. To obtain these generalizations, no additional concepts needed to be introduced.

Thus the choice of binary *vs.* gradient distinctions should depend simply on *what* one is trying to model.

2.4 Properties of phonological patterns

What are the properties of phonological patterns? When we consider alternations, it is the case that they can be described with any grammar capable of describing regular relations (Johnson 1972; Kaplan and Kay 1981, 1994; Karttunen 1993, 1998; Eisner 1997a, 1997b; Riggle 2004).[2] It is known that all regular relations have regular domains and co-domains, so it follows that all phonotactic patterns are regular as well. This is a striking hypothesis in light of the fact that some syntactic patterns appear to belong to higher levels of the Chomsky Hierarchy (Chomsky 1956; Joshi 1985; Shieber 1985; Kobele 2006).

Though limiting phonology to regular patterns is a significant restriction, it is not nearly restrictive enough. For instance, consider a hypothetical stress pattern consisting of all forms with an even number of stressed syllables. This pattern is regular, but it is wildly unlike those observed in natural language (see e.g. Hayes 1995; Gordon 2002; CHAPTER 41: THE REPRESENTATION OF WORD STRESS). Furthermore, though assuming that phonology is regular provides significant structure to the hypothesis space, there are many learning frameworks where this is still too little structure to guarantee learnability.

Learning theorists are interested in the properties that make patterns learnable. Linguistic properties are just now beginning to be investigated for their contributions to learnability. In the case of phonological patterns, it seems likely that the relevant properties will be subregular; that is, properties that carve out some proper subclass of the regular languages. Rogers and Pullum (2007) draw attention to the Subregular Hierarchy (McNaughton and Papert 1971), which classifies regular patterns according to the properties of different kinds of grammars capable of generating them. Additional recent work which attempts to relate phonological patterns to their place in the Subregular Hierarchy include Edlefsen *et al.* (2008), Graf (2010), and Heinz (2010).

3 Learning theory

3.1 Goals

There are many good resources on formal learning theory for phonologists. Nowak *et al.* (2002) provides an excellent, short introduction. Niyogi (2006) and de la Higuera (2010) provide detailed, accessible treatments, and Anthony and

[2] The notable exception to this is reduplication (CHAPTER 100: REDUPLICATION; CHAPTER 119: REDUPLICATION IN SANSKRIT), which is arguably a morphological process (Inkelas and Zoll 2005). For regular (finite-state) approaches to reduplication, see Roark and Sproat (2007).

Biggs (1992), Kearns and Vazirani (1994), and Jain *et al.* (1999) provide technical introductions. Here we summarize some of the main ideas and common results.

Learning theory characterizes learning, and the necessary and sufficient conditions required for learning strategies to be successful, or to exhibit some other particular behavior. This focus on characterizing a learner's behavior helps us understand precisely *why* a particular learning strategy succeeds in some cases, and helps us to characterize the class of cases where it may fail.

Learning theory defines learners as functions which map experience to grammars. The experience of the learner is necessarily finite, but the target languages typically are not. Any learning procedure can be characterized in this way, including learners that are connectionist (e.g. Rumelhart and McClelland 1986) or Bayesian (Griffiths *et al.* 2008), and learners based on maximum entropy (e.g. Goldwater and Johnson 2003), as well as those embedded within generative models such as Recursive Constraint Demotion (Tesar 1995; Tesar and Smolensky 1998) and minimal generalization (Albright and Hayes 2002; Albright 2009). Results of formal learning theory apply to all of these particular proposals and many others.

By characterizing learning algorithms as functions, it is possible to focus on the functional behavior of a learning strategy rather than its procedural description. This allows one to identify relevant properties of the mapping – like the linguistic typology predicted by a function's range – that are independent of the algorithm's implementation. Moreover, these properties are often crucial in understanding precisely what kinds of patterns learners are guaranteed to learn, and where they can fail.

Learning functions can also be characterized in terms of their computational complexity. Some learning procedures may require unreasonable resources and time. The exact meaning of "unreasonable" is studied in a number of works, including Pitt (1989) and de la Higuera (1997).

3.2 Learning frameworks

In §3.2.1–§3.3 we survey three learning frameworks: Identification in the Limit from Positive Data (Gold 1967), Probably Approximately Correct learning (Valiant 1984), and the Mistake Bounds model (Littlestone 1988). Other frameworks are discussed in §3.3.1, and the major results of learning theory are given in §3.4. Across the frameworks, precisely the same conclusion explicates the necessity of (some form of) Universal Grammar: namely, without a structured, restricted hypothesis space, feasible learning is impossible.

3.2.1 Identification in the Limit from Positive Data

In the Identification in the Limit from Positive Data (ILPD) framework, there are no limits on the learner's computational resources or time, and the input is assumed to consist of an infinitely long noise-free text that contains at least one instance of every form in the target pattern. Learners are partial functions, which map initial finite portions of these texts to grammars. A learner is said to *converge* to a grammar G if and only if at some finite point every future hypothesis is G. The learner is said to identify a language (or class of languages) in the limit just in case the learner converges to a grammar that generates the target language for *any* text from the target language (for any member of the class of languages).

Though the learner's input is generously assumed to include every potentially useful finite collection of forms from the target, the criterion for success is very strict: for *any* logically possible text for the language, the learner must find a grammar that generates the target pattern without a single deviation. This scenario focuses the learning problem squarely on generalization. Given as much *finite* experience as desired, can any learning device, no matter how powerful, exactly learn some language pattern, which may be *infinite* in size? How can the learning device cover the gap between its finite experience and the infinite set which represents the capacity of a normal speaker?

3.2.2 *Probably Approximately Correct learning*

Another model that has received a great deal of attention in learning theory is the Probably Approximately Correct (PAC) framework, first introduced by Valiant (1984) and subsequently developed by Kearns *et al.* (1987), Angluin (1988a), and Blumer *et al.* (1989), among others.[3] This model offers a probabilistic perspective on efficiently learning a class of languages in terms of the probability of attaining a hypothesis that has a low likelihood of making errors modulo the number of training samples observed.

The input to the learners is determined by drawing elements from the instance space X of data points, according to a probability distribution Π. Instead of exact identification, the quality of a learner's hypothesis h is evaluated in terms of the probability that h disagrees with the target language l for any $x \in X$, randomly drawn according to distribution Π. The ingenious aspect of the PAC model is that it does not matter what the distribution is, only that the same distribution Π is used for training and for evaluation.

The *error* of a hypothesis, denoted *error(h)*, is the sum of the probability that Π assigns to data points where h disagrees with the target. Analysis of learnability in the PAC model centers on the following question:

(9) For a given level of error ε, if a learner is presented with m samples drawn from X at random according to Π, what is our confidence δ that *error(h)* of the learner's hypothesis h is less than ε?

For a language class $\mathcal{L}$ and any given learning strategy, the *sample complexity* is the number of samples m needed to ensure that, for any $l \in \mathcal{L}$ and *any* distribution Π, the likelihood is at least δ that a learner will generate a hypothesis whose error is at most ε. This leads to the following definition of PAC-learnability:

(10) $\mathcal{L}$ is PAC-learnable iff the sample complexity of $\mathcal{L}$ is a polynomial function of ε and δ.

The PAC-learning framework differs from IDLP in two important respects. In one sense, the PAC model is more stringent because the required training data and computation must be *feasible* (i.e. polynomial). But, in another sense, the PAC model

3 See Haussler *et al.* (1992) and Haussler (1995) for overviews of work in learnability theory and insights into the deep connections between the PAC, Bayesian, and mistake-bound perspectives.

is less stringent than the Gold model, in that it loosens the definition of success from exact identification of a language to *approximate* identification that is likely to be correct *most* of the time.

3.3 Mistake bounds

Littlestone (1988) observes that, in many cases of interest, learnability can be characterized by the fact that the number of mistaken classifications – and subsequent corrections – is bounded. In this online framework, a learning algorithm $\mathcal{A}$ must classify each form it observes according to its current hypothesis h, which may be updated after the correct classification is revealed. The *mistake bound* for $\mathcal{A}$ on language class $\mathcal{L}$, denoted $M_{\mathcal{A}}(\mathcal{L})$, is the number of mistaken classifications that $\mathcal{A}$ might make when facing a diabolical adversary who knows $\mathcal{A}$'s strategy and has boundless computing resources to choose the hardest language in $\mathcal{L}$, and the least helpful presentation of examples. The *optimal mistake bound* for $\mathcal{L}$, denoted $Opt(\mathcal{L})$, is the smallest $M_{\mathcal{A}}(\mathcal{L})$ for any possible $\mathcal{A}$.

Littlestone (1988) shows that if $Opt(\mathcal{L})$ is finite, then it is the case that the class $\mathcal{L}$ is both identifiable in the limit and PAC-learnable. The converse, however, does not hold; neither PAC nor Gold learnability guarantees a finite mistake bound. In the former case there might be an infinite sequence of imperfect hypotheses that all have error less than ε, and in the latter case one might be able to guarantee that the number of mistakes will be finite without being able to give a specific bound.

3.3.1 *Other frameworks*

There are other learning frameworks. Some enrich the learner's input in particular ways, which gives the learner more information and generally leads to stronger positive results. For example, Gold (1967) also considers the case of learning from positive and negative data. In this scenario, the entire class of recursively enumerable languages is learnable in principle, though no learner is efficiently computable for even the regular languages (Gold 1978). Gold also shows that restricting texts to those with certain useful kinds of structure (for example, by only allowing texts whose structure is describable with primitive recursive functions; see also Rogers 1967) can also guarantee the learnability of the recursively enumerable languages. This means that knowing crucial properties of the presentations of the data can, like negative evidence, make a huge contribution to pattern learning. However, it is highly doubtful that the natural language data children observe have either of these properties (note that occasional overt corrections do not necessarily constitute negative evidence).

Similarly, Horning (1969) shows that, when learning stochastic languages (distribution learning), if it is the case that learners are required to succeed only on texts generable by the target distribution then it follows that probabilistic context-free languages can be learned (see also Osherson *et al.* 1986). Angluin (1988b) supersedes Horning's result, and shows that the entire class of recursively enumerable distributions is in fact learnable in this sense. Like Gold's result above, these results suggest that knowing properties of the presentations of the input data dramatically increases what is learnable in principle. Crucially, however, the learners in these proofs are not remotely feasible, so these results do not inform human language learning.

3.4 *Main results*

In the ILPD, PAC, and MB frameworks surveyed above, there are significant negative results: none of the major classes in the Chomsky Hierarchy is learnable. In the case of identification in the limit from positive data, no class which is a proper superset of the finite languages is learnable. In the PAC and MB models, not even the finite class of languages is learnable. In other words, there is no learner, not even in principle, that can PAC-learn or identify in the limit from positive data all regular, context-free, or context-sensitive language patterns.

There are many ways to interpret this result (see, for example, Pinker 1979). Gold mentions restricting the problem so that not *all* regular (or context-sensitive) patterns are permitted in natural language. This possibility is promising for three reasons. First, the field of grammatical inference has identified many classes of languages that are ILDP and PAC-learnable (Angluin 1982; Muggleton 1990; Clark and Eyraud 2007; Heinz 2008; de la Higuera 2010). Many of these classes contain infinitely many patterns, and some include context-free, even context-sensitive patterns. In virtually every case, the successes occur because the language classes *are non-arbitrary* in important ways: the hypothesis space is structured. Secondly, this possibility makes sense from the studies of distribution learning above: while recursively enumerable distributions are learnable in principle they are not feasibly learnable in practice. The efficiency issues can be overcome by restricting the class of distributions to be learned (if doing so adds sufficient structure to the hypothesis space). Finally, this possibility also matches well with language typologists' repeated observations that the extensive variation that exists in natural languages appears to be limited, though stating exact universals is difficult (Greenberg 1963, 1978; Mairal and Gil 2006; Stabler 2009).

The results surveyed above lead to the following conclusion: structure matters. In particular, if the collection of language patterns to be learned has the right kind of structure – the right kind of properties – then learning is possible. The most interesting learners will use the structure or properties in the language class to license the right generalizations from their finite experience to an infinite pattern. Conversely, these results show that there is essentially no hope of learning in cases where the range of possible patterns is too unstructured.

4 The role of structure in generalization

The *structure* of the hypothesis space is what allows for generalization. In this section, we discuss very general structural properties important to learnability. We begin with a discussion of finite hypothesis spaces, then turn to structure related to what has been called the *subset problem*, and conclude with a general metric of structure known as the Vapnik-Chervonenkis Dimension.

4.1 *Finiteness as a kind of structure*

Many linguistic theories, such as Principles and Parameters and Optimality Theory, only allow finitely much variation in the typology, thereby providing a finite collection of languages. This property of hypothesis spaces is a sufficient property for success in many frameworks, including PAC and IDLP. A common brute-force strategy for any finite hypothesis space is to essentially match all grammars with the learning data, and choose the one that is the most consistent.

Although it is a sufficient property for learning, finiteness is hardly an interesting property. To see why, recall the earlier discussion of quantity insensitive (QI) stress patterns, and let us artificially place an upper bound on word length so that we only have finitely many patterns to consider. For ease of illustration we set the bound at four). If we restrict ourselves to just one string of each length, then there are 2^{10} = 1,024 logically possible patterns, of which eight are shown in (11).

(11) *Some logically possible stress patterns over 1–4 syllables*

	natural QI stress systems	*unnatural systems*
Initial	1 10 100 1000	0 01 100 0101
Final	1 01 001 0001	1 10 101 0110
Edges	1 11 101 1001	1 00 000 1010
Binary-initial	1 10 101 1010	1 01 101 1100

The artificial bound limits the class in a very significant but uninteresting way, because almost all of these 1,024 patterns belong in the "unnatural" column. The properties that determine which patterns belong to the "natural" column are going to be precisely those same linguistic properties that hold regardless of whether the class is finite or infinite. It is of far greater interest how those properties – and not finiteness – structure the hypothesis space.

Finiteness is hardly a necessary property for learnability – many infinite language classes are efficiently learnable because they have structure that learners can utilize (Jain *et al.* 1999; de la Higuera 2010). On the other hand, brute-force learners that simply traverse an enumeration of all hypotheses are not generally feasible (since finite classes can still be very large). Even for the finite case, the interesting learners are those that make use of structure (see e.g. Recursive Constraint Demotion; Tesar and Smolensky 2000).

4.2 Tell-tale sets and the subset problem

Angluin (1980) provides one benchmark for necessary and sufficient structure in a hypothesis space. If every language pattern L in the hypothesis space contains a finite set S, such that no other language pattern L' in the hypothesis space is simultaneously a superset of S and proper subset of L (see Figure 3.1), then this hypothesis space is sufficiently structured to be identified in the limit from positive data. The finite set S is called a *tell-tale set*, and we call the above property of hypothesis spaces the *tell-tale property*.

The tell-tale property is sufficient for learning, because a learner that guesses L after exposure to its tell-tale set is guaranteed to have hypothesized the smallest language in the class consistent with the sample. Characterizing the tell-tale sets of a hypothesis space – and more generally, characterizing the finite experience a learner needs to generalize correctly to the language patterns in a hypothesis space – is one of the important lessons of learning theory. It adds to the functional characterization of the learner. This is because once the tell-tale sets are characterized, when given a learner and a language pattern L from the learner's

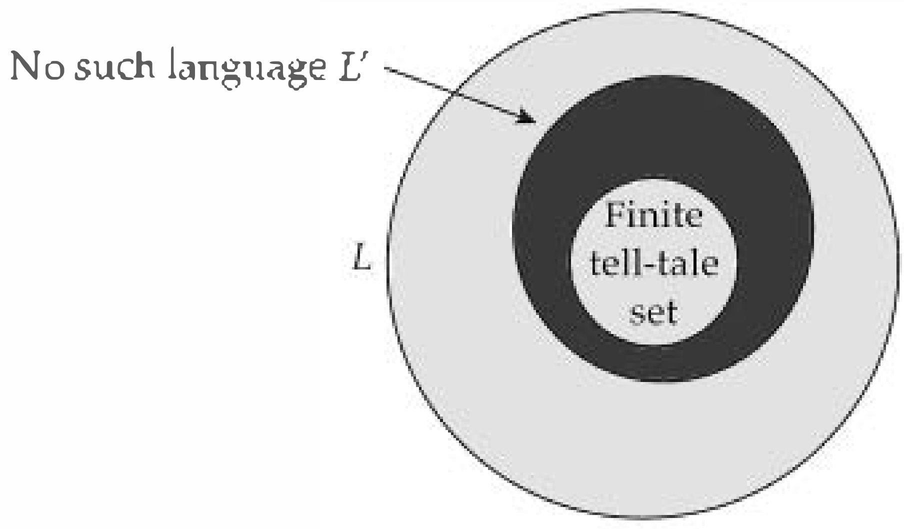


Figure 3.1 The tell-tale property

hypothesis space, one knows whether the learner will succeed given some finite sample S (by checking whether S is a tell-tale set).

Identifying properties of the tell-tale sets is important to phonologists for two reasons. First, it provides an additional way to evaluate learning proposals, since the tell-tale sets can be compared with the actual linguistic forms available to children. Secondly, knowledge of the properties of tell-tale sets allows one to understand how the learner generalizes, and may provide insight into stages of the learning process.

4.3 The Vapnik–Chervonenkis dimension

One particularly simple and robust metric of structure for concept classes is the combinatorial measure of complexity known as the Vapnik–Chervonenkis (VC) dimension (Vapnik 1998; Vapnik and Chervonenkis 1971). For a given concept class $\mathcal{L}$, the VC dimension (VCD) of $\mathcal{L}$ is the cardinality of the largest set of data S such that there is at least one language in $\mathcal{L}$ for each of the $2^{|S|}$ possible ways of labeling the data points in S as "ungrammatical" or "grammatical." If S has this property it is said to be *shattered* by $\mathcal{L}$.[4] If sets of arbitrary size are shatterable then the VCD is said to be infinite.

For an illustration, suppose that we represent coda clusters as points in $\mathcal{R}^2$ (the x–y plane) where the x-axis encodes the sonority of the second consonant and the y-axis the sonority of the first. Suppose further that $\mathcal{L}$ is the (infinite) set of languages corresponding to "half-spaces" defined by straight lines that split $\mathcal{R}^2$ into two regions, one for licit clusters and the other for illicit clusters. Figure 3.2 provides a rough example that situates the clusters *sn*, *pl*, *pt* in $\mathcal{R}^2$. In this example, a grammar that includes all three clusters can be obtained by drawing a line off to one side so that the illicit (shaded) area does not include the points. Grammars that include any two of the points can be obtained by drawing a line between the point to be excluded and the other two, and shading the side with the excluded point. These four possibilities make up the top row of Figure 3.2. The other four possibilities are illustrated in the bottom row of Figure 3.2; these are obtained by inverting the grammars in the top row. Since there is a grammar (i.e. a half-space)

[4] Formally, sample $S = \{x_1, \ldots, x_n\} \subseteq \mathcal{X}_n$ is shatterable if $\forall (v_1, \ldots, v_n) \in \{0, 1\}^n$, $\exists l \in \mathcal{L}$ such that $\forall i\ c(xi) = v_i$. The VC dimension of $\mathcal{L}$ is the cardinality of the largest shatterable sample: $vcd(L) = max\{|S| : S \text{ is shatterable}\}$.

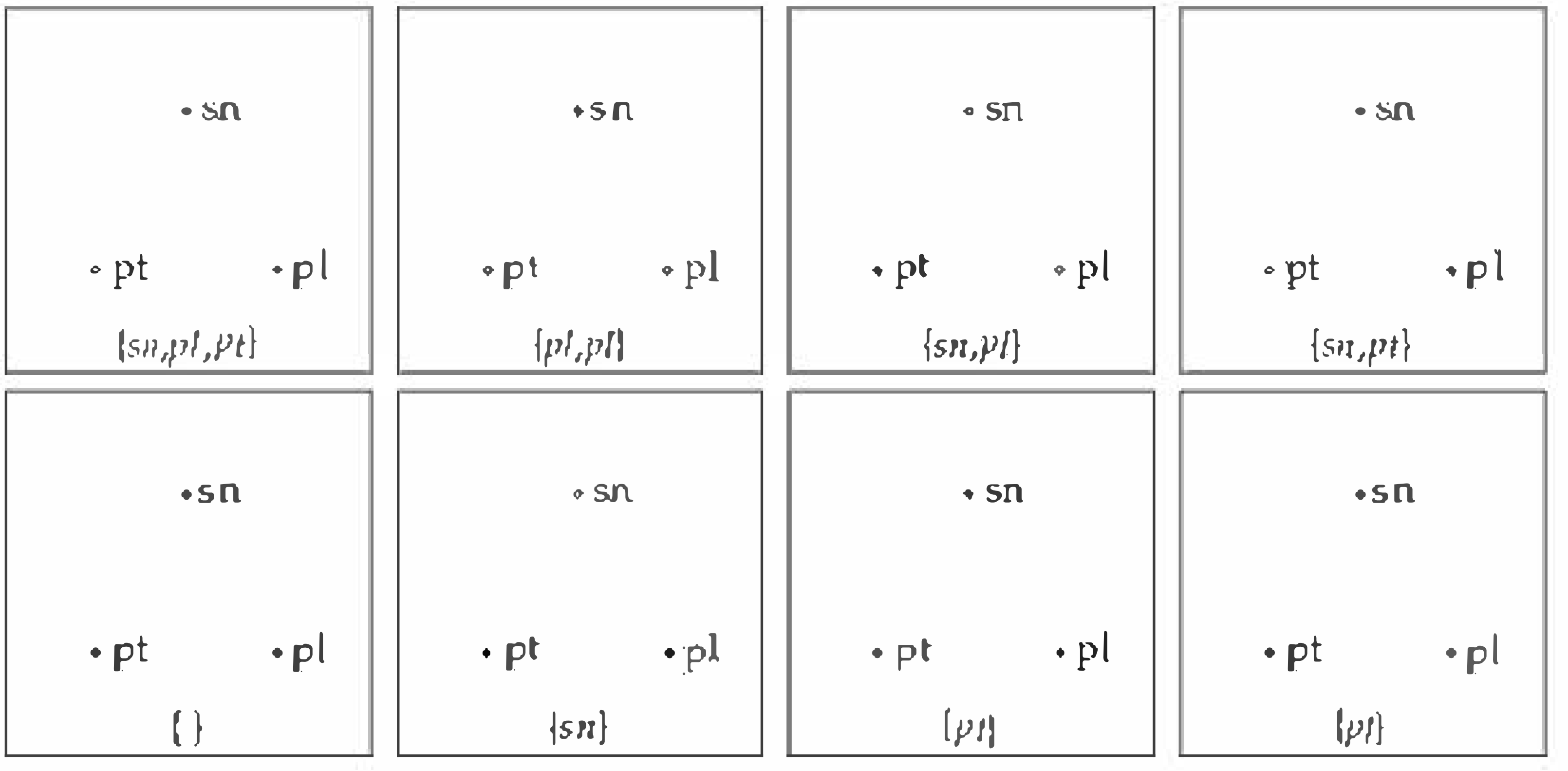


Figure 3.2 A set of three points that is shatterable by half-spaces in $\mathcal{R}^2$

for each of the eight logically possible grammatical/ungrammatical distinctions over these three points, we say that this set of points is shattered by the class $\mathcal{L}$.

Because there is a shatterable set of three points, the VC dimension of $\mathcal{L}$ is at least three. This does not entail that *every* set of three points is shatterable. For instance, any set of three collinear points cannot be shattered, because one of the points lies directly between the other two and thus cannot be separated by a half-space. This same logic explains why no set of four points in $\mathcal{R}^2$ is shatterable; either one point is inside a triangle whose corners are the other points, or the four points are the corners of a convex polygon. In the former case, no hyperplane can include the interior point while excluding all the points at the corners and, in the latter case, no hyperplane can include two opposing corners while excluding both points at the other corners. The fact that there are shatterable sets of three points, but no shatterable sets of four, places the VC dimension of half-spaces in $\mathcal{R}^2$ at three (for $\mathcal{R}^n$ it is n+1). What this means in terms of learnability is that, for any dataset with more than three points, there *must* be points whose grammaticality is interdependent.

To understand the role this structure plays in learning, consider a class $\mathcal{L}$ whose VC dimension is d and a learner with m data points. As long as $m \leq d$, it is possible that the labeling of every point is totally independent of the others. But, as soon as $m > d$, some generalization/prediction is always possible, because there are fewer than 2^m distinct ways to label the data points as grammatical or ungrammatical according to languages in $\mathcal{L}$ (otherwise the VCD would be higher). Furthermore, it turns out that when $m > d$, the number of possible labelings is a *polynomial* function of m. In essence, there is a sort of "phase transition" from exponentially many labelings when $m \leq d$ to only polynomially many when $m > d$, which makes the complexity of the hypothesis space a polynomial function of m when the VCD is finite.[5]

[5] By complexity, we mean information-theoretic complexity in the sense that $\mathcal{L}$ makes it possible to describe any labeling of $m > d$ data points with fewer than $\log_2 m$ bits. See Kearns and Vazirani (1994) for more discussion and for proofs.

A finite VCD is both necessary and sufficient for PAC-learnability (Blumer *et al.* 1989), and thus not even the class of finite languages (which has infinite VCD) is learnable in the PAC framework. It follows that none of the major classes of the Chomsky Hierarchy are PAC-learnable, which again suggests that the right characterization of the class of patterns in phonology is some class that cross-cuts the Chomsky Hierarchy.

Finally, we should note that substantive linguistic restrictions will shrink the VC dimension below the upper bound that follows from more general structural properties of the class. For example, in the grammars of Figure 3.2, phonetic factors such as ease of articulation or perception might rule out the possibility of languages that admit [pt] clusters while excluding [sl] and [sn] clusters. When additional properties such as implicational universals over sonority sequencing restrict a concept class, the VCD can quantify the structure that such factors bring to the learning problem.

4.4 Summary

The three kinds of structure surveyed here – finiteness, the tell-tale property, and the VC dimension – provide a foundation for phonologists to investigate the contribution phonological properties make to learning. Phonologists widely agree that there is intricate structure in phonological patterns. How this phonological structure relates to the structures that are relevant to learnability is a promising new research area.

5 Phonological learners

5.1 Learning rule-based alternations from pairs

Johnson (1984) presents an algorithm that takes as input a set of $\langle u, s\rangle$ pairs and returns segment substitution rules and their orderings that are logically consistent with the data. The class $\mathcal{L}$ of all languages (sets of $\langle u, s\rangle$ pairs) that are representable by ordered sequences of substitution rules is superfinite, and thus we know that this strategy cannot identify $\mathcal{L}$ in the limit from positive data. Johnson notes that this set of induced rules and orderings can be reduced via evaluation metrics and other heuristics grounded in language universals. The need for the latter shows that, while logical properties of phonological rules can restrict the hypothesis space, additional structure in linguistic systems must play a role in choosing among hypotheses.

Gildea and Jurafsky (1996) present an algorithm that takes as input $\langle u, s\rangle$ pairs from a dataset with some alternation, and returns a rule, which unlike Johnson's system can include deletion and epenthesis. Their work begins with a result from Oncina *et al.* (1993), who present an algorithm dubbed OSTIA, which identifies in the limit from positive data a subclass of regular relations describable by subsequential finite state transducers. Since the flapping rule of English can be represented this way, Gildea and Jurafsky ask whether OSTIA will acquire the

[6] Since the CMU dictionary does not include allophonic information, Gildea and Jurafsky modified the dictionary to replace [t] and [d] with [ɾ] in every instance where the rule would apply.

flapping rule from an appropriately modified version of the *Carnegie Mellon University Pronouncing Dictionary* (CMU 1993).[6] Essentially, they ask whether the CMU contains a tell-tale set (for OSTIA). Because OSTIA fails to learn the flapping rule from the CMU dictionary, the answer is no, probably because a tell-tale set would need to include non-English forms like *ttt*.[7]

Gildea and Jurafsky then augment OSTIA with three phonologically motivated principles. These are Faithfulness: underlying–surface pairs tend to be similar; Community: similar segments tend to behave similarly; and Context: phonological rules can access variables in their context. This modified OSTIA algorithm gets much closer to acquiring a rule that represents the English flapping alternation. Gildea and Jurafsky conclude that these biases aid learning, and argue for a research program for evaluating the contributions of such biases. We agree wholeheartedly; domain-appropriate biases that add structure to or otherwise reduce the hypothesis space are likely to aid learning by also reducing the size of tell-tale sets. However, it is critical to ask exactly how and why this occurs, and most crucially what class of rules are learnable with the biases in place. To our knowledge, neither of these interesting questions has been addressed.

Albright and Hayes (2003) also aim to learn alternations expressed by phonological rules. Their algorithm takes as input $\langle u, s\rangle$ pairs and returns a set of rewrite rules with confidence scores. A central idea in their rule construction procedure is a strategy called minimal generalization. The idea is that if two sounds are known to undergo some alternation, then one may conclude that all sounds in the smallest natural class containing those two sounds also undergoes the alternation (cf. the Community principle). In addition, the algorithm assigns a confidence score to each rule based on the frequency of the rule's application in the corpus. The confidence score can be used to analyze free variation, or phonologically conditioned allomorphy (as with the irregular English past tense).

Albright and Hayes do not focus on an analytical characterization of the class of languages that their algorithm can learn, but instead compare the behavior of their algorithm to the judgments of native speakers on "wug" tests (Berko 1958; CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY). These comparisons reveal intriguing correlations, but they are somewhat difficult to interpret. On the one hand, a shift in focus from the analysis of properties that define various learnable classes of languages to the behavior of humans is undoubtedly appealing to any who feel that the results of learnability theory are too abstract and remote from real-world learning problems. On the other hand, having observed that an algorithm $\mathcal{A}$ and human subject $\mathcal{H}$ give similar responses for a particular set of test items T after being exposed to a set of training data D, it is not clear what we can conclude about $\mathcal{H}$ or the relationship between $\mathcal{A}$ and $\mathcal{H}$, because they might wildly diverge for some other data T' and D'. The goal of determining which properties of the data critically underlie learnability – or in this case the correlation between $\mathcal{A}$ and $\mathcal{H}$ – is precisely why learning theory focuses mainly on the

[7] It should be emphasized that OSTIA learns a rule that is consistent with the data. It is just that the alternation that this rule describes is not the same infinite set of ⟨underlying form, surface form⟩ pairs that phonologists think the flapping rule ought to describe.

properties of classes of languages or the *general behavior of specific algorithms*, as opposed to the specific behavior of specific algorithms.

5.2 *Learning OT grammars*

Optimality Theory (Prince and Smolensky 1993) is a theory of grammar which characterizes alternations by a strict ranking of constraints which evaluate possible $\langle u, s\rangle$ pairs. A $\langle u, s\rangle$ pair belongs to the alternation just in case it is optimal among the (possibly infinite) range of $\langle u, s'\rangle$ pairs according to the ranked constraints. (CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS).

For a fixed (universal) set of k constraints there are at most $k!$ languages and thus any set of constraints defines a finite class of languages that is learnable in the limit. Though the members of any finite class of languages can be identified in the limit by enumerating the languages, such an approach is not feasible in practice. An early positive result for OT learning was provided by Tesar and Smolensky's (1993) Recursive Constraint Demotion (RCD) algorithm. Tesar and Smolensky (1996: 26) subsequently showed that the structure that ranked constraints given to the hypothesis space guarantees that RCD will successfully identify languages with a polynomial mistake bound (unlike a brute-force enumeration).

5.3 *The VC dimension of OT and HG*

As mentioned earlier, finitude is itself a very simple kind of structure for concept classes. With regard to the VC dimension, this is reflected by the fact that the VCD of any finite set of grammars is at most $\log_2$ of the cardinality of the set. This follows because it takes at least 2^n concepts to shatter a set of n data points. Hence the VCD of any set of OT grammars over a fixed set of k constraints is at most $\log_2 k!$, because there are only $k!$ possible rankings. By contrast, if we take the same constraints and consider grammars defined by real-valued *weightings* (as in Harmonic Grammar; HC)[8] there are infinitely many possible grammars and thus no a priori bound on the VCD.

This pair of cases proves to be quite illuminating. Though the finitude of $\mathcal{L}$ (or lack thereof) provides some information about its learnability, the characterization is both coarse and incomplete. In the case of OT, the finitude of the concept class bounds the VC dimension at $\log_2$ of $k!$ (which is on the order of $k \log_2 k$). Unsurprisingly, the hypothesis space has more structure than its mere finitude, and this structure bounds OT's VC dimension at $k-1$ (Riggle 2009). By contrast, one might expect the infinite hypothesis space of HG to have much less structure, but it turns out that learning weightings can be represented as the problem of learning half-spaces in $\mathcal{R}^k$ (as in Figure 3.2), so the VC dimension cannot be greater than $k+1$ and in fact is $k-1$ (Bane *et al.* 2010). This parity means not only that both models are efficiently learnable, but that the learning problems are essentially of equal complexity (recalling Vapnik's observation in §2.3).

[8] In addition to HG (Legendre *et al.* 1990; Smolensky and Legendre 2006; Pater 2009), a range of weighted models have been proposed by Goldsmith (1990, 1991, 1993a, 1993b) and a few others.

5.4 *PAC learning of rankings and weightings in OT and HG*

Both OT and HG have the same VC dimension: k–1 for grammars with k constraints. For a concrete example of what this means in terms of learnability, consider the three hypothetical tableaux in (12).

(12)

input 1	c_1	c_2	c_3	c_4	
Cand *a*		*			implication: $a > b$ iff $w_1 > w_2$ in HG or $c_1 >> c_2$ in OT
Cand *b*	*				implication: $b > a$ iff $w_2 > w_1$ in HG or $c_2 >> c_1$ in OT

input 2	c_1	c_2	c_3	c_4	
Cand *c*			*		implication: $c > d$ iff $w_2 > w_3$ in HG or $c_2 >> c_3$ in OT
Cand *d*		*			implication: $d > c$ iff $w_3 > w_2$ in HG or $c_3 >> c_2$ in OT

input3	c_1	c_2	c_3	c_4	
Cand *e*				*	implication: $e > f$ iff $w_3 > w_4$ in HG or $c_3 >> c_4$ in OT
Cand *f*			*		implication: $f > e$ iff $w_4 > w_3$ in HG or $c_4 >> c_3$ in OT

In both OT and HG it is possible to formulate sets of k–1 binary tableaux like those in (12), in which each of the exponentially many (i.e. 2^{k-1}) ways to choose a set of winners is possible under some grammar. However, as soon as a learner has seen k or more tableaux – in either model – there are only polynomially many ways to choose a set of winners (i.e. there is no set of four tableaux in which all patterns of winners are possible). The remarkable consequence of this fact is that any learner that meets the simple condition that its hypotheses are always consistent with all previous observations is guaranteed to PAC-learn a ranking/weighting from a set of training data whose size is a linear in the number of constraints.[9]

Given a constraint set and a dataset comprising ⟨winner, loser⟩ pairs, Recursive Constraint Demotion (Tesar and Smolensky 1993, 1998, 2000; Tesar 1995, 1997, 1998a, 1998b) constructs a stratified hierarchy $\mathcal{H}$ (i.e. a weak ordering) that is consistent with the data by constructing strata consisting of constraints for which, in each remaining ⟨w, l⟩ pair, w has no more violations than l, and then discarding any pair in which w is optimal according to the $\mathcal{H}$ constructed thus far. This process is reiterated until all ⟨w, l⟩ pairs are gone (or until no constraint favors a winner, in which case no ranking is consistent with the data). If, in addition to $\mathcal{H}$, RCD records the ranking conditions that support its correct predictions, then it can generate hypotheses consistent with all previous observations and thereby be guaranteed to PAC-learn rankings from in the order of k random samples (the extra record-keeping is needed to ensure consistency because "accidentally" correct predictions can be undone by subsequent updates to $\mathcal{H}$).

[9] The bound on sample complexity m, according to VC dimension d, is $m \leq \lceil (4/\varepsilon)\ (d \ln (12/\varepsilon) + \ln (2/\delta)) \rceil$; see e.g. Blumer *et al.* (1989).

For HG grammars, Potts *et al.* (2010) propose a consistent learner that finds a constraint-weighting $w = \langle w_1, w_2, \ldots, w_k \rangle \in \mathcal{R}^k$ that simultaneously satisfies all the linear inequalities that correspond to a set of ⟨winner, loser⟩ pairs – such as those in (12) – using a technique from linear programming called the *simplex algorithm* (see e.g. Papadimitriou and Steiglitz 1998: chapter 2). Though their learner is intended to operate over batches of $\langle w, l \rangle$ pairs, they could conceivably be recast as an "error-driven" learner, so that, rather than generating a new hypothesis for each new datum based on all prior observations, a new hypothesis would be generated only in the event of an erroneous prediction.

RCD also has an error-driven formulation, and an especially useful property of error-driven learners is that they only need to "remember" data points that they misclassified (often called "supports") in order to faithfully reconstruct correct predictions for all forms in the training sequence. This allows a mistake bound to double as a memory bound on the amount of information that a learner could ever need to store.

Pater (2008) observes that Rosenblatt's (1958) "perceptron" can be straightforwardly applied to HG learning. The perceptron is an error-driven learner that maintains a weighting $w = \langle w_1, w_2, \ldots, w_k \rangle \in \mathcal{R}^k$, with which they make predictions as follows. For candidates a and b, the value $\Delta(a, b) \in \mathcal{Z}^k$ is the result of subtracting b's violations from a's violations (e.g. in (10), $\Delta(a, b) = \langle -1, 1, 0, 0 \rangle$). This point in k-dimensional space is "in" just in case it lies within the half-space described by w (i.e. if the inner product $w \cdot \Delta(a, b)$ is greater than zero; this is a linear-classifier like the ones in Figure 3.2). Upon misclassifying a data point, the hyperplane represented by the weight-vector w is nudged in the direction of $\Delta(a, b)$. Though multiple errors on the same data point are possible (i.e. the update rule is non-corrective), the perceptron is guaranteed to eventually converge to a correct weighting if one exists. In the general case, the perceptron is not a PAC-learner, because the sample complexity can be exponential in k when the probability mass of Π is concentrated on positive and negative data points that are packed too close to the hyperplane that separates them. Moreover, though the perceptron will converge eventually, it is precisely these "hard" probability distributions that lead to many mistakes.

5.5 Mistake bounds in OT and HG

Regarding optimal mistake bounds, Littlestone (1988) shows that, while a lower bound on $Opt(\mathcal{L})$ is set by $\mathcal{L}$'s VC dimension, in cases where $\mathcal{L}$ is finite, the upper bound of $Opt(\mathcal{L})$ is $\log_2|\mathcal{L}|$. This follows because the strategy of making predictions that accord with a plurality of the hypotheses consistent with previous observations only errs on data points that half or fewer of the remaining hypotheses correctly classify (else the correct prediction would have been made) and, as such, each error halves the set of viable hypotheses which allows no more than $\log_2|\mathcal{L}|$ errors.

This suggests room for improvement over RCD's quadratic mistake bound of $k(k-1)/2$, which follows from the maximum number of stratified hierarchies that RCD can entertain on the way from all k constraints in a single stratum to a total order (Tesar and Smolensky 1996: 26). To implement Littlestone's halving algorithm for OT, Riggle (2008) proposes a recursive function for calculating the fraction of the space of possible rankings that are consistent with a set of optimal candidates, a quantity he calls the r-volume. For just two candidates a and b, if A denotes the constraints for which *a* has fewer violations and *B* those for which *b* has fewer violations, then the fraction of rankings that select *a* is precisely |A| / (|A|+|B|).

(13)

input	c_1	c_2	c_3	c_4	
Cand *a*		*	*	*	the *r*-volume of candidate *a* is ⅓ (i.e. 8 rankings)
Cand *b*	**		*		the *r*-volume of candidate *a* is ⅔ (i.e. 16 rankings)

Unfortunately, though computing *r*-volume for larger sets of candidates can often be done in ways vastly more efficient than exhaustive search, there are "hard" cases where computation will always be intractable.[10] This highlights the core tension between power and efficiency in learning; RCD's mistake bound may be sub-optimal but it is still polynomial and it is obtainable at amazingly low computational cost, whereas the halving algorithm yields a nearly optimal mistake bound (i.e. within a logarithmic factor of $k-1$), but does so by introducing computation that is intractable in the worst case.

Analysis of mistake bounds illuminates a significant point of divergence between OT and HG. Though the two models have the same VC dimension, the mistake bound of the former is finite, while the mistake bound of the latter is not. This is so because it is possible to construct a sample sequence of arbitrary length in which each new data point causes an error that leads to an ever smaller change in the weighting. Thus, though learners that use strategies such as the perceptron algorithm will eventually converge to a correct constraint weighting for any HG grammar (see Pater 2008), there is no general bound on the rate of convergence (i.e. the number of mistakes along the way) that holds for all possible sets of training data.

Partially due to this fact, much of the work on learning linear classifiers has focused on the way that specific properties of samples affect learnability. For instance, the quantity γ, known as the *margin*, measures the distance (in high dimensional space) between the grammatical and ungrammatical points and the line that separates them. Given γ, one can derive bounds on the number of mistakes and the rate of convergence. In fact, if the margin is large enough, it supplants the dimensionality of the sample space in determining the VC dimension of the learning problem. Thus, with large margins, HG grammars with thousands of constraints might nonetheless have very low mistake bounds and sample complexity, suggesting that searching for so-called large-margin classifiers might provide linguistic insights.

5.6 *Learning segmental adjacency patterns*

Hayes and Wilson (2008) develop a learner that takes as input a list of words and outputs a maximum entropy grammar consisting of a finite set of weighted constraints that define a probability distribution over forms. The algorithm has several properties of interest. First, the constraints it returns are essentially *n*-grams and thus, in its simplest form, the algorithm can learn adjacency patterns, but not harmony patterns. Secondly, the units in these constraints are feature bundles denoting natural classes. Thirdly, the algorithm is designed to first search for more general constraints (i.e. those with smaller *n* and fewer features). Fourthly, following the

[10] This follows from the fact that pairs of candidates can be used to define partial orders over the constraints and the fact that the problem of counting the linear extensions of partial orders is #*P*-Complete.

principle of maximum entropy, the model weights constraints so that their observed number of violations in the training data matches the expected number.

The authors provide case studies using corpus data suggesting that phonological features play a crucial role in generalization. However, Albright (2009) explores feature-based generalization in Hayes and Wilson's model, as well as one based on minimal generalization, and shows that the specific contribution features make to learning remains unclear (CHAPTER 17: DISTINCTIVE FEATURES). This is an interesting class of models, and the phonological biases with the hypothesis space are in many ways appealing. However, as with the biases in Gildea and Jurafsky (1996), formal analysis of their contribution is needed.

5.7 *Learning Harmony Patterns*

Hayes and Wilson (2008) show that when representations are enriched by allowing segments with certain features to project onto tiers (where segments without such features are not projected) (see CHAPTER 14: AUTOSEGMENTS; CHAPTER 105: TIER SEGREGATION), if the algorithm is allowed to search for *n*-gram-like constraints on these additional levels of representation then it is possible to learn long-distance phonotactic constraints (i.e. harmony; see CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS). Hayes and Wilson (2008: 32) conclude that "in controlled comparative simulations, [tiers] makes phonotactic learning possible where it would not otherwise be so." It is, however, critical to bear in mind that this result tells us something about a particular *algorithm*, and not something about the linguistic phenomenon of *harmony* (i.e. a class of languages). Indeed, Heinz (2007, 2010) shows that long-distance phonotactic constraints can be learned without tiers (see below). Furthermore, the tiers that are critical to the success of the algorithm are taken by Hayes and Wilson to be antecedently given, but this does not entail (nor do the authors claim) that they must be antecedently given. Goldsmith and Riggle (forthcoming) offer a strategy for learning long-distance patterns that has many similarities to Hayes and Wilson's approach, but begins with an algorithm from Goldsmith and Xanthos (2009) for "discovering" tiers via unsupervised categorization of the sounds of corpus into vowels and consonants.

Heinz (2007, 2010) shows that phonotactic patterns derived from long-distance agreement patterns (Hansson 2001; Rose and Walker 2004) can be learned without tiers, using the notion of a discontiguous subsequence of length two. This idea is similar to bigram learning where learners keep track of contiguous subsequences of length two. Heinz provides proofs and formal analysis of classes of patterns this algorithm is able to identify in the limit. Unfortunately, the absence of analysis of what classes are learnable by the previously discussed phonotactic learners hinders comparisons of the models.

5.8 *Learning stress patterns*

Stress patterns can be thought of as word-well-formedness conditions, and hence a kind of phonotactic pattern. Since stress typologies are diverse and well established, learning stress patterns has become a popular and challenging proving ground for learning algorithms (CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE; CHAPTER 41: THE REPRESENTATION OF WORD STRESS; CHAPTER 44: THE IAMBIC–TROCHAIC LAW; CHAPTER 57: QUANTITY-SENSITIVITY).

Dresher and Kaye (1990) propose a learning model in the Principles and Parameters framework for learning stress patterns. In this framework, a grammar is a vector of parameters. The learner takes as input a list of words, and for each word, sets parameters as determined by checking whether the word consists of particular properties, called cues. Gillis *et al.* (1995) implement the model with interesting discussion regarding what constitutes an appropriate cue. They only provide input words up to four syllables in length, and demonstrate that the learner succeeded in learning 75 percent of the patterns. Related work includes Gibson and Wexler's (1994) Triggering Learning Algorithm (see also Frank and Kapur 1996 and Niyogi 2006: chapter 3 for discussion).

Goldsmith (1994) and Gupta and Touretzky (1994) investigate how quantity-insensitive stress patterns can be learned using dynamic networks. Although the models differ in their specifics – Goldsmith employs a different updating procedure than Gupta and Touretzky, who use a standard perceptron – these methods achieve a certain level of success in learning the patterns for which data is presented.

Tesar and Smolensky (2000) discuss twelve OT constraints which yield a typology of quantity-sensitive stress patterns. The OT constraints make reference to feet (CHAPTER 40: THE FOOT), which are not part of the learning input. Consequently, another procedure is necessary to parse the learner's input data, so that it can be processed by RCD (the underlying form is assumed to be a string of the right number of unstressed syllables). This procedure is non-trivial, as there may be different parses (i.e. foot assignments) for a given stress pattern. Tesar (1998a) proposes a procedure called *robust interpretive parsing*. To test their system, Tesar and Smolensky hand-selected a test set consisting of 124 languages containing most of the "familiar metrical phenomena" analyzable with their constraints (Tesar and Smolensky 2000: 68). Note, however, that they acknowledge this set is not necessarily representative of the whole typology generated by their constraints. Using robust interpretive parsing, they report that if the initial state of the learner is monostratal – that is, no a priori ranking – then the learner succeeds on about 60 percent of the languages in the test set. When a particular initial constraint hierarchy is adopted, the learner achieves ~97 percent success. So in this case, robust interpretive parsing (mostly) addresses the problem RCD has with hidden structure (for this particular set of test data).

Heinz (2007, 2009) proposes that all phonotactic patterns are neighborhood-distinct, which is a locality condition defined in automata-theoretic terms. It is shown that all but two of 109 descriptions of the world's stress patterns are neighborhood-distinct and that a particular learner that uses this property can learn 100 of these 109 patterns exactly. Although not every pattern can be learned, the patterns acquired in the "failure" cases differ only slightly from the target patterns. Heinz concludes that this particular notion of locality structures the hypothesis space in a way that makes a significant contribution to phonotactic learning.

6 Conclusions

We have argued that learning theory affirms the role of structure as a solution to the problem of generalization, and that there are ideas and methods within learning theory that allow one to measure this structure and the class of languages which have such structure. These tools offer phonologists a way to characterize the contribution various structural properties of phonological patterns can make to learning.

With the exception of a substantial amount of work on learning in Optimality Theory (and Heinz 2010, on phonotactics), it is striking that most proposed learning algorithms have been evaluated only with case studies. Though such studies are suggestive and can be vital in the development of models, in order to know whether a given case study illustrates general properties of a problem we need analytical results that show *why* the algorithm succeeds, what properties of the training sample are critical to success, and how the algorithm maps experience to grammars.

Finally, we have emphasized what we believe to be the most fruitful direction for future research. Phonologists ought to identify properties of phonological patterns that structure the hypothesis space or reduce its size (cf. Heinz 2009; Tesar, forthcoming). This approach works in tandem with, rather than in lieu of, formal analysis.

REFERENCES

Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26. 9–41.

Albright, Adam & Bruce Hayes. 2002. In Mike Maxwell (ed.) *Proceedings of the 6th Meeting of the ACL Special Interest Group on Computational Phonology*, 58–69. Philadelphia: Association for Computational Linguistics.

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119–161.

Angluin, Dana. 1980. Inductive inference of formal languages from positive data. *Information Control* 45. 117–135.

Angluin, Dana. 1982. Inference of reversible languages. *Journal for the Association of Computing Machinery* 29. 741–765.

Angluin, Dana. 1988a. Queries and concept learning. *Machine Learning* 2. 319–342.

Angluin, Dana. 1988b. Identifying languages from stochastic examples. Unpublished ms., Yale University.

Anthony, Martin & Norman Biggs. 1992. *Computational learning theory*. Cambridge: Cambridge University Press.

Bane, Max, Jason Riggle & Morgan Sonderegger. 2010. The VC dimension of constraint-based grammars. *Lingua* 120. 1194–1208.

Barton, G. Edward. 1986. Computational complexity in two-level morphology. *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 53–59. Available (July 2010) at http://aclweb.org/anthology-new/P/P86.

Barton, G. Edward, Robert C. Berwick & Eric Sven Ristad. 1987. *Computational complexity and natural language*. Cambridge, MA: MIT Press.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.

Blumer, Anselm, Andrzej Ehrenfeucht, David Haussler & Manfred K. Warmuth. 1989. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM* 36. 929–965.

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory*. 113–124.

Clark, Alexander & Rémi Eyraud. 2007. Polynomial identification in the limit of substitutable context-free languages. *Journal of Machine Learning Research* 8. 1725–1745.

CMU. 1993. *The Carnegie Mellon University Pronouncing Dictionary* v0.1. Available (July 2010) at www.speech.cs.cmu.edu/cgi-bin/cmudict.

Coetzee, Andries W. 2008. Grammaticality and ungrammaticality in phonology. *Language* 84. 218–257.

Coleman, John & Janet B. Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In John Coleman (ed.) *Proceedings of the 3rd Meeting of the ACL Special Interest Group in Computational Phonology (SIGPHON 3)*, 49–56. Somerset, NJ: Association for Computational Linguistics.

Dresher, B. Elan & Jonathan Kaye. 1990. A computational learning model for metrical phonology. *Cognition* 34. 137–195.

Edlefsen, Matt, Dylan Leeman, Nathan Meyers, Nathaniel Smith, Molly Visscher & David Wellcome. 2008. Deciding Strictly Local (SL) languages. In *Proceedings of the Midstates Conference for Undergraduate Research in Computer Science and Mathematics* 6, 66–75.

Eisner, Jason. 1997a. Efficient generation in primitive Optimality Theory. In *Proceedings of the 35th Annual Meeting of the ACL and 8th Conference of the European Chapter of the Association for Computational Linguistics*, 313–320. Morristown, NJ: Association for Computational Linguistics.

Eisner, Jason. 1997b. What constraints should OT allow? Handout from paper presented at the 71st Annual Meeting of the Linguistic Society of America, Chicago (ROA-204).

Frank, Robert & Shyam Kapur. 1996. On the use of triggers in parameter setting. *Linguistic Inquiry* 27. 623–660.

Gibson, Edward & Kenneth Wexler. 1994. Triggers. *Linguistic Inquiry* 25. 407–454.

Gildea, Daniel & Daniel Jurafsky. 1996. Learning bias and phonological-rule induction. *Computational Linguistics* 22. 497–530.

Gillis, Steven, Gert Durieux & Walter Daelemans. 1995. A computational model of P&P: Dresher & Kaye (1990) revisited. In Maaike Verrips & Frank Wijnen (eds.) *Approaches to parameter setting*, 135–173. Amsterdam: University of Amsterdam.

Gold, E. M. 1967. Language identification in the limit. *Information and Control* 10. 447–474.

Gold, E. M. 1978. Complexity of automata identification from given data. *Information and Control* 37. 302–320.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Oxford & Cambridge, MA: Blackwell.

Goldsmith, John A. 1991. Phonology as an intelligent system. In Donna Jo Napoli & Judy Kegl (eds.) *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman*, 247–267. Hillsdale, NJ: Lawrence Erlbaum.

Goldsmith, John A. 1993a. Harmonic phonology. In Goldsmith (1993b), 221–269.

Goldsmith, John A. 1993b. *The last phonological rule: Reflections on constraints and derivations*. Chicago: University of Chicago Press.

Goldsmith, John A. 1994. A dynamic computational theory of accent systems. In Jennifer Cole & Charles W. Kisseberth (eds.) *Perspectives in phonology*, 1–28. Stanford: CSLI.

Goldsmith, John A. & Jason Riggle. Forthcoming. Information theoretic approaches to phonological structure: The case of Finnish vowel harmony. *Natural Language and Linguistic Theory*.

Goldsmith, John A. & Aris Xanthos. 2009. Learning phonological categories. *Language* 85. 4–38.

Goldwater, Sharon & Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenador, Anders Eriksson & Östen Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–120. Stockholm: Stockholm University.

Gordon, Matthew. 2002. A factorial typology of quantity-insensitive stress. *Natural Language and Linguistic Theory* 20. 491–552.

Graf, Thomas. 2010. Comparing incomparable frameworks: A model theoretic approach to phonology. *University of Pennsylvania Working Papers in Linguistics* 16(1). Available at http://repository.upenn.edu/pwp/vol16/iss1.

Greenberg, Joseph H. 1963. Some universals of grammar, with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.) *Universals of language*, 73–113. Cambridge, MA: MIT Press.

Greenberg, Joseph H. 1978. Initial and final consonant sequences. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 2: *Phonology*, 243–279. Stanford: Stanford University Press.

Griffiths, Thomas L., Charles Kemp & Joshua B. Tenenbaum. 2008. Bayesian models of cognition. In Ron Sun (ed.) *The Cambridge handbook of computational cognitive modeling*, 59–100. Cambridge: Cambridge University Press.

Gupta, Prahlad & David Touretzky. 1994. Connectionist models and linguistic theory: Investigations of stress systems in language. *Cognitive Science* 18. 1–50.

Hansson, Gunnar Ólafur. 2001. Theoretical and typological issues in consonant harmony. Ph.D. dissertation, University of California, Berkeley.

Harrison, Michael A. 1978. *Introduction to formal language theory*. Reading: Addison Wesley.

Haussler, David. 1995. Part 1: Overview of the Probably Approximately Correct (PAC) learning framework. Available (July 2010) at http://citeseer.ist.psu.edu/haussler92part.html.

Haussler, David, Michael Kearns & Robert Schapire. 1992. Bounds on the sample complexity of Bayesian learning using information theory and the VC dimension. Technical Report LRC-91-44.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440.

Heinz, Jeffrey. 2007. The inductive learning of phonotactic patterns. Ph.D. dissertation, University of California, Los Angeles.

Heinz, Jeffrey. 2008. Learning left-to-right and right-to-left iterative languages. In Alexander Clark, François Coste & Laurent Miclet (eds.) *Grammatical inference: Algorithms and applications*, 84–97. Berlin: Springer.

Heinz, Jeffrey. 2009. On the role of locality in learning stress patterns. *Phonology* 26. 303–351.

Heinz, Jeffrey. 2010. Learning long distance phonotactics. *Linguistic Inquiry* 41. 623–661.

Higuera, Colin de la. 1997. Characteristic sets for polynomial grammatical inference. *Machine Learning* 27. 125–138.

Higuera, Colin de la. 2010. *Grammatical inference: Learning automata and grammars*. Cambridge: Cambridge University Press.

Hopcroft, John E., Rajeev Motwani & Jeffrey D. Ullman. 1979. *Introduction to automata theory, languages, and computation*. Boston: Addison-Wesley.

Horning, J. J. 1969. A study of grammatical inference. Ph.D. dissertation, Stanford University.

Idsardi, William J. 1998. Tiberian Hebrew spirantization and phonological derivations. *Linguistic Inquiry* 29. 37–73.

Inkelas, Sharon & Cheryl Zoll. 2005. *Reduplication: Doubling in morphology*. Cambridge: Cambridge University Press.

Jain, Sanjay, Daniel Osherson, James S. Royer & Arun Sharma. 1999. *Systems that learn: An introduction to learning theory*. 2nd edn. Cambridge, MA: MIT Press.

Johnson, C. Douglas. 1972. *Formal aspects of phonological description*. The Hague & Paris: Mouton.

Johnson, Mark. 1984. A discovery procedure for certain phonological rules. In *Proceedings of the 10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics*, 344–347.

Joshi, Aravind K. 1985. Tree-adjoining grammars: How much context sensitivity is required to provide reasonable structural descriptions? In David R. Dowty, Lauri Karttunen & Arnold M. Zwicky (ed.) *Natural language parsing: Psychological, computational, and theoretical perspectives*, 206–250. Cambridge: Cambridge University Press.

Kaplan, Ronald & Martin Kay. 1981. Phonological rules and finite state transducers. Paper presented at the 55th Annual Meeting of the Linguistics Society of America, New York.

Kaplan, Ronald & Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20. 331–378.

Karttunen, Lauri. 1993. Finite-state constraints. In Goldsmith (1993b), 173–194.

Karttunen, Lauri. 1998. The proper treatment of optimality in computational phonology. In *Proceedings of the International Workshop on Finite State Methods in Natural Language Processing*, 1–12. Ankara: Bilkent University.

Kearns, Michael & Umesh Vazirani. 1994. *An introduction to computational learning theory*. Cambridge, MA: MIT Press.

Kearns, Michael, Ming Li, Leonard Pitt & Leslie G. Valiant. 1987. On the learnability of Boolean formulae. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing*, 285–295. New York: ACM Press.

Kobele, Gregory. 2006. Generating copies: An investigation into structural identity in language and grammar. Ph.D. dissertation, University of California, Los Angeles.

Koskenniemi, Kimmo. 1983. *Two-level morphology: A general computational model for word-form recognition and production*. Helsinki: Department of General Linguistics, University of Helsinki.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky. 1990. Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness – theoretical foundations. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, 388–395. Mahwah, NJ: Lawrence Erlbaum.

Littlestone, Nick. 1988. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning* 2. 285–318.

Mairal, Ricardo & Juana Gil (eds.) 2006. *Linguistic universals*. Cambridge: Cambridge University Press.

McNaughton, Robert & Seymour A. Papert. 1971. *Counter-free automata*. Cambridge, MA: MIT Press.

Muggleton, Stephen. 1990. *Inductive acquisition of expert knowledge*. Wokingham: Addison-Wesley.

Niyogi, Partha. 2006. *The computational nature of language learning and evolution*. Cambridge, MA: MIT Press.

Nowak, Martin A., Natalia L. Komarova & Partha Niyogi. 2002. Computational and evolutionary aspects of language. *Nature* 417. 611–617.

Oncina, José, Pedro García & Enrique Vidal. 1993. Learning subsequential transducers for pattern recognition interpretation tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15. 448–458.

Osherson, Daniel, Scott Weinstein & Michael Stob. 1986. *Systems that learn*. Cambridge, MA: MIT Press.

Papadimitriou, Christos H. & Kenneth Steiglitz. 1998. *Combinatorial optimization: Algorithms and complexity*. Mineola, NY: Dover Publications.

Pater, Joe. 2008. Gradual learning and convergence. *Linguistic Inquiry* 39. 334–345.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 36. 999–1035.

Pinker, Steven. 1979. Formal models of language learning. *Cognition* 7. 217–283.

Pitt, Leonard. 1989. Inductive inference, DFAs and computational complexity. In *Proceedings of the International Workshop on Analogical and Inductive Inference*, 18–44. Heidelberg: Springer.

Popper, Karl. 1959. *The logic of scientific discovery*. New York: Basic Books.

Potts, Christopher, Joe Pater, Karin Jesney, Rajesh Bhatt & Michael Becker. 2010. Harmonic Grammar with linear programming: From linear systems to linguistic typology. *Phonology* 27. 77–117.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Riggle, Jason. 2004. Generation, recognition, and learning in finite state Optimality Theory. Ph.D. dissertation, University of California, Los Angeles.

Riggle, Jason. 2009. The complexity of ranking hypotheses in Optimality Theory. *Computational Linguistics* 35. 47–59.

Roark, Brian & Richard Sproat. 2007. *Computational approaches to morphology and syntax*. Oxford: Oxford University Press.

Rogers, Hartley. 1967. *Theory of recursive functions and effective computability*. New York: McGraw Hill.

Rogers, James & Geoffrey K. Pullum. 2007. Aural pattern recognition experiments and the subregular hierarchy. In Marcus Kracht (ed.) *Proceedings of the 10th Mathematics of Language Conference*, 1–7. Los Angeles: University of California, Los Angeles.

Rose, Sharon & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80. 475–531.

Rosenblatt, Frank. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65. 386–408.

Rumelhart, D. E. & J. L. McClelland. 1986. On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart & the PDP Research Group (eds.) *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 2, 216–271. Cambridge, MA: MIT Press.

Shieber, Stuart M. 1985. Evidence against the context-freeness of natural language. *Linguistics and Philosophy* 8. 333–343.

Smolensky, Paul & Géraldine Legendre (eds.) 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, MA: MIT Press.

Sober, Elliot. 2008. *Evidence and evolution*. Cambridge: Cambridge University Press.

Stabler, Edward P. 2009. Computational models of language universals: Expressiveness, learnability and consequences. In Morten H. Christiansen, Chris Collins & Simon Edelman (eds.) *Language universals*, 200–223. Oxford: Oxford University Press.

Tesar, Bruce. 1995. Computational Optimality Theory. Ph.D. dissertation, University of Colorado, Boulder.

Tesar, Bruce. 1997. Multi-recursive constraint demotion. Unpublished ms., Rutgers University.

Tesar, Bruce. 1998a. An iterative strategy for language learning. *Lingua* 104. 131–145.

Tesar, Bruce. 1998b. Error-driven learning in Optimality Theory via the efficient computation of optimal forms. In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis & David Pesetsky (eds.) *Is the best good enough? Optimality and competition in syntax*, 421–436. Cambridge, MA: MIT Press.

Tesar, Bruce. Forthcoming. Learning phonological grammars for output-driven maps. *Papers from the Annual Meeting of the North East Linguistic Society* 39.

Tesar, Bruce & Paul Smolensky. 1993. The learnability of Optimality Theory: An algorithm and some basic complexity results. Unpublished ms., University of Colorado, Boulder.

Tesar, Bruce & Paul Smolensky. 1996. Learnability in Optimality Theory (long version). Technical Report 96:3, Department of Cognitive Science, Johns Hopkins University.

Tesar, Bruce & Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29. 229–268.

Tesar, Bruce & Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Thomas, Wolfgang. 1997. *Languages, automata, and logic*, vol. 3. New York: Springer.

Valiant, Leslie G. 1984. A theory of the learnable. *Communications of the ACM* 27. 1134–1142.

Vapnik, Vladimir. 1998. *Statistical learning theory*. New York: Wiley.

Vapnik, Vladimir & Alexey Chervonenkis. 1971. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications* 16. 264–280.

Zuraw, Kie. 2000. Patterned exceptions in phonology. Ph.D. dissertation, University of California, Los Angeles.

4 Markedness

Elizabeth Hume

1 Introduction

Markedness is one of the most widely used concepts in phonology and other areas of linguistics. The picture is complicated, however, by the fact that the term is used in different ways, as summarized in (1) (see also Haspelmath 2006).

(1) *Markedness usages*

a. *Descriptive markedness*
An abstract relation holding over members of a set of observations displaying asymmetry, such that one subset is unmarked and the other is marked.

b. *Theoretical markedness*
A universal principle or laws that guide language acquisition, loss, inventory structure, processes, rules, etc. toward the unmarked form.

c. *Markedness constraints*
A technical term in Optimality Theory referring to a category of constraints that evaluate the well-formedness of output structures.

As stated in (1c), one usage of markedness appears in Optimality Theory (OT), where the term describes a category of constraints that evaluate output structures. The technical use of markedness in OT is distinct from other usages: "A markedness constraint in OT may produce results related to [the] descriptive or typological sense of markedness, but the formal constraint and the typological observation are two different things" (McCarthy 2002: 15). Markedness constraints are not discussed in this chapter; the reader is referred to CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS, where the topic is addressed in detail.

A more common usage of the term, given in (1a), is descriptive in nature, referring to an abstract relation holding over a set of language observations. Descriptive markedness has served as an analytic tool to categorize sounds and other linguistic elements. It has also been commonly used as an instrument for comparing different language systems leading to the creation of language typologies. This concept of markedness dates back to linguists from the Prague School, most notably Nikolai Trubetzkoy and Roman Jakobson. Trubetzkoy (1939) used the term in order to describe the relations among members of a sound opposition: one member of the opposition bears some property or "mark" that the other member lacks. For example, given the set of consonants [m n b d], the nasal sounds [m n] can be described as being in opposition to the oral sounds [b d] regarding the property *nasal*: [m n] bear the "mark" of nasality, while [b d] do not. Thus, nasal consonants can be referred to as the marked category and oral consonants the unmarked category (see also Jakobson 1932; Jakobson and Pomorska 1990; CHAPTER 11: THE PHONEME).

In contemporary linguistics, the terms "marked" and "unmarked" have come to carry much more meaning than simply "bearing some property," as exemplified in (2) by terms drawn from the literature. (See §5 for discussion of some of these terms, including the noted contradiction regarding perceptual salience.) Thus, describing an observation as *unmarked* is often taken to mean that it is, for example, more frequent, natural, simple, and predictable than the marked observation of the comparison set. The unmarked is also often referred to as the *default* member of a class; that is, it is the member to be assumed, the most basic member of the set, barring further requirements or information.

(2) *Markedness descriptors*

unmarked	*marked*
natural	less natural
normal	less normal
general	specialized
simple	complex
inactive	active
more frequent	less frequent
optimal	less optimal
predictable	unpredictable
acquired earlier	acquired later
more phonetically variable	less phonetically variable
articulatorily simple	articulatorily difficult
perceptually strong	perceptually weak
perceptually weak	perceptually strong
universal	language-specific
ubiquitous	parochial

A further meaning of the term "markedness" is theoretical in nature, referring to a universal principle or laws that guide language toward the unmarked, (1b). The theoretical use of markedness has been at the core of modern phonology and the focus of much debate. It is also the main focus of this chapter. Some key areas of controversy are listed in (3), though it should be pointed out that given the enormity of this topic, it is difficult to do justice in a single chapter to all the debates surrounding markedness. Consequently, the reader is encouraged to consult the references cited throughout and related *Companion* chapters for additional discussion.

(3) *Some key areas of controversy*

a. General approaches to predicting markedness patterns.
b. The relation of markedness to phonological theory.
c. The formal expression of markedness.
d. Criteria for predicting markedness patterns.

2 General approaches to predicting markedness patterns

It is uncontroversial that asymmetrical patterns exist within and across languages. Consider vowel epenthesis, for example (CHAPTER 67: VOWEL EPENTHESIS). It has been widely observed that in some languages, one vowel from the language's inventory is consistently used by speakers to break up ill-formed consonant clusters. In English, this vowel is typically schwa or [ɪ] (CHAPTER 26: SCHWA). For example, as a native English speaker, I pronounce the Polish city name *Gdansk* as [gədænsk], with schwa inserted in the [gd] cluster, rather than [i e a o] or any other English vowel. Thus, we can say that, given the set of English vowels, schwa patterns asymmetrically from other vowels within that set. Expanding our investigation to include other languages, we observe that the epenthetic vowel is [i] in Maltese, [e] in Spanish, and [œ/ø] in French. By defining the set of observations as the epenthetic vowels in English, French, Maltese, and Spanish, we can observe a further asymmetry (among possibly others) having to do with the nature of the vowel selected: in three languages (English, Maltese, Spanish) the epenthetic vowel is unrounded, and in one (French) it is rounded (see §5.5 for additional discussion). Asymmetrical patterns such as these are commonly called markedness observations. That asymmetries such as these exist is not in question; the controversy surrounds how to predict them.

There are two general approaches to this issue. The first draws on markedness itself as the explanation for the asymmetries. In this sense, markedness laws form part of a person's innate knowledge of language (i.e. *competence*; Chomsky 1965, 1986), the position widely adopted by generative phonologists. In this view, markedness forms part of Universal Grammar, and is thus predetermined for speakers of all languages. Further, markedness is a formal issue, in that markedness patterns are predicted from the formalism. For simplicity, I will refer to this as the *markedness-through-formalism* approach.

An alternative, which I will call the *markedness-through-mechanism* approach, attributes markedness patterns to a confluence of factors that interact with grammatical systems, and relate to physical, cognitive, and social mechanisms shared by all humans (e.g. Lass 1975; Stampe 1979; Comrie 1983; Menn 1983; Boersma 1998; Blevins 2004; Hume 2004b; Mielke 2008; Bybee, forthcoming). Comrie (1983), for example, argues that markedness can be explained in terms of human interaction with other humans and with the world, and not as an accidentally inherited or a purely formal property of language. Since independent factors are able to explain the patterns, he argues, there is no need for recourse to Universal Grammar, and markedness as a universal guiding principle need not exist.

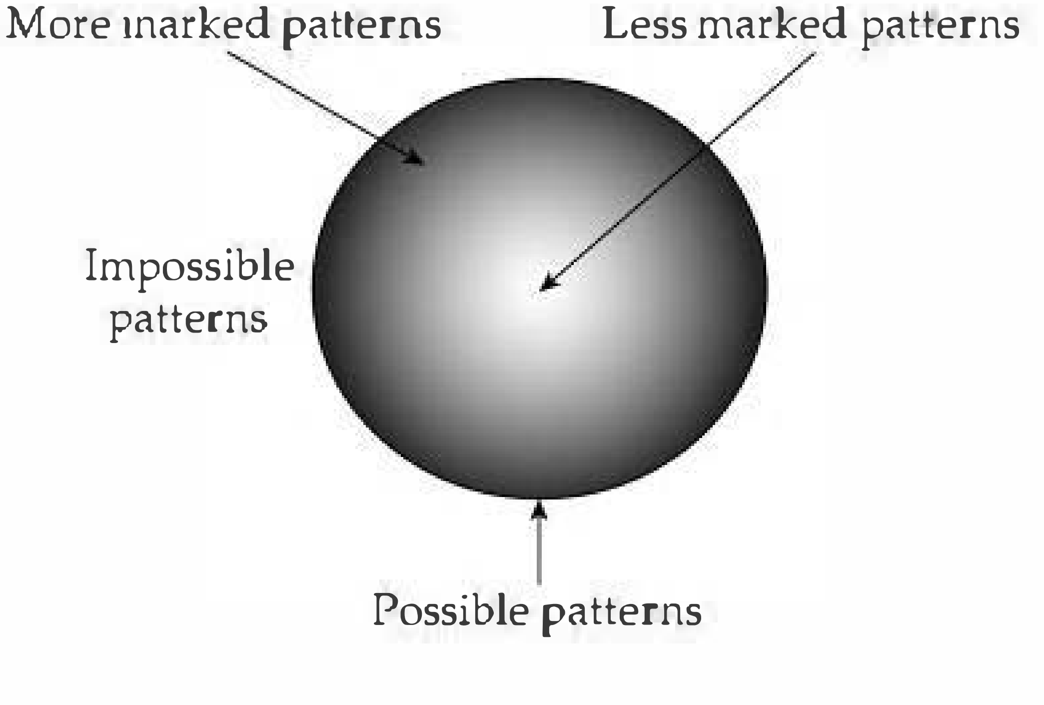
Asymmetrical patterns in syllable structure exemplify the nature of the debate. It has long been observed that there are preferred syllable types cross-linguistically (e.g. Cairns and Feinstein 1982; Vennemann 1988). For example, onsets with sonority rising into the nucleus (e.g. [blV]) are more common than those with sonority falling into the nucleus (e.g. [lbV]) (CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE; CHAPTER 49: SONORITY; CHAPTER 55: ONSETS). Such cross-linguistic preferences for syllable type are commonly described in markedness terms: an onset cluster with rising sonority is less marked than one with falling sonority. Berent *et al.* (2007) tested the claim that listeners have innate knowledge of such patterns. Results from their perception experiments indicate that the asymmetrical patterning of syllable types observed cross-linguistically is reflected in the linguistic behavior of their English listeners: the more an onset cluster is marked (i.e. onsets with sonority falling into the nucleus), the more likely it is to be perceived with an illusory epenthetic vowel (presumably as a means of creating a less marked onset) (see also CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). Their interpretation of the findings is in keeping with the markedness-through-formalism approach to predicting markedness patterns: the results provide evidence that knowledge about the markedness of onset clusters is part of Universal Grammar (UG). Proponents of the markedness-through-mechanism approach would interpret the findings differently, as expressed in Peperkamp's (2007: 634) response to Berent *et al.*'s conclusion: "A possible account of the reported markedness effect that makes no appeal to innate knowledge lies with phonetic differences across onset clusters with varying sonority profiles." In particular, the human perceptual mechanism is responsible for a human's ability to detect phonetic differences, rather than assuming that the individual is born with knowledge about these differences.

It is important to note that it is not the notion of innateness that differentiates these approaches; both assume innateness. Rather, the difference lies in *what* is considered innate. In the markedness-through-mechanism approach, it is the physical, cognitive, and, for some, social mechanisms guiding language acquisition, loss, and usage that are innate. In the markedness-through-formalism approach, on the other hand, it is the formalism from which markedness patterns are derived that is innate. Depending on the particular theory, the formalism may be rooted in the physical and cognitive mechanisms assumed in the former approach.

3 The relation of markedness to phonological theory

Markedness has played an important role in the development of phonological theories; however, approaches have differed with regards to where the domain of explanation for markedness patterns resides. In one approach, markedness observations are directly expressed in a theory of phonology, while in another, they are the basis for a separate theory of markedness. By means of illustration, assume that the sphere in (4) contains all possible sound patterns; impossible patterns reside outside of the sphere. Further, the space within the sphere is graded, with more marked patterns occurring further away from the core.

(4) *Language patterns*



One approach assumes separate theories to account for the set of language patterns in (4): a theory of markedness and a formal theory of phonology (or grammar). Examples include Chomsky and Halle (1968: ch. 9), Kean (1975), Cairns and Feinstein (1982), Mohanan (1993), Calabrese (1995), Steriade (1995), and Boersma (1998). The goal of markedness theory, according to Chomsky and Halle (1968), is to distinguish between more and less natural segments and rules, and to determine the degree of admissibility of a given lexical item. In other words, the domain of markedness theory is internal to the sphere in (4). A theory of grammar serves a different purpose: to distinguish between possible and impossible items, i.e. between elements that do and do not occur within the sphere (see Mohanan 1993 and Calabrese 1995 for related discussion).

This model of markedness was first proposed for phonology in chapter 9 of Chomsky and Halle (1968; *SPE*). Like Jakobson (1971a) and Greenberg (1966), *SPE* treats markedness as a universal principle which guides language acquisition and the formation of phoneme inventories. In addition, it serves as an evaluation metric on the selection of analytic options in the formulation of phonological rules and underlying representations: "children construct grammars to account for the data they are exposed to within the constraints imposed by the formalism, and an evaluation metric selects the simplest possible grammar for the given data" (Mohanan 1992: 639, based on *SPE*). Marked options impose a cost on an analysis, while unmarked ones do not; qualifiers such as *simpler, preferred, optimal,* and *more highly valued* are commonly used to convey the less costly nature of the unmarked option. Steriade (1995: 118) suggests that a theory of markedness should also be able to "document the validity of . . . universal statements" such as feature co-occurrence constraints, "and seek an explanation for their universal status."

A second approach to predicting markedness patterns assumes a single theory. In this view, the goal of an adequate theory of grammar is to be able to predict not only possible grammars (inside *vs.* outside the sphere), but markedness observations as well (gradations within the sphere). Such an approach is seen in, for example, Archangeli (1984), Prince and Smolensky (1993), Calabrese (1995), Rice (1996), de Lacy (2002, 2006), and made explicit in Sagey's (1986: 9) influential dissertation on feature organization:

> It should be possible to represent within [a theory of phonology] any phonological process or form that is a possible human language, and it should be impossible to represent phonological forms and processes that do not exist in human language . . . Another requirement on the theory is that the relative simplicity of describing in the representation each process or form that occurs should reflect its relative naturalness, in the sense of its frequency of occurrence in the languages of the world. That is, more marked forms and processes should be correlated with more marked representations.

As discussed further below, this perspective is fundamental to some theories of underspecification (e.g. Archangeli 1984; CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION) and feature organization (e.g. Clements 1985; Avery and Rice 1989; CHAPTER 27: THE ORGANIZATION OF FEATURES).

4 The formal expression of markedness

The representation of markedness, in a theory of markedness or grammar, typically assumes two components: (i) a formal device denoting whether an element is marked or unmarked, and (ii) a mechanism encoding the relative markedness of the elements in question. Examples of each are given in (5). Implicational statements (§5.2) can be viewed as incorporating both components.

(5) *Representing markedness*

a. *Formal devices*
Feature values (e.g. Jakobson *et al.* 1952; Jakobson 1971a; *SPE*)
Diacritics (e.g. *SPE*; Calabrese 1995)
Distributional statements (e.g. Mohanan 1993)
Redundancy rules (e.g. *SPE*; Archangeli 1984)
Constraints (e.g. Prince and Smolensky 1993; Calabrese 1995; Boersma 1998; Causley 1999; de Lacy 2002, 2006)

b. *Relational mechanisms*
Ordering (e.g. Calabrese 1995; CHAPTER 74: RULE ORDERING)
Constraint ranking (e.g. Prince and Smolensky 1993)
Subset relations (de Lacy 2002, 2006)
Quantitative measures (e.g. *SPE*; Archangeli 1984; Clements 1985; Kiparsky 1985; Avery and Rice 1989).

Given space limitations, the discussion below provides an overview of only some of these components.

In *SPE*, the diacritics *m* and *u* are formal devices assigned to features in order to indicate whether a feature value is intrinsically "marked" or "unmarked," respectively. All phonemes are assigned a value for each feature, thus *m* and *u* values are present in the lexicon. The combination of feature values within a segment determines whether that sound is marked or unmarked. Prior to the application of phonological rules, markedness conventions, as in (6), translate *m* and *u* into the values + and –. For example, the redundancy rule in (6a) states that the unmarked value of the feature [high] for vowels is [+high], while (6b) states that the unmarked value for [back] when coupled with [+low] is [+back]. Importantly, unmarked rules are considered universal statements that cannot be violated, while marked rules can be violated, but add a cost to the grammar; unmarked values do not contribute to the complexity of a grammar.

(6) *Sample marking conventions for vowels*

a. [u high] → [+high]

b. [u back] → [+back] / $\begin{bmatrix} \overline{} \\ \text{+low} \end{bmatrix}$

In *SPE*, the relative markedness among elements in a system is a quantitative issue, in that the complexity of a system is equal to the sum of the marked features of its members. However, as the authors point out (Chomsky and Halle 1968: 409ff.), counting features alone is not sufficient to predict optimal systems; other considerations, such as system symmetry and simplicity, also need to be taken into account.

Correlating markedness with quantity is also central to some theories of underspecification, in that the least marked element is represented with the least amount of theoretical machinery (e.g. Kiparsky 1982; Archangeli 1984; Pulleyblank 1988; Archangeli and Pulleyblank 1989; CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION). Most models of feature geometry adopt a similar approach (Clements 1985; McCarthy 1988; Avery and Rice 1989; Clements and Hume 1995), as exemplified by Clements's (1985) proposal that natural (unmarked) phonological rules are expressed as single operations. As such, an unmarked rule of assimilation involves the addition of a single association line (CHAPTER 81: LOCAL ASSIMILATION), and an unmarked deletion rule is presented by the delinking of a single line (CHAPTER 68: DELETION). In this light, predicting markedness patterns is a representational problem (Rice 1999).

Associating quantity of structure with unmarkedness is generally linked to the view that unmarked feature values are absent in underlying representation (for discussion see, e.g. Archangeli 1984; Steriade 1987, 1995; Clements 1988, 1993; Odden 1992; Rice 1992; Mohanan 1993; Calabrese 1995; CHAPTER 1: UNDERLYING REPRESENTATIONS; CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION; CHAPTER 27: THE ORGANIZATION OF FEATURES). Thus, underlying representations provide the vehicle for determining markedness relations. For example, Paradis and Prunet (1991: 3) argue that "the special status of coronals lies in the fact that they lack specifications for place features in UR" (see also CHAPTER 12: CORONALS). Similarly, Rice (1992: 64) remarks: "Segment structure combined with absence of universally unmarked features at a node . . . indicates markedness relations, with more structure indicating a more marked segment and less structure a less marked segment." This approach can be traced to early writings by Jakobson and colleagues (e.g. Cherry *et al.* 1953; see §5.5), who sought to maximize the efficiency of phoneme systems by minimizing redundant information, a view reflected in *SPE*. These ideas are further developed in Kiparsky (1982), Archangeli (1984), Pulleyblank (1988), and Archangeli and Pulleyblank (1989), and is expressed in the following quote (Kiparsky 1982: 54–55) (see also CHAPTER 8: SONORANTS):

> The theory of grammar will provide a set of universal redundancy rules functionally analogous to the markedness principles of Chomsky and Halle (1968), but formally identical to ordinary phonological rules. In particular, assume that for every feature F there is minimally a rule
>
> (42) [ ] → [α F]
>
> Where α (+ or −) is the "unmarked" value. In addition, other rules may be applicable in specific syntagmatic or paradigmatic contexts. For example, for voicing we may have the rules
>
> (43) a. [ ] → [+voiced]
> b. [+obst] → [−voiced]
>
> putting the unmarked value as [−voiced] for obstruents and [+voiced] elsewhere. We now say that voiceless obstruents and voiced sonorants are represented as [0 voiced], that is, unspecified for voicing, and that their respective specifications for voicing are filled in by the application of rule (43). This much is quite in the spirit of traditional markedness theory.

Critical discussion of the connection between underspecification and markedness can be found in e.g. Mohanan (1993) and Steriade (1995). Among the critiques, Steriade (1995: 199–200) questions:

> whether the asymmetric distribution between the marked and the unmarked value justifies eliminating the unmarked value from underlying structure. Is underlying privativity the faithful representation of markedness facts? . . . Should the unmarked value be represented at all on the surface?

In response to these questions, she suggests that there are segments in most languages that start out and end up without features, such as place of articulation in the case of the laryngeal consonants [h] and [ʔ] and schwa-like vowels.

A further issue concerns the assumption that only one member of a class can be unmarked and thus least specified. For example, Kiparsky (1985) argues that coronal nasals are unmarked for place in Catalan, since only coronals assimilate to the place of articulation of any following consonant (Mascaró 1976). Rice (2007), however, points out that this approach is untenable, a claim supported by evidence in Hume (2003) showing that virtually any combination of dorsal, labial, and coronal can be unmarked, given their patterning as targets of assimilation, a common markedness diagnostic (see §5).

In the 1990s, the shift in focus from underlying to surface representations further impacted the use of underspecification as a means of expressing markedness. Instead, surface-oriented devices were developed. Calabrese (1995: 374) proposes the use of marking statements, similar to the marking conventions of *SPE*, though formalized as constraints rather than rules. He argues that "the structure of inventories is determined by restrictions on phonological segments in the form of constraints on pairs of features." Co-occurrence restrictions occur within marking statements, as shown in (7). The property of being marked is encoded by underlining of the relevant feature in the statement, as is the case for value β of feature G in combination with another feature value αF, where α and β range over + and –.

(7) *Marking statement* (Calabrese 1995)

$[\alpha F, \underline{\beta G}]$

Constraints are also central to accounts of markedness developed within the framework of Optimality Theory (e.g. Prince and Smolensky 1993; Steriade 1995; Hamilton 1996; Lombardi 1997; Causley 1999; de Lacy 2002, 2006; Hayes and Steriade 2004). For example, Hayes and Steriade (2004: 1) propose that "markedness laws characterising the typology of sound systems play a role, as grammatical constraints, in the linguistic competence of individual speakers."

In some constraint-based approaches, degrees of markedness are expressed through formal devices that organize constraints hierarchically. Calabrese, for instance, draws on constraint ordering to create a hierarchy of markedness statements. In some OT approaches, markedness relations are expressed through the harmonic ordering of members of a correlational set which imposes a universally fixed order on a set of constraints (Prince and Smolensky 1993). For example, as shown in (8), it is claimed that the harmonic ordering for place of articulation expresses the markedness relations provided by UG, such that dorsal is more marked than labial, which is more marked than coronal, and so forth. The corresponding grammatical constraints and ranking appear in (8b). Despite the assumed universality of such rankings, this approach also provides a formal means for describing the observation that a given feature need not be unmarked in all languages: an additional constraint may dominate the fixed ranking, thus having the effect of overruling the unmarked status of a lower-ranked constraint (see e.g. Lombardi 1997 for an analysis along these lines; also CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION).

(8) a. *Harmonic ordering for place of articulation* (> = "is less marked than")
glottal > coronal > labial > dorsal

b. *Corresponding fixed universal ranking* (>> = "is ranked above")
*dorsal >> *labial >> *coronal >> *glottal

De Lacy's (2002, 2006) OT-based theory of markedness makes use of markedness hierarchies tied to sets of constraints. Place of articulation features are represented by the hierarchy in (9a), which projects onto the output constraint sets in (9b).

(9) a. *Markedness hierarchy*
(hierarchy enclosed by | |; > = "is more marked than")
| dorsal > labial > coronal > glottal |

b. *Each hierarchy is related to a set of constraints:*

i. *[dorsal]
Assign a violation for each [dorsal] feature.

ii. *[dorsal, labial]
Assign a violation for each [dorsal] and each [labial] feature.

iii. *[dorsal, labial, coronal]
Assign a violation for each [dorsal], each [labial], and each [coronal] feature.

iv. *[dorsal, labial, coronal, glottal]
Assign a violation for each [dorsal], each [labial], each [coronal], and each [glottal] feature.

Unlike (8b), the ranking of the constraint sets in (9b) is not fixed. Instead, de Lacy proposes that markedness patterns are generated by virtue of the subset relations present in each set of constraints. For example, since in (9) the marked feature [dorsal] is listed in each constraint of the entire set, a violation of any of the constraints will necessarily target [dorsal]. Conversely, the absence of [glottal] in all constraints but *[dorsal, labial, coronal, glottal] preserves the feature, except when all other feature types are involved. In other words, [dorsal] is least preferred, while [glottal] is most preferred. I refer the reader to de Lacy (2006) for discussion concerning the predictive power of the approaches represented in (8) and (9).

5 Criteria for predicting markedness patterns

Determining what constitutes reliable evidence for markedness relations remains an area of controversy. One issue concerns whether one can actually diagnose markedness: are markedness diagnostics to be treated as "criterial for determining asymmetry or as merely correlative" Battistella (1996: 14)? Generative phonological approaches have tended to view markedness diagnostics as criterial.

Others view such diagnostics as correlative, to the extent that they reveal reliable statistical tendencies (e.g. Lass 1975; Menn 1983; Ohala 1990; Bybee 2001; Hume 2004a, 2004b; Haspelmath 2006; Mielke 2008).

Controversy also surrounds the issue of whether evidence for markedness comes from synchrony or diachrony (CHAPTER 93: SOUND CHANGE). In theories of generative phonology, the status of synchronic patterns is considered paramount, given the fundamental premise that such observations provide a window into an individual's innate knowledge of language. Further, since the individual does not have direct access to knowledge of changes from the past, only synchrony can provide valid sources of evidence for markedness patterns (see e.g. Rice 1999; de Lacy 2002, 2006). Conversely, Blevins (2004: 20) argues that

> there is no clear role for markedness within synchronic phonology. Absolute universals and universal tendencies in sound patterns emerge from general pathways of language change, and have no independent status in the grammar . . . most proposed universals and their counterexamples have straightforward diachronic explanations.

(See also Mielke 2008; CHAPTER 17: DISTINCTIVE FEATURES.)

Another area of debate concerns which criteria are appropriate for determining the markedness value of a given element. The types of evidence used to support claims about phonological markedness include those in (10).

(10) *Types of evidence for predicting markedness patterns*

a. Acquisition.
b. Phonological patterns, e.g. alternations, inventory structure, distribution, implicational relations.
c. Phonetics, e.g. quality of acoustic/auditory cues to the identification of a pattern, articulatory difficulty in producing a pattern.
d. Usage factors, e.g. statistical patterns.
e. Cognitive factors, e.g. information content, entropy.

In the remainder of this section we take a closer look at these criteria.

5.1 *Language acquisition*

Since Jakobson's seminal papers on language acquisition (1971a, 1971b), the view that there are universal principles, such as markedness, that guide the order in which a child learns language has been highly influential (CHAPTER 101: THE INTERPRETATION OF PHONOLOGICAL PATTERNS IN FIRST LANGUAGE ACQUISITION). Central to this approach is the proposal that acquisition of a marked sound category, defined in terms of distinctive features, presupposes the acquisition of the corresponding unmarked category (see implicational relations: §5.2). According to Jakobson (1971b: 11), the order in which children acquire sound categories "corresponds exactly to the general laws of irreversible solidarity (implication) which govern the synchrony of the languages of the world." For example, under the assumption that dorsal place of articulation is more marked than coronal, the acquisition of dorsals presupposes that of coronals. Jakobson (1971b: 7) acknowledges in places that the term *general laws* might be "more prudently formulated" as *tendencies*, though his proposed orders of acquisition have typically been interpreted as absolutes: all children acquire sounds in the same order, regardless of the language that they are learning.

Controversy has surrounded at least two issues arising from the proposed link between acquisition and markedness: (i) the extent to which there is a universal order of acquisition such that the marked implies the unmarked; and (ii) whether acquisition data should be used as evidence for claims regarding markedness.

With regard to the first point, experimental evidence indicates that there is considerable variability in the order of acquisition across languages. For example, Morrisette *et al.* (2003) tested the claim that dorsal place of articulation is more marked than coronal by examining inventory structure and substitution patterns from 211 English-learning children. They found that

> the full range of logical possibilities was found to occur with regard to inventory structure; that is, some children included both coronals and dorsals in their inventories, others included coronals but not dorsals, and yet others included dorsals but not coronals. In terms of the children's substitution patterns, dorsals were replaced by coronals . . . in a large proportion of the cases. However, a small proportion of the children replaced coronals with dorsals. (Morrisette *et al.* 2003: 351)

(See also Menn 1983; Vihman 1993; Beckman *et al.* 2003.)

Beckman *et al.*'s (2003) study of the acquisition of place of articulation also challenges the view that there is a universal order of acquisition for place of articulation. Japanese-learning children made more than twice as many "backing" errors for /t/ (i.e. /t/ pronounced as /k/) as they made "fronting" errors for /k/ (/k/ pronounced as /t/), which runs counter to the claim that back consonants like /k/ are universally marked and likely to be replaced by front consonants like /t/.

However, Beckman *et al.* (2003) also suggest that some universals may exist, "if we take the term 'universal' to mean a strong numerical tendency rather than an absolute rule." They suggest that the best example involves laryngeal features, thus supporting Jakobson's claim that voiceless unaspirated stops are mastered before aspirated or voiced stops. Drawing on evidence from Kewley-Port and Preston (1974), they propose that this universal of acquisition is phonetically grounded in the relative diffiiculty of satisfying aerodynamic requirements for the different stop types.

A second issue concerns whether or not observations from acquisition, including second language acquisition, should be used as evidence for a theory of markedness at all. In other words, if there is empirical evidence that children acquire one class of sounds prior to another, should these observations be used as evidence for or against a theory of markedness? There has been less of an overt debate regarding this issue; rather, authors typically assume one position or the other. For example, Calabrese (1995) draws on the order of acquisition of segments (and segment loss) in his markedness theory to establish the order of marking statements. Conversely, de Lacy (2006) assumes that acquisition data do not in and of themselves provide evidence for a markedness pattern. He takes issue with the findings of Beckman *et al.* (2003), who argue that velar /k/ patterns are unmarked among Japanese-learning children, claiming that in the absence of synchronic alternations in the language that also attest to the unmarked status of velars, the apparent unmarked behavior of velars could be explained by phonetic factors. A crucial difference between de Lacy's and Beckman *et al.*'s approaches is that in the former, markedness constitutes a fundamental part of an individual's grammatical competence, independent of performance factors. In this view, any observations that might be explained by factors external to the grammar and not supported by synchronic processes do not constitute empirical evidence for a formal theory of markedness. Conversely, in the latter approach, there is no strict division between competence and performance; linguistic knowledge comprises and is influenced by both types of information.

5.2 *Phonological patterns*

From the earliest discussion of markedness in Trubetzkoy (1939) to more recent research on the topic, the asymmetrical patterning of features and sounds in inventories and phonological processes has served as the basis for predicting markedness relations. Indeed, in Rice's (1999) evaluation of markedness criteria, she concludes that the strongest and most compelling arguments come from phonological processes (see also de Lacy 2006). Thus, when comparing features within a class, if one feature patterns asymmetrically with respect to the others, it is this feature that is deemed the unmarked member of the relation. Consider an example from Yoruba. There are three tones in the language's tonal inventory (High, Mid, Low) (CHAPTER 45: THE REPRESENTATION OF TONE). The Mid tone, unlike High and Low, never appears in the structural descriptions or changes of phonological rules. Mid thus patterns asymmetrically with respect to other members of the tonal class and, as a result, is considered the unmarked member (Akinlabi 1985). The following discussion provides an overview of the types of evidence used in diagnosing markedness, but cannot do justice to the vast literature that has been devoted to this topic. Additional works that discuss the different types of evidence include Battistella (1990, 1996); Rice (1999, 2007); de Lacy (2002, 2006); Hume (2004b, 2006, 2008); Haspelmath (2006).

In terms of phonological patterning, it is widely assumed that, to the extent that there is asymmetry, the unmarked member of an opposition patterns differently than other members in being the *target* of phonological processes, such as reduction, deletion, and assimilation (though cf. de Lacy 2006). With regard to assimilation (CHAPTER 81: LOCAL ASSIMILATION; CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS), for example, Rice (1999: 4) points out:

> the unmarked pole of an opposition is lost or obscured, with the marked pole remaining . . . In assimilation, the marked features within a class are active . . . the unmarked features, on the other hand, are passive, or inert . . . overridden by other features.

Put another way, marked features are assumed to resist modification, while unmarked features are subject to change. Rice (1999, 2007) labels this criterion "submergence of the unmarked." Recall the Catalan example from above, in which only coronals assimilate to the place of articulation of a following consonant (Mascaró 1976); in this case, coronal is considered unmarked. A more complex yet frequently cited example comes from place assimilation in Korean, as shown in (11), where a final obstruent stop assimilates in place to a following consonant, with the following restrictions. A morpheme-final coronal assimilates to a following dorsal or labial consonant (11a). A morpheme-final labial also assimilates to a following dorsal, but fails to assimilate to a following coronal, as in (11b). As the examples in (11c) show, a final dorsal consonant does not assimilate to either a following labial or coronal consonant. According to the view that markedness is correlated with resistance to modification, the dorsal consonant is considered most marked, followed by labial, then coronal (Iverson and Kim 1987), i.e. dorsal > labial > coronal.

(11) *Korean place assimilation*

a.	/mit+ko/	[mikk'o]		'believe and'
	/mit^h+pota/	[mipp'ota]		'more than the bottom'
b.	/ip+ko/	[ikk'o]		'wear and'
	/nop+ta/	[nopt'a]	*[nott'a]	'high'
c.	/nok+ta/	[nokt'a]	*[nott'a]	'melt'
	/kuk+pota/	[kukp'ota]	*[kupp'ota]	'more than soup'

The *output* of certain phonological processes including epenthesis and neutralization has also been drawn on as evidence for identifying the unmarked member of a class. The output of *neutralization* (CHAPTER 80: MERGERS AND NEUTRALIZATION) is one of the original diagnostics for markedness, proposed in Trubetzkoy (1939): the result of neutralization does not bear the mark of the relevant property, i.e. it is unmarked. For example, in languages with final devoicing such as German, Maltese, Polish, and Russian, the contrast between voiced and voiceless consonants is arguably neutralized in word-final or coda position; only one member of the opposition survives, the voiceless consonant, and thus this member is considered unmarked. There is lack of consensus, however, regarding the use of the output of neutralization as evidence for the unmarked. Croft (2003), for example, claims that there is no uniformity across languages concerning which value is to be considered unmarked (see also de Lacy 2002, 2006; Vaux and Samuels 2005). And for epenthesis, de Lacy (2002, 2006) considers consonant, but not vowel, epenthesis to be a valid criterion for markedness.

Distributional evidence has also long been considered a criterion for determining markedness relations. Battistella (1990) distinguishes between paradigmatic distribution and syntagmatic distribution. With regard to the former, it is claimed that the unmarked category of a contrastive pair is used to distinguish more words than the marked category (see also Waugh and Lafford 1994). As for syntagmatic distribution, it is widely held that the unmarked segment (or feature) is more widely distributed than its marked counterpart (see, e.g. Trubetzkoy 1939; Hockett 1955; Greenberg 1966; Battistella 1990; Stemberger 1992; for related discussion, see Rice 1999). In fact, Hockett (1955: 166) proposes that distributional differences underlie the unmarkedness (simplicity, in his terms) associated with the result of neutralization (see also Trubetzkoy 1939; Hume 2004b). In this light, the view that voiceless obstruents in languages such as Russian are simpler (unmarked) as compared to their voiced counterparts is explained in terms of their wider distribution: the voiceless member of the contrast occurs in non-final contexts, as does the voiced member, but also in word-final or coda positions.

One of the original sources of evidence for markedness comes from *implicational relations*, as noted in §5.1: the presence of *x* implies the presence of *y* (Trubetzkoy

1939; Greenberg 1966; Jakobson 1971a). Jakobson (1971a) treats implicational relations as a cornerstone of markedness, in that they state a compulsory connection between two related properties of language; for example, the occurrence of nasal vowels implies that of nasal consonants. Implicational relations have been claimed to reveal markedness patterns in sound change, acquisition, and inventories. Thus, in sound change, the loss of a marked category implies the loss of the corresponding unmarked one (Bailey 1973). Similarly, in child language acquisition, the acquisition of a marked category implies prior acquisition of the corresponding unmarked category (Jakobson 1971a).

Yet it has been in phoneme inventories that implicational relations have been most widely cited as evidence for markedness patterns (see e.g. Chomsky and Halle 1968; Kean 1975; Lass 1975; Cairns and Feinstein 1982; Calabrese 1995; Hamilton 1996; Causley 1999; de Lacy 2002, 2006). For example, Kean (1975) claims that if a language has voiced obstruents, it also has voiceless obstruents. Conversely, Rice (1999) considers implication problematic as evidence for markedness relations given the assumption that language learners do not have access to implicational relationships between segments in consonant and vowel inventories. Rather than being a part of an individual's knowledge of language, implication in Rice's view is a consequence of properties relating to phonetics and usage.

Interestingly, Lass (1975: 501), who opposes the claim that markedness observations provide evidence for innate properties of language, nonetheless views implicational relations as relevant for a theory of "universal phonetics." He suggests that there is a hierarchy made up of

> necessary choices which appear to be universal, and a larger (unordered) set of contingent choices which any language can make after it has made the non-contingent ones. This reflects such facts as the apparent lack of languages without vowels, the lack of languages with nasalized vowels but no oral ones, or glottalized consonant but no non-glottalized one . . . and so on. . . . Under this view we return to something very like the Praguian notion of "marking," in which (at least in some cases) the relation U : M presupposes not "optimalness" or "simplicity" vs. "non-optimalness" or "complexity," but rather (irreflexive) implication: If M, then U; but not vice versa.

Structural asymmetries in phoneme inventories have also been used to identify the unmarked member of the inventory (e.g. Rice and Avery 1993; Rice and Causley 1998; Causley 1999). Building on insights from Trubetzkoy (1939) and Jakobson (see Battistella 1990), Rice and Causley argue that non-contrastive features are inactive, and that such inactivity signals the unmarked status of features and segments. To illustrate, compare two hypothetical sound systems in (12). System A is comprised of six segments; we may assume that each of the following pairs are distinguished by some feature [F]: /a – d/, /b – e/, and /c – f/. System B is made up of five segments; like system A, feature [F] is contrastive for /b – e/ and /c – f/, but not for segment /a/.

(12) *Structural asymmetries*

System A			*System B*		
a	b	c	a	b	c
d	e	f		e	f

Dyck (1995) proposes that phonological patterns occur in which the /a/-like segment in system A triggers processes, whereas the /a/-like segment in system B does not (see the discussion in Rice 2007). Further, there is some support for the role of contrast in predicting markedness in vowel place specifications. Causley (1999: 76–77) argues that vowel systems that

> lack a central vowel have a front or back vowel patterning as unmarked. When a language has a front–central–back contrast, the central vowel is chosen as the least marked. Thus the markedness status of different vowel specifications is linked to the inventory.

Using contrast as a predictor of markedness values is not consistently reliable, however. Rice (2007) presents a number of cases showing variation in the segment that emerges as unmarked among languages with similar systems of contrast.

5.3 *Phonetic properties*

There is general consensus that phonetic factors play a role in predicting markedness patterns, a view dating back to the earliest writings on the topic (e.g. Trubetzkoy 1939; Jakobson and Halle 1956; Greenberg 1966; Chomsky and Halle 1968). Acceptance of this view is seemingly independent of whether or not one associates markedness with Universal Grammar. Hayes and Steriade (2004) note that universal markedness laws are rooted in phonetic knowledge. Consistent with this view, Mohanan (1993) assumes that phonological patterns are influenced by three phonetic principles: minimize the number of articulatory gestures; minimize the deviation of articulatory gestures from configurations of least effort; maintain the phonetic distinctions that distinguish between different words. Similarly, Boersma (2000: 18) notes that markedness, in the sense of cross-linguistic rarity, "emerges from an interaction between principles of articulatory effort and perceptual confusion, and [are] not encoded directly in our phonological language device." Further, Bybee (2001: 14) proposes that the "underlying explanation for sound changes that create markedness relations is phonetic in nature." A more nuanced approach appears in de Lacy (2002, 2006: 1), where two theories of markedness are assumed: a performance theory and a competence theory. "Markedness is part of grammatical Competence (I-language). Markedness in Competence is distinct from sometimes apparently similar Performance-related phenomena." While asymmetrical patterns are assumed to exist in both competence markedness and performance markedness, according to de Lacy their explanations lie in different domains: the former are explained in terms of Universal Grammar, and the latter by factors external to the grammar, such as those relating to phonetics and usage.

The remainder of this section focuses on the phonetic factors most commonly used as criteria for identifying markedness values, as listed in (13).

(13) *Phonetic factors*

a. phonetic variability in production
b. articulatory simplicity
c. perceptual distinctiveness

5.3.1 *Phonetic variability in production*

The unmarked member of an opposition has been claimed to be produced with greater phonetic variability than the marked counterpart, a view first attributed to Trubetzkoy (1939). Similarly, Greenberg (1966) considers greater allophonic variability to be a criterion of the unmarked category, and Rice (1999: 19) suggests that fixedness of phonetic realization correlates with markedness such that the marked pole is more clearly defined: marked members show little variation, while unmarked members show more.

5.3.2 *Articulatory factors*

The unmarked member of an opposition is widely considered to be easier and less complex than the marked counterpart in terms of production. Calabrese (1995: 376), for example, states that the less marked nature of certain feature combinations can be explained by ease of articulation and perceptual saliency: for instance, he claims that

> the [+continuant, −strident] coronal fricatives [θ, ð] are complex because of the articulatory adjustments needed to maintain absence of stridency in fricatives. Still other combinations are phonologically simple. For example, the combination [+continuant, +strident] is simple, since stridency is a natural consequence of the type of constriction found in fricatives.

Similar assumptions hold for sequences of sounds. Along these lines, Hamilton (1996: ii) proposes that

> each phonotactic constraint is phonetically grounded. Unmarked clusters correspond to structures which are gesturally and/or perceptually simple and marked structures are gesturally and/or perceptually complex.

For related discussion, see Archangeli and Pulleyblank (1994), Jun (1995), and Boersma (1998), among others.

The use of articulatory simplicity to predict markedness patterns is not without issues. For example, despite wide acknowledgment in the phonological literature that such factors play an important role, the actual phonetic evidence showing *x* to be simpler than *y* is often not rigorously demonstrated. Further, while it may seem intuitively evident in some cases that two elements are appreciatively different in terms of articulatory complexity (e.g. [p] *vs.* [pʕ]), gradations of difficulty are less evident among more similar elements (e.g. [p] *vs.* [t] *vs.* [k]), particularly for adult speakers with years of experience producing the sounds (see Stampe 1979: 10 for related discussion). An additional issue concerns the interaction between articulatory and perceptual factors, a point returned to in the following section.

5.3.3 *Perceptual distinctiveness*

The perceptual distinctiveness of sounds and strings of speech has also been drawn on to predict observed asymmetries in phonological systems (e.g. Trubetzkoy 1939; Jakobson and Waugh 1987; Lindblom 1990; Flemming 1995; Jun 1995; Steriade 1995; Boersma 1998; Hume 1999, 2004a; Hume and Johnson 2001; Blevins 2004; Bybee, forthcoming; CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). It is clear that distinctiveness is important for the correct identification of sounds and strings of speech (e.g. Ohala 1981; Kawasaki 1982; Lindblom 1990). In this regard, Kawasaki (1982) proposes that sharper changes in the speech signal serve to increase the salience of cues in the portion of the signal where the modulation takes place: the greater the magnitude of the modulation, the better a given signal is detected. Interestingly, this proposal relates to claims regarding phonological markedness in two seemingly conflicting ways.

The first concerns the observation that patterns with *larger modulations* tend to recur across languages, since they are more resistant to change than those with smaller modulations. In this case greater syntagmatic distinctiveness correlates with unmarkedness, an interpretation consistent with the common assumption that the CV syllable is universally preferred over other syllable types, e.g. V, CVC, VC (Cairns and Feinstein 1982; Clements and Keyser 1983). Paradigmatically, strong perceptual distinctiveness has also been proposed as the explanation for recurring phonological inventories. Liljencrants and Lindblom (1972) propose that the structure of vowel systems is determined largely by the principle of maximal perceptual contrast, thus favoring systems in which vowels are perceptually distinct, e.g. [i a u] (see Flemming 1995). Perceptual distinctiveness has also been drawn on extensively to account for asymmetries in phonological patterns. As Hamilton (1996: 12ff.) observes, following Jun (1995) and Steriade (1995),

> Simple percepts are unmarked and complex percepts are marked . . . simple speech sounds in perceptual terms are those with robust spectral cues . . . The empirical correlate to this is that perceptually robust speech sounds are highly valued cross-linguistically, while languages only grudgingly elaborate perceptually opaque structures.

The flip side of the coin is that structures with *small modulations* due to weak acoustic/auditory cues will tend to be modified more often than those with larger modulations (Kawasaki 1982). As a result, sounds lacking distinctiveness are subject to phonological processes such as assimilation, reduction, and deletion to a greater degree than sounds with robust cues (see e.g. Kohler 1990; Hura *et al.* 1992; Jun 1995; Steriade 1997; Boersma 1998; Hayes *et al.* 2004). Thus weak perceptual distinctiveness is also correlated with being unmarked. This interpretation is consistent with the claim that the target of phonological processes such as deletion is the unmarked member of the comparison set.

The use of perceptual distinctiveness thus appears contradictory as a markedness diagnostic: unmarkedness is associated with *strong* perceptual distinctiveness as well as *weak* perceptual distinctiveness. Consequently, it is seemingly impossible to predict a priori whether a sound is marked or unmarked given its salience. Hume (2006), and Hume and Mailhot (forthcoming) propose that these observations are consistent with a model of markedness conceptualized in terms of entropy, a measure of the uncertainty in a system (Shannon and Weaver 1949); see §5.5 for discussion.

A further issue concerns the interaction between production and perceptual demands, and the role of these interactions in predicting markedness observations. Given the complexity of this issue, it is understandable that the topic has received less attention in the literature than the roles of perception and production in markedness independently. The subject is addressed in considerable detail in Boersma (1998), who proposes a production theory of markedness and a perception theory of markedness. He observes (Boersma 1998: 17) that production and perceptual constraints can make conflicting predictions regarding markedness, citing fricatives as an example:

> In gestural terms, fricatives are complex, since the correct spatial relationship between the active and passive articulators must be very precisely controlled in order to maintain turbulent airflow . . . At the same time, fricatives are spectrally very robust. Their random high frequency energy patterns are very distinct from the spectral properties of the other manners of articulation, and also act as a very salient cue to place of articulation.

To resolve such conflicts, Boersma claims (1998: 17) that languages adopt different strategies:

> They may follow the [production theory of markedness] and elaborate gesturally harmonic features, or follow the [perception theory of markedness] and elaborate perceptually harmonic features, or opt for some combination of the two. Therefore in the case of fricatives some languages have more stops than fricatives while others (including English) have fricatives which lack stop counterparts.

5.4 Usage

The most common usage factor cited in relation to markedness is the statistical frequency with which a particular pattern occurs, a criterion dating back to Zipf (1932), Trubetzkoy (1939), Hockett (1955), and Greenberg (1966), among others. The claim is that the unmarked is more frequent cross-linguistically than the marked (CHAPTER 90: FREQUENCY EFFECTS). Despite its long history, the relation between frequency and markedness remains controversial. Three of the main issues are given in (14).

(14) *Issues regarding frequency and markedness*

a. whether or not frequency should play a role in predicting markedness patterns;
b. to the extent that frequency is used as a predictor of markedness, whether frequency should be calculated across all languages, within a single language, or involve both calculations; and
c. what type of frequency measure is relevant.

Trubetzkoy (1939), inspired by Zipf (1932), drew a link between unmarked status and frequency: since the unmarked member of an opposition emerges in cases of neutralization, the distributional asymmetry would contribute to the unmarked sound's greater frequency over the marked sound (see also Hume 2004b). For Jakobson, frequency was foundational to the use of markedness in modern linguistics, reflected in the frequency of a sound across languages, and through implicational relations. This view continues to be commonly accepted today, though not uniformly so. Hume and Tserdanelis (2002) point to the higher frequency of labial place in Sri Lankan Portuguese Creole nasals as support for the unmarkedness of the labial nasal in that language. Rice (1999), however, considers a pattern's frequency to be a consequence of emergent properties (i.e. factors influencing language change), rather than a diagnostic for markedness (see also de Lacy 2002, 2006).

Nonetheless, the link between frequency and markedness is widely attested in the generative phonology literature. Citing cross-linguistic frequency as evidence for markedness, Kean (1975) states that since almost all languages have [t a] but fewer have [kp y], [t a] are unmarked and [kp y] are marked. Further, as noted in §4, Sagey (1986: 9) claims that a requirement of a theory of phonology

> is that the relative simplicity of describing in the representation each process or form that occurs should reflect its relative naturalness, in the sense of its frequency of occurrence in the languages of the world. That is, more marked forms and processes should be correlated with more marked representations.

Similarly, Paradis and Prunet (1991: 10) point to three types of statistical patterns which, they argue, support the view of the feature coronal being unmarked. As listed in (15), the evidence comes from cross-linguistic measures (typological frequency) as well as language-specific ones (inventory frequency, token frequency).

(15) *Evidence from frequency for the unmarked status of coronal* (Paradis and Prunet 1991: 10)

a. Inventory frequency: the number of coronals in the consonant inventory of a given language.
b. Typological frequency: the number of coronals attested in a universal phonemic inventory.
c. Occurrence frequency (i.e. token frequency): the number of times coronals are produced in a representative speech corpus.

Paradis and Prunet's broad use of frequency measures relates to another area of controversy: whether frequency should be calculated across all languages, within a single language, or both. Trubetzkoy (1939) considered the frequency of one member of an opposition within a language as a signal of unmarked status. Greenberg (1966) used statistical methods to claim that the frequencies with which phonemes show up in languages are the expression of a universal tendency to avoid marked phonemes; unmarked phonemes have higher cross-linguistic frequencies. This view is also reflected in more recent literature; Hamilton (1996: 5–6) states that "unmarked features are more frequent than marked features . . . features with a wider cross-linguistic distribution also occur at higher frequencies language-internally." Yet Meier (1999) challenges the view that cross-linguistic frequency is a valid criterion for markedness, arguing that "presumably marked values (e.g. glottality in stop consonants) have a very different distribution in different languages" (cited in Elšík and Matras 2006: 16). In this view, frequency is best calculated on a language-specific basis (see also Hume 2006).

The use of statistical measures is complicated by a number of issues. One has to do with what the appropriate measure is, e.g. type or token frequency. Type frequencies are calculated over a lexicon, for example, how frequently a phoneme occurs in the dictionary of a language; see for example Trubetzkoy (1939). Token frequency is calculated over a written or spoken corpus, for example, how frequently a phoneme occurs in a corpus of natural speech. Type frequencies generally provide more information about the structure of a language, whereas token frequency reveals patterns of usage. The extent to which the two properties affect phonological patterns differently remains an empirical question (see e.g. Munson 2000 for related discussion). A further issue concerns which elements and levels of language frequency should be measured over, for example, features, segments, syllables, words. Does frequency affect all elements of language uniformly? If not, what are the empirical consequences for a language's phonological system? A further issue concerns the interaction of statistical patterns with other factors such as production, perception, word structure, and word similarity. For additional discussion see, for example, Bybee (2001), Phillips (2006), and CHAPTER 90: FREQUENCY EFFECTS.

5.5 *Entropy and information content*

The concepts of entropy and information content have been connected to markedness in at least two ways: as a foundational concept for the dichotomous view of the unmarked and marked, and as a predictor of markedness patterns. Information content and the more widely used term "entropy" are tools of Information Theory (IT; Shannon and Weaver 1949), and well established in the field of computational linguistics. IT is concerned with representing mathematically how much information, measured in terms of binary choices (bits), is needed to efficiently convey a message given the constraints imposed on the system. Entropy is a measure of the amount of uncertainty associated with selecting outcomes in a given system. A higher entropy value is associated with greater uncertainty among a set of possible outcomes; those outcomes that are relatively unexpected contribute more information, or complexity, to a message.

Jakobson's theory of distinctive features and the binary nature of markedness were inspired by Information Theory. According to Cherry *et al.* (1953: 34), analyzing a language requires that "we must determine the minimum set of [distinctive] features that the listener needs in order to recognize and distinguish all except homonymic morphemes, without help from context or situation." Each binary feature represents a choice regarding a particular sound (is it voiced? is it nasal?), with the response encoded as plus (yes) and minus (no) (zero indicates redundant information). A plus value is generally interpreted as marked, and minus as unmarked. The goal of this approach was to determine how many binary choices were needed to identify a given phoneme within the language system. Consistent with Information Theory, the most efficient system was the one that required the fewest number of binary questions (i.e. features) to convey the identity of a particular phoneme. Many of these ideas became cornerstones of subsequent theories of feature specification, and the information-theoretic concept of binarity emerged as a fundamental assumption of markedness theories.

Entropy has also been applied to other aspects of phonological knowledge; for applications of information-theoretic concepts to sound patterns, see Hockett (1955), Broe (1996), Goldsmith (1998, 2002), Hume and Bromberg (2005), Hume (2006), Hall (2009), Goldsmith and Riggle (forthcoming), Hall *et al.* (forthcoming), Hume and Mailhot (forthcoming); also CHAPTER 6: SELF-ORGANIZATION IN PHONOLOGY. For example, Hume (2006), Hall *et al.* (forthcoming), and Hume and Mailhot (forthcoming) suggest that entropy and information content can be used to make predictions about the preferred contexts and likely outcomes of a range of phonological processes, hence markedness patterns. Specifically, the most likely targets are predicted to be those that contribute little to the total entropy of the system, or in the case of repair processes, those that would significantly increase entropy if left unrepaired. The most likely outcomes are predicted to be ones that result in the least change in the pre-existing entropy of the system. One consequence of this approach is that it unites a range of otherwise disparate observations regarding influences on phonological patterns within a given system. For instance, Hall *et al.* (forthcoming) argue that in cognition, the information content of an element is influenced not only by its frequency, but also by its phonetic salience and by the attention a language user allocates to it in context. These are factors that linguists have independently argued influence phonological patterns. Laboratory studies of a number of languages show that listeners tend to allocate more attention to beginnings of words. Correspondingly, many languages restrict marked elements or structures to initial syllables of words (e.g. Beckman 1997).

To briefly illustrate, consider a language system L with a set of elements $X = \{x_i, x_{ii}, \ldots x_n\}$. Predicting which element will occur in a given context can be measured as a function of its probability, determined in part by its frequency of occurrence (see below for the role of acoustic salience in defining information content). Each element thus has its own probability of occurrence: sound x_i has probability $p(x_i)$. The information content of an element x_i is the negative logarithm base 2 of its probability p. This corresponds to our intuition that elements with higher probability are more expected than elements with lower probability; in information-theoretic terms, it takes less information (fewer binary decisions) to determine whether or not it will occur. In the vowel epenthesis example below, for instance, information content is used to compare the probability that a particular vowel occurs in a given context with the probabilities for all vowels in the system that comprise the relevant probability distribution.

Comparisons can also be made at the level of systems, rather than individual items: the total entropy of a system, or its complexity, is the sum of the information contributed by all members. To the extent that systems from different languages are compared, this measure can address questions relating to universal markedness patterns. To measure the contribution of a particular element to the entropy of the system, its information content is multiplied by the probability of its occurrence in that system, that is, $-p(x_i) \log_2 p(x_i)$. Note that an element's probability appears twice in this expression, with opposite effects. As a result, elements with either very high or very low probability contribute little to total system entropy: elements with very low probability have high information content, but contribute little because of their low rate of occurrence; conversely, elements that occur very frequently contribute little because their information content is correspondingly low. Elements with intermediate probability contribute relatively more to the complexity of the system, that is, its entropy, through their balance of moderate information content and frequency of occurrence.

The common markedness diagnostic, epenthesis, illustrates how these concepts can be used to answer the questions: (i) why epenthesize?; and (ii) why epenthesize a particular vowel? Consider the case of English vowel epenthesis as used to repair a non-occurring word-initial cluster. As noted above, a word-initial [gd] cluster as in the Polish city name Gdansk is often repaired as [gəd] (or [gɪd]). The first question concerns the motivation for epenthesis. In this approach, the answer is that adding [gd] to the language would, all else being equal, increase the complexity of the system, measured as entropy. This conclusion derives from considering the information content of the structures [gd] and [gəd], measured as the negative $\log_2$ of the conditional probability of the respective phonotactic sequences in English and the contribution of each to the entropy of the system. (In this simple case, information content can be approximated solely in terms of the type or token frequency of each sequence in a corpus of English.) Prior to borrowing, [gd] has a probability of zero, and contributes nothing to the entropy of the system; adding this sequence increases systemic entropy sharply. In contrast, increasing the frequency of an already occurring element such as [gəd] has less of an effect on the entropy of the system.

Parallel reasoning can be used to predict why one segment is the preferred epenthetic vowel in a given context over other vowels in the language. Why, for example, is the initial [gd] cluster typically repaired in English as [gəd] or [gɪd], as opposed to [gad], [god], [gid], among other possibilities? Common explanations point to the epenthetic vowel as unmarked, less salient, more frequent, predictable or somehow simpler than other vowels in the language. The information-based approach is consistent with these insights. A simple model takes into account the probability of each vowel in the system as a function of frequency and mutual information (context-dependent frequencies), and the phonetic nature of the sound in the relevant context, measured in terms of the quality of its acoustic cues. In this model, the information content of a given vowel is a function of both its probability of occurrence and its acoustic salience. The results indicate that [ə ɪ] have the lowest information content of all vowels in the system, consistent with the observation that these vowels commonly emerge as epenthetic.

While English epenthesis provides a relatively straightforward example, the approach extends to cases such as vowel epenthesis in French where, as noted above, a mid front rounded vowel serves as the epenthetic vowel; as such, the initial [gd] cluster of *Gdansk* would commonly be repaired as [gød] or [gœd]. French epenthesis is of particular interest, given the widely held view that front rounded vowels are universally marked (Chomsky and Halle 1968; Causley 1999; de Lacy 2006). However, the front rounded vowels are not only among the most frequent vowels in French, but also have been shown to be the most confusable with other vowels in the system. The information-based approach correctly predicts the mid front rounded vowels to have the lowest information content among vowels in the language, and to be the preferred epenthetic vowels (Hume and Bromberg 2005).

This approach provides an answer to the conflict regarding perceptual salience noted above: unmarked status is associated both with patterns with strong salience and those with weak salience. The salience of sounds in a system can be calculated as the probability of a given element being correctly identified. To the extent that the probability distribution is asymmetrical such that some sounds have a high probability of being misidentified (weak salience), these sounds will have a low value for information content and thus contribute little to the entropy of the system (as defined by the relevant probability distribution). All else being equal, such sounds are likely candidates for reduction and deletion. Interestingly, the observation that sounds with low information content are prone to deletion as well as epenthesis also addresses another apparent conflict among markedness criteria: the unmarked segment is not only the most likely to be deleted, it is also the most likely to be inserted. As noted above, the proposed explanation is because changes to the entropy of the system would be minimally affected by those elements that contribute little to its complexity. Sound patterns with strong salience, on the other hand, have higher information content; as a result, their loss would result in a more drastic change to the system. Thus, all else being equal, patterns with stronger salience are predicted to be more stable and avoid change.

6 Conclusion

As should be evident from the discussion above, the concept of markedness has had a long and controversial history in phonology. The goal of understanding why some elements pattern differently from others is at the heart of the linguistic and, more generally, cognitive enterprise. As a result, I suspect that markedness, whatever "it" is, will continue to be a much debated topic for years to come.

ACKNOWLEDGMENTS

My sincere thanks to Marc van Oostendorp, Keren Rice, and an anonymous reviewer, whose valuable comments strengthened this chapter. I would also like to thank Colin Ewen for his assistance in preparing the final version. This chapter is dedicated to the memory of G. Nick Clements who, on my generals exam at graduate school, posed the question of what markedness is, thus starting me on my quest to get to the bottom of it.

REFERENCES

Akinlabi, Akinbiyi. 1985. Tonal underspecification and Yorùbá tone. Ph.D. dissertation, University of Ibadan.

Archangeli, Diana. 1984. Underspecification in Yawelmani phonology and morphology. Ph.D. dissertation, MIT.

Archangeli, Diana & Douglas Pulleyblank. 1989. Yoruba vowel harmony. *Linguistic Inquiry* 20. 173–217.

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Avery, Peter & Keren Rice. 1989. Segment structure and coronal underspecification. *Phonology* 6. 179–200.

Bailey, Charles-James N. 1973. *Variation and linguistic theory*. Arlington, VA: Centre for Applied Linguistics.

Battistella, Edwin. 1990. *Markedness: The evaluative superstructure of language*. Albany: State University of New York Press.

Battistella, Edwin. 1996. *The logic of markedness*. Oxford: Oxford University Press.

Beckman, Jill N. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14. 1–46.

Beckman, Mary E., Kiyoko Yoneyama & Jan Edwards. 2003. Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. *Journal of the Phonetic Society of Japan* 7. 18–28.

Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104. 591–630.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Boersma, Paul. 1998. Functional phonology: Formalizing the interactions between articulatory and perceptual drives. Ph.D. dissertation, University of Amsterdam.

Boersma, Paul. 2000. Phonetically-driven acquisition of phonology. Unpublished ms., University of Amsterdam.

Broe, Michael B. 1996. A generalized information-theoretic measure for systems of phonological classification and recognition. In Richard Sproat (ed.) *Computational Phonology in Speech Technology: 2nd meeting of the ACL Special Interest Group in Computational Phonology, Proceedings of the Workshop*, 17–24. Philadelphia: Association for Computational Linguistics.

Bybee, Joan. 2001. *Phonology and language use.* Cambridge: Cambridge University Press.

Bybee, Joan. Forthcoming. Markedness: Iconicity, economy and frequency. In Jae Jung Song (ed.) *The Oxford handbook of language typology.* Oxford: Oxford University Press.

Cairns, Charles & Mark Feinstein. 1982. Markedness and the theory of syllable structure. *Linguistic Inquiry* 13. 193–225.

Calabrese, Andrea. 1995. A constraint-based theory of phonological markedness and simplification procedures. *Linguistic Inquiry* 26. 373–463.

Causley, Trisha. 1999. Complexity and markedness in Optimality Theory. Ph.D. dissertation, University of Toronto.

Cherry, E. Colin, Morris Halle & Roman Jakobson. 1953. Toward the logical description of languages in their phonemic aspect. *Language* 29. 34–46.

Chomsky, Noam. 1965. *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use.* New York: Praeger.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. 1988. Toward a substantive theory of feature specification. *Papers from the Annual Meeting of the North East Linguistic Society* 18. 79–93.

Clements, G. N. 1993. Underspecification or nonspecification? *Proceedings of the Eastern States Conference on Linguistics (ESCOL)* 10. 58–80.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable.* Cambridge, MA: MIT Press.

Comrie, Bernard. 1983. Markedness, grammar, people and the world. In Eckman *et al.* (1983), 85–106.

Croft, William. 2003. *Typology and universals.* 2nd edn. Cambridge: Cambridge University Press.

de Lacy, Paul. 2002. The formal expression of markedness. Ph.D. dissertation, University of Massachusetts, Amherst.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology.* Cambridge: Cambridge University Press.

Dyck, Carrie. 1995. Constraining the phonetics–phonology interface. Ph.D. dissertation. University of Toronto.

Eckman, Fred, Edith Moravcsik & Jessica Wirth (eds.) 1983. *Markedness.* New York & London: Plenum Press.

Elšík, Viktor & Yaron Matras. 2006. *Markedness and language change: The Romani sample.* Berlin & New York: Mouton de Gruyter.

Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory.* Cambridge, MA & Oxford: Blackwell.

Goldsmith, John A. 1998. On information theory, entropy, and phonology in the 20th century. Paper presented at the Royaumont CTIP II Round Table on Phonology in the 20th Century.

Goldsmith, John A. 2002. Probabilistic models of grammar: Phonology as information minimization. *Phonological Studies* 5. 21–46.

Goldsmith, John A. & Jason Riggle. Forthcoming. Information theoretic approaches to phonological structure: The case of Finnish vowel harmony. *Natural Language and Linguistic Theory.*

Greenberg, Joseph H. 1966. *Language universals, with special reference to feature hierarchies.* The Hague: Mouton.

Hall, Kathleen Currie. 2009. A probabilistic model of phonological relationships from contrast to allophony. Ph.D. dissertation, Ohio State University.

Hall, Kathleen Currie, Elizabeth Hume, Dahee Kim, Frédéric Mailhot, Adam Ussishkin & Andrew Wedel. Forthcoming. Entropy as an organizing principle in phonology. *Theoretical Linguistics.*

Hamilton, Philip. 1996. Phonetic constraints and markedness in the phonotactics of Australian aboriginal languages. Ph.D. dissertation, University of Toronto.

Hardcastle, W. J. & A. Marchal (eds.) 1990. *Speech production and speech modeling.* Dordrecht: Kluwer.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42. 25–70.

Hayes, Bruce & Donca Steriade. 2004. Introduction: The phonetic bases of phonological markedness. In Hayes *et al.* (2004), 1–33.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology.* Cambridge: Cambridge University Press.

Hockett, Charles F. 1955. *A manual of phonology.* Baltimore: Waverly Press.

Hume, Elizabeth. 1999. The role of perceptibility in consonant/consonant metathesis. *Proceedings of the West Coast Conference on Formal Linguistics* 17. 293–307.

Hume, Elizabeth. 2003. Language specific markedness: The case of place of articulation. *Studies in Phonetics, Phonology and Morphology* 9. 295–310.

Hume, Elizabeth. 2004a. The indeterminacy/attestation model of metathesis. *Language* 80. 203–237.

Hume, Elizabeth. 2004b. Deconstructing markedness: A predictability-based approach. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 13. 182–198.

Hume, Elizabeth. 2006. Language specific and universal markedness: An information-theoretic approach. Paper presented at the 80th Annual Meeting of the Linguistic Society of America, Albuquerque.

Hume, Elizabeth. 2008. Markedness and the language user. *Phonological Studies* 11. 295–310.

Hume, Elizabeth & Ilana Bromberg. 2005. Predicting epenthesis: An information-theoretic account. Paper presented at 7èmes journées internationales du réseau français de phonologie, Aix-en-Provence.

Hume, Elizabeth & Keith Johnson. 2001. A model of the interplay of speech perception and phonology. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 3–26. San Diego: Academic Press.

Hume, Elizabeth & Frédéric Mailhot. Forthcoming. Entropy and surprisal in phonologization and language change. In Alan C. L. Yu (ed.) *Origins of sound patterns: Approaches to phonologization.* Oxford: Oxford University Press.

Hume, Elizabeth & Georgios Tserdanelis. 2002. Labial unmarkedness in Sri Lankan Portuguese Creole. *Phonology* 19. 441–458.

Hura, Susan L., Björn Lindblom & Randy L. Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35. 59–72.

Iverson, Gregory K. & Kee-Ho Kim. 1987. Underspecification and hierarchical feature representation in Korean consonantal phonology. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 182–198.

Jakobson, Roman. 1932. *Zur Struktur des russischen Verbums. Charisteria Gvilelmo Mathesio qvinqvagenario a discipulis et Circuli Lingvistici Pragensis sodalibus oblata.* Prague. Reprinted as Structure of the Russian verb, in Linda R. Waugh & Morris Halle (eds.) (1984), *Roman Jakobson: Russian and Slavic grammar. Studies 1931–1981*, 1–14. Berlin: Mouton.

Jakobson, Roman. 1971a. The sound laws of child language. In Roman Jakobson (ed.) *Studies on child language and aphasia*, 7–20. The Hague & Paris: Mouton. Reprinted in Waugh & Monville-Burston (1990), 294–304.

Jakobson, Roman. 1971b. Implications of language universals for linguistics. In Roman Jakobson. *Selected writings*, vol. 2, 580–592. The Hague: Mouton.

Jakobson, Roman & Morris Halle. 1956. *Fundamentals of language*. The Hague: Mouton.

Jakobson, Roman & Krystyna Pomorska. 1990. The concept of mark. In Waugh & Monville-Burston (1990), 134–140.

Jakobson, Roman & Linda R. Waugh. 1987. *The sound shape of language*. 2nd edn. Berlin: Mouton de Gruyter.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Jun, Jongho. 1995. Place assimilation as the result of conflicting perceptual and articulatory constraints. *Proceedings of the West Coast Conference on Formal Linguistics* 14. 221–237.

Kawasaki, Haruko. 1982. An acoustical basis for universal constraints on sound sequences. Ph.D. dissertation, University of California, Berkeley.

Kean, Mary-Louise. 1975. The theory of markedness in generative grammar. Ph.D. dissertation, MIT.

Kewley-Port, Diane & Malcolm S. Preston. 1974. Early apical stop production: A voice onset time analysis. *Journal of Phonetics* 2. 195–210.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Kohler, Klaus J. 1990. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In Hardcastle & Marchal (1990), 69–92.

Lass, Roger. 1975. How intrinsic is content? Markedness, sound change, and "family universals." In Didier L. Goyvaerts & Geoffrey K. Pullum (eds.) *Essays on The Sound Pattern of English*, 475–504. Ghent: E. Story-Scientia.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48. 839–862.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H & H theory. In Hardcastle & Marchal (1990), 403–439.

Lombardi, Linda. 1997. Coronal epenthesis and markedness. *University of Maryland Working Papers in Linguistics* 5. 156–175.

Mascaró, Joan. 1976. Catalan phonology and the phonological cycle. Ph.D. dissertation, MIT.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

Meier, Tinka. 1999. Phonological markedness revisited: Evidence from Caucasian languages. In Helma van den Berg (ed.) *Studies in Caucasian linguistics*, 181–196. Leiden: CNWS.

Menn, Lise. 1983. Language acquisition, aphasia and phonotactic universals. In Eckman *et al.* (1983), 241–256.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Mohanan, K. P. 1992. Emergence of complexity in phonological development. In Charles A. Ferguson, Lise Menn & Carol Stoel-Gammon (eds.) *Phonological development: Models, research, implications*, 635–662. Timonium, MD: York Press.

Mohanan, K. P. 1993. Fields of attraction in phonology. In John A. Goldsmith (ed.) *The last phonological rule: Reflections on constraints and derivations*, 61–116. Chicago & London: University of Chicago Press.

Morrisette, Michele, Daniel Dinnsen & Judith Gierut. 2003. Markedness and context effects in the acquisition of place features. *Canadian Journal of Linguistics* 48. 329–355.
Munson, Benjamin. 2000. Phonological pattern frequency and speech production in children and adults. Ph.D. dissertation, Ohio State University.
Odden, David. 1992. Simplicity of underlying representation as motivation for under-specification. *OSU Working Papers in Linguistics* 41. 85–100.
Ohala, John J. 1981. The listener as a source of sound change. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 17(2). 178–203.
Ohala, John J. 1990. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26(2). 319–338.
Paradis, Carole & Jean-François Prunet. 1991. Introduction: Asymmetry and visibility in consonant articulations. In Carole Paradis & Jean-François Prunet (eds.) *The special status of coronals: Internal and external evidence*, 1–28. San Diego: Academic Press.
Peperkamp, Sharon. 2007. Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin. *Cognition* 104. 631–637.
Phillips, Betty S. 2006. *Word frequency and lexical diffusion*. New York: Palgrave MacMillan.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.
Pulleyblank, Douglas. 1988. Vocalic underspecification in Yoruba. *Linguistic Inquiry* 19. 233–270.
Rice, Keren. 1992. On deriving sonority: A structural account of sonority relationships. *Phonology* 9. 61–99.
Rice, Keren. 1996. Default variability: The coronal–velar relationship. *Natural Language and Linguistic Theory* 14. 493–543.
Rice, Keren. 1999. Featural markedness in phonology: Variation. *GLOT International* 4.7: 3–6, 4.8: 3–7.
Rice, Keren. 2007. Markedness in phonology. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 79–97. Cambridge: Cambridge University Press.
Rice, Keren & Peter Avery. 1993. Segmental complexity and the structure of inventories. *Toronto Working Papers in Linguistics* 12. 131–153.
Rice, Keren & Trisha Kathleen Causley. 1998. Asymmetries in featural markedness: Place of articulation. Paper presented at the 21st Generative Linguistics in the Old World Conference, Tilburg University.
Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.
Shannon, Claude Elwood & Warren Weaver. 1949. *The mathematical theory of communication*. Urbana-Champaign: University of Illinois Press.
Stampe, David. 1979. *A dissertation on natural phonology*. New York: Garland.
Stemberger, Joseph P. 1992. Vocalic underspecification in English language production. *Language* 68. 492–524.
Steriade, Donca. 1987. Redundant values. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 339–362.
Steriade, Donca. 1995. Underspecification and markedness. In Goldsmith (1995), 114–174.
Steriade, Donca. 1997. Phonetics in phonology: The case of laryngeal neutralization. Unpublished ms., University of California, Los Angeles.
Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.
Vaux, Bert & Bridget Samuels. 2005. Laryngeal markedness and aspiration. *Phonology* 22. 395–436.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin.* Berlin: Mouton de Gruyter.

Vihman, Marilyn M. 1993. Variable paths to early word production. *Journal of Phonetics* 21. 61–82.

Waugh, Linda R. & Barbara A. Lafford. 1994. Markedness. In R. E. Asher (ed.) *Encyclopedia of languages and linguistics*, 2378–2383. Oxford: Pergamon Press.

Waugh, Linda R. & Monique Monville-Burston (eds.) 1990. *On language: Roman Jakobson.* Cambridge, MA: Harvard University Press.

Zipf, George K. 1932. *Selected studies of the principle of relative frequency in language.* Cambridge, MA: Harvard University Press.

5 The Atoms of Phonological Representations

Marianne Pouplier

1 Introduction

Phonology describes spoken language as a combinatorial system typically comprising 30–40 basic units. The smallest known inventories make use of about 10 contrastive units, the largest no more than 150 (Maddieson 1984). Phonological theories capture, on the one hand, the nature of these basic units themselves (the atoms of speech production and perception) and, on the other hand, the language-specific patterns that the units show when they combine into larger structures. One of the key problems in relating this combinatorial system to the phonetic, physical side of speaking lies in the spatio-temporal structure of speech, which does not seem to mirror the units of linguistic analysis. When we inspect instrumental records of speech, the discrete combinatorial system that underlies these records is by no means obvious. We see continuously changing, highly context-dependent articulator motion and spectral energy patterns, with no clear boundaries between the individual sounds. The discrepancy between the symbolic units used to capture grammatical regularities and the dynamic complexities of speech as a motor act is so fundamental that some researchers have adopted the view that phonology and phonetics best be studied as separate sciences, since it seems difficult to see how the insights of each discipline could be relevant for the other. Current phonological and phonetic theories are, however, informed by the insight that phonetics and phonology are intertwined to such an extent that one cannot be understood without the other. How the relationship between phonetics and phonology can be negotiated has therefore been one of the key research topics in both disciplines (Pierrehumbert 1990; Boersma 1998; Flemming 2001; Hume and Johnson 2001; Keating *et al.* 2003; Hayes *et al.* 2004; Prince and Smolensky 2004; to name but a few); see also Kingston and Beckman (1990) and subsequent Laboratory Phonology volumes.

Several theories have tackled this issue in a new fashion since the 1980s, most prominently by positing the phonetic grounding of constraints in optimality-theoretic approaches and by recasting phonological primitives as dynamic entities within gestural theories. Within the optimality-theoretic framework, phonological constraints can directly encode speakers' knowledge of the kinematics of speech production and of characteristics of the auditory system. For instance, sounds can be ranked according to a scale of articulatory effort, and constraints may target different points along that scale (in particular so in the functional phonology school, e.g. Jun 1995; Boersma 1998; Flemming 2001; Steriade 2009). In addition, the perceptibility of sounds can directly be referred to in phonological constraints, often in the form of a requirement of contrast maintenance (CHAPTER 2: CONTRAST). Assimilation patterns (CHAPTER 81: LOCAL ASSIMILATION) have for instance been modeled as affecting primarily the least perceptible segments. In this view, speakers conserve articulatory effort (reduce a given articulation) if they know that it is likely to go undetected by a listener or if the perception of a phonological contrast is not endangered (Hura *et al.* 1992; Steriade 2001). Continuous constraint evaluation is also used to predict undershoot and hyper- or hypoarticulation in the sense of Lindblom (1963, 1990); see Boersma (1998) and Jun (2004). These theoretical frameworks capture how phonological constraints exploit the physiological givens of speech production and perception, but many accounts still rely on discrete phonological units that are implemented phonetically at some unspecified later translation stage.

Flemming (2001) advocates that these phonetically informed constraints also operate on continuous, phonetically enriched representations. Category effects arise from constraints requiring maximal contrast, i.e. contrast will only arise if there is sufficient phonetic distinctiveness. He assumes that consonants and vowels have acoustic targets that are fixed in terms of formants and duration. Contextual deviations from the target fall out from effort avoidance constraints and the interaction of effort avoidance with constraints requiring the maintenance of contrast. Flemming (2001) illustrates his account with an example using an F2 target; how the account could be expanded to the full wealth of acoustic cues and their trading relations remains open. He further relies on a direct correlation between extent of articulator movement and acoustic change, but, due to the many nonlinearities in the acoustic–articulatory relationship (Stevens 1989), this can only be a first approximation (for the general controversy surrounding the assumption of acoustic targets in speech production see Fowler 1980, 1986; Guenther *et al.* 1998). Flemming's model shows how in principle both gradient and categorical effects may arise from the same representations. His model remains incomplete, however, in that it fails to incorporate the relative timing of articulatory events beyond a linear sequencing of target specifications.

A full-fledged model of how the temporal orchestration of articulatory events is shaped by linguistic structure and how phonological representations can carry temporal specifications while still providing means for expressing phonological contrast has been developed by gestural approaches. Gestural theories like Articulatory Phonology (Browman and Goldstein 1985) have departed from the assumption of symbolic representations: the basic units (gestures) are dynamically specified, goal-directed movements in the vocal tract and are coordinated with respect to one another to form larger molecules. Coupling provides the "glue" that makes gestures cohere into larger molecules. Phonology, in this view, directly governs coordination relationships (relative timing) among articulatory gestures.

We can quickly appreciate why the claim that phonological units should be abstract dynamic specifications of articulator motion provides a formidable theoretical challenge (and has been by no means uncontroversial) when we consider that the generation of any single sound as much as of any utterance requires a highly skilled coordination among numerous components of the respiratory, laryngeal, and supralaryngeal systems. For the production of a single sound, it has been estimated that 70–80 pairs of muscles are required; for instance, 12 muscle pairs alone participate in lip movement (Hardcastle 1976). The physical production of speech, in contrast to the phonological inventory of a language, involves a large number of subcomponents. Yet numbers express only a (perhaps small) part of the complexity issue: the temporal discreteness of phonological representations ("beads on a string") can only predict the temporal orchestration of articulatory events in the vocal tract to a limited extent.

Just prior to the development of the gestural approach, Autosegmental Phonology (Goldsmith 1976, 1990) (see also CHAPTER 14: AUTOSEGMENTS) was a major phonological development driven by the insight that the absolute linearity of phonological representations (Chomsky and Halle 1968) does not do justice to the temporal structure of phonological specifications. Goldsmith (1976) refers to this as the "Absolute Slicing Hypothesis," meaning that feature matrices belonging to each segment all have exactly the same domain. Autosegmental Phonology abandoned the idea of absolute slicing or absolute linearity (see also CHAPTER 54: THE SKELETON). Instead, Goldsmith used the analogy of a musical score that represents different information channels simultaneously. Non-linear phonology was motivated by the insight that the domains of linguistic specifications (e.g. the sequence of feature columns specifying the CV content of a word and the tonal specifiications of a word) do not necessarily line up in a 1:1 fashion. Temporal complexity of phonological representations was therefore expanded by introducing different tiers, which allow each linguistic specification to have a larger or smaller scope than a feature column on another tier (smaller in that, for example, two tonal features can be associated with a single vowel for a contour tone). Yet Autosegmental Phonology was not concerned with the translation problem: the basic units of representation remain symbolic and a-temporal. Non-linear phonologies still assume some unspecified translation stage that converts phonological symbols into the physics of speech.[1]

Articulatory Phonology, which spearheaded the gestural approach in the 1980s, is a much more radical departure from absolute linearity, by virtue of its claim that phonological representations are dynamic specifications of articulatory goals. Articulatory Phonology and Task Dynamics, as well as Motor Theory and Direct Realism, all belong to the early group of gestural theories. These theories have provided insights into different aspects of a skilled action approach to speech: Articulatory Phonology is a theory of how phonological processes can be understood by reference to the physical act of producing speech, and Task Dynamics provides a theory of speech motor control that enables us to understand how a continuous speech stream can arise from underlying combinatorial action units (Fowler 1980, 1984, 1996; Browman and Goldstein 1985, 1989, 1990a, 1992, 1995a; Liberman and Mattingly 1985; Fowler and Smith 1986; Kelso *et al.* 1986; Saltzman 1986; Saltzman and Munhall 1989; Liberman 1997; Goldstein *et al.* 2006). A treatment of Direct Realism and Motor Theory – the theoretical approaches that seek to explain how listeners parse these combinatorial action units from the continuous

[1] Another way to represent categorical temporal differences in phonological representations is Moraic Theory (Hayes 1989). Moraic Theory, however, does not provide an account of articulatory timing; furthermore, the mora needs to be translated into a phonetic-physical specification. For a gestural re-interpretation of moras as units of syllable weight, see the section on split-gesture dynamics.

speech stream – is outside the scope of this chapter, but useful overviews can be found in Hawkins (1999), Jusczyk and Luce (2002), and Galantucci *et al.* (2006).

We will elaborate on the basic concepts of gestural approaches in more detail in this chapter, and review some aspects of the controversies surrounding such a conceptualization of phonological representations. Recent theoretical proposals regarding the nature of gestures will also be presented. We will then discuss the modeling of syllable structure, as well as fluent speech phenomena as case studies illustrating the difference between segmental and gestural views. The last section discusses work that has combined gestural representations with optimality-theoretic constraints.

2 Gestures as units of action: Task Dynamics

Task Dynamics draws on the principles and insights from general motor behavior; the general model of skilled movement control has been adopted from non-speech domains to linguistics by Elliot Saltzman and colleagues (Saltzman 1979, 1986, 1991, 1995; Kelso and Tuller 1984; Browman and Goldstein 1985; Kelso *et al.* 1986; Saltzman and Kelso 1987; Saltzman and Munhall 1989; Turvey 1990). Task Dynamics provides a model of speech motor control showing how a continuous speech stream can arise from underlyingly discrete (action) units, without positing intrinsically timeless (symbolic) entities. These underlying units are called *gestures*. Task Dynamics addresses the nature of these primitive action units, the consequences of their partial or complete spatial and temporal overlap, and how coupling provides cohesion for larger units consisting of multiple gestures (gestural molecules).

The central problem in movement control – in speech as much as in any other area of skilled action – lies in the numerous degrees of freedoms that need to be controlled in an abstract fashion. Turvey *et al.* (1978) illustrate this on the basis of the five hinged parts controlling the spatial orientation of an airplane: to control the positioning of these parts individually would be a task stretching the limits of human information processing. By yoking the individual parts such that they move in a coordinated fashion, the angle of ascent/descent and the turn of a plane can be controlled by simply moving a joystick forward/backward or left/right. Also complex movements (skilled actions) in biological systems can be controlled macroscopically on the basis of so-called coordinative structures that yoke together many subparts into a single action unit (Bernstein 1967; Kelso and Tuller 1984; Kugler and Turvey 1987; Kay 1988; Kelso 1995). Action units are discrete, but their parameter settings (see below) give rise to continuous movement in time and space. Coordinative structures are functionally specific in that they are flexibly assembled depending on the requirements of a particular task; they are not hard-wired. In terms of speech, a speech task or gesture (e.g. Lip Closure) calls on several articulators (e.g. upper lip, lower lip, jaw) as a single coordinative structure; the articulators are grouped temporarily and flexibly into a functional unit that will accomplish the linguistic task. In this manner, gestures macroscopically govern the formation and release of local constrictions in the vocal tract. The task dynamic model importantly distinguishes *gestures* from *articulators* as two distinct but interacting levels. Put simply, gestures are phonological units, while articulators are not. To reach their tasks (Lip Closure), gestures call on several articulators (upper lip, lower lip, jaw) as a single coordinative structure. In this sense, gestures are abstract: they define global tasks (Lip Closure), not point-by-point movement. Each gesture is associated with a set of tract variables, specifying constriction location and degree, and a set of articulators that will act to perform the task (Saltzman and Munhall 1989).

Which articulatory component should be specified at the gestural level and which at the coordinative structure/articulator level is open to debate. For example, Mooshammer *et al.* (2007) propose that at least some consonants (specifically sibilants) should be specified for a jaw gesture, instead of the contribution of the jaw arising only by virtue of it being part of the coordinative structure. The authors base their argument on the fact that consonants with different manners of articulation differ systematically in the relative contribution of the jaw (Keating *et al.* 1994). However, which vocal-tract control component should be specified at the gestural or articulator level will usually be a phonological matter. Phonological processes like assimilation (CHAPTER 81: LOCAL ASSIMILATION) or harmony patterns (CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS; CHAPTER 118: TURKISH VOWEL HARMONY; CHAPTER 123: HUNGARIAN VOWEL HARMONY) probably do not refer to the jaw, and for these reasons the jaw component should be relegated to the articulator level.

Tasks are modeled as dynamical systems (hence the name Task Dynamics). A dynamical system provides us with the simplest and most abstract description of the forms of motion over time produced by a system on the basis of an invariant control structure. Note that tasks are abstract, but they are not symbolic: the gestural task is modeled as a virtual mass-spring system whose rest position depends on the linguistic task. For Lip Closure, the rest position is set to the Euclidean distance between the lips being zero. The position of the individual articulators may vary by context but the lip closure corresponds to the rest position of the task that the virtual springs will inevitably return to, irrespective of initial conditions or perturbation on the way. Motor skills are generally modeled on the basis of critically damped mass-spring systems precisely because they provide a good model of this *equifinality* (the rest position/target is reached independently of the initial position, the amount of displacement, or number of perturbations along the way). Coordinative structures allow for the necessary context-specific variability; the task can be reached with many different articulator configurations (a relative lower jaw position, but more raising of the lower lip, etc.) without the details of the configuration requiring central replanning or a task-level reparameterization (Turvey 1977; Abbs and Gracco 1984; Kelso and Tuller 1984; Kelso *et al.* 1984).

Task Dynamics is an intrinsic timing theory, meaning that the proposed units themselves carry inherently a temporal dimension. In contrast, extrinsic timing theories posit that the dimension of time is irrelevant for defining the phonological units themselves, or their relations in planned utterances. Symbolic segmental representations do not carry an intrinsic temporal component; the temporal evolution of segments during their phonetic implementation is provided externally through a timekeeper that is extrinsic to the units themselves and not explicitly represented in the plan (Fowler 1980). In Task Dynamics, however, the extension of sounds in time and their relative timing are intrinsic consequences of the system's dynamical organization. Distinct gestures for a given vocal organ are associated with a distinct set of dynamical parameter values that are fixed over time. A gesture's activation interval refers to the time interval during which these specifications are active. Each tract variable is associated with linear second order dynamics, that is, activation of a tract variable entails a specific damping ratio, a natural frequency (stiffness), and a target position (Saltzman and Munhall 1989). Gestures therefore carry temporal specifications in several ways: they have an *activation interval* during which they control the vocal tract; second, *intragestural timing* refers to each gesture's intrinsic temporal properties; and third, *intergestural timing* refers to the temporal coordination of multiple gestures with respect to each other.

Serial order arises in two ways. First, the different intrinsic temporal properties of gestures may give rise to serial order. For example, vowels are "slower" compared to consonants (they have a lower stiffness parameter setting): for a production of /ba/, for example, the vowel gesture may start before the lip gesture has reached its target (Perkell 1969; Löfqvist and Gracco 1999), but the vowel extends temporally beyond the consonant. Allophonic variation may also arise as a consequence of this temporal overlap of gestures. The allophonic variation of English /k/ in a front and back vowel context arise due to the velar constriction being blended with the co-occurring vowel constriction, since both vowels and /k/ control the tongue body (Öhman 1966); the temporal overlap of gestures is seen as the cause of co-articulation (Fowler and Saltzman 1993).

The other way to capture serial order is through coupling between multiple gestures; coupling provides the glue that makes gestures cohere into larger molecular structures. These coupling relations between gestures are expressed as relative timing or phase relations between gestures (intergestural timing). Since gestures are conceptualized as underlyingly oscillatory systems, larger gestural structures such as segments, syllables, and lexical items are modeled as coupled dynamical systems. Two specific coupling modes have a privileged status: it is well known that coupled dynamical systems in general exhibit naturally preferred coordination modes in terms of stability (Strogatz and Stewart 1993; Turvey 1990): Most stable is synchronous coordination (in-phase, 1:1 frequency) followed by a coordination 180° out-of-phase. Synchronous coordination has been observed to emerge spontaneously when skilled actions fail to be produced accurately, for example under increasing speed. While humans routinely perform highly complex patterns of coordination, these take time to learn and will, under the right conditions, revert to the stable synchronous coordination. There has also been some evidence for the significance of these preferred coordination modes in speech. For example, when subjects are instructed to repeated /ip/ at increasing speaking rate, we observe a spontaneous phase shift to the timing pattern characteristic for /pi/ (Kelso *et al.* 1986). Speech errors have also been argued to arise from a dynamic synchronization of articulatory gestures: they may result in the gestures for two consonants being produced on top of each other (Goldstein *et al.* 2007; Pouplier and Goldstein 2010). Gestural theories argue that these coordination patterns are not only part of speech motor control, but they are part of grammar and provide the basis for expressing cross-linguistic preferences such as a preference of CV over VC. They also constitute the basis of the gestural account of syllable structure (Browman and Goldstein 1988, 2000), speech errors (Goldstein *et al.* 2007; Pouplier 2008), and moraic structure (Nam 2006). Certain aspects of phonological development in language acquisition have likewise been modeled on the basis of in-phase and out-of-phase coordination (Studdert-Kennedy and Goldstein 2003).

Task Dynamics is a theory of how continuous articulator motion can be divided into discrete underlying action units (gestures). These action units can provide the small inventory of discrete units required for phonological contrasts. How gestures can function as combinatorial phonological units was first developed in the theory of Articulatory Phonology and will be detailed in the next section.

3 Gestures as units of phonological contrast

The gestural structure is the most abstract representation level of gestural planning. It captures the gestural tasks, the underlying coordination or coupling relationships, and the shapes and duration of each gesture's activation function that gives rise to the gestural score (Goldstein *et al.* 2006). The activation of gestures is not on-or-off; gestural activation rises and falls gradually (Byrd and Saltzman 1998). The dynamic parameterization (target, stiffness, and damping) of a gesture as lexically specified in the gestural score is invariant and context-independent, although there seems to be some notion that speaking rate or prosodic boundaries can affect the stiffness parameterization of a gesture. Crucially, however, due to the temporal co-existence of gestures, the articulator movement that the gestural score gives rise to is ***not*** invariant but varies with context (Browman and Goldstein 1990b; Fowler and Saltzman 1993). The gestural score is the driving input for the articulator level (although the levels interact), and the activation waves serve to insert a gesture's parameter settings (constriction location and degree) into the articulator dynamical system. Figure 5.1 shows an example gestural structure and score. The relatively independent articulatory subsystems (termed "vocal organs" by Browman and Goldstein) are arranged on separate tiers. The gestural structure

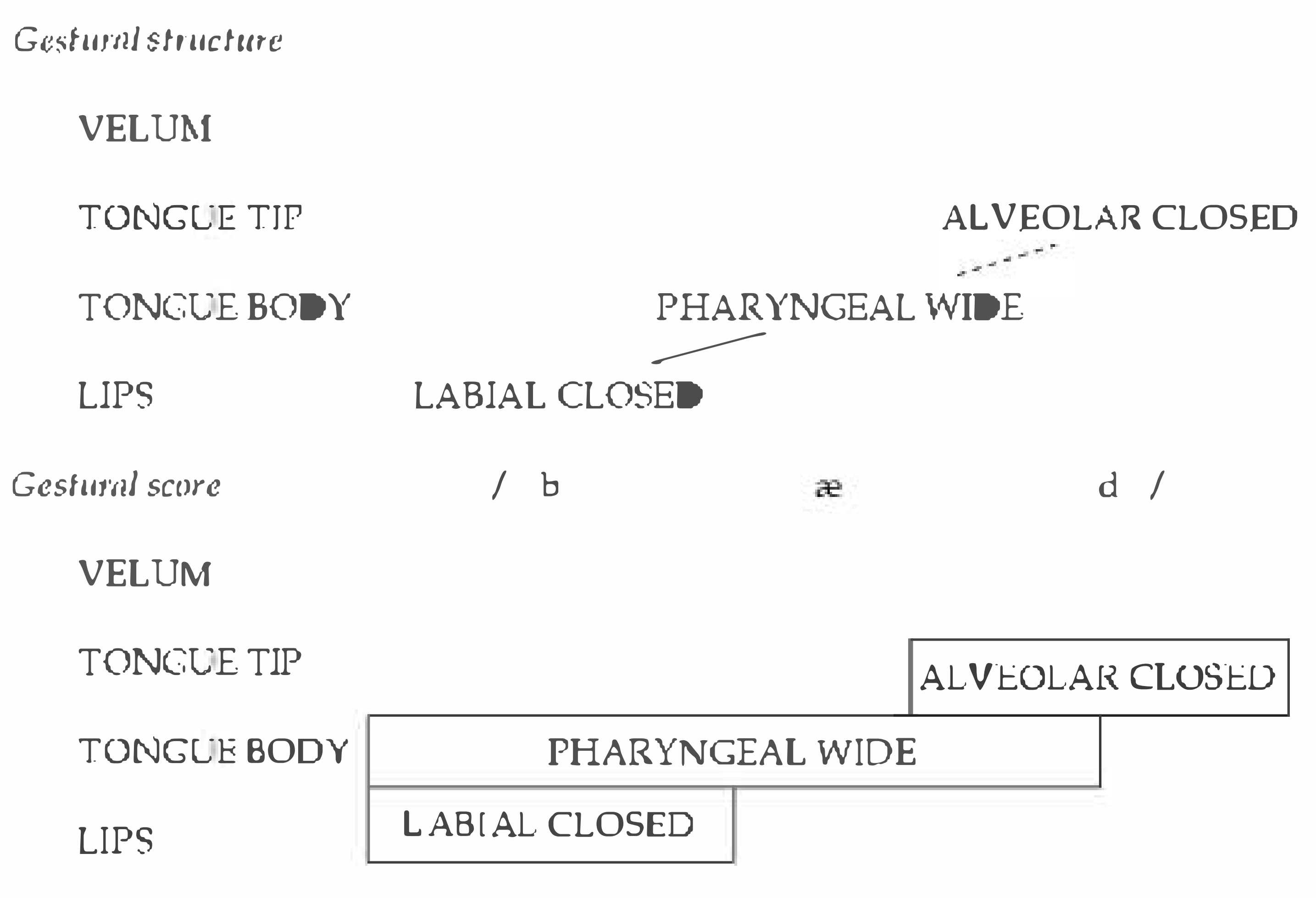


Figure 5.1 Gestural structure (top) and gestural score (bottom) for the English word *bad*. In the gestural structure, dashed lines indicate anti-phase coupling relations, while solid lines indicate in-phase coupling relations

specifies the coupling relations between the gestures of a lexical item. The gestural score is the relative timing pattern of gestures that arises from the gestural structure. Again, the vocal organs are arranged on separate tiers. The boxes indicate the time interval during which a given gesture actively controls the vocal tract.

The visual impression of the gestural structure highlights the metaphor of gestural atoms and molecules that both gestural and non-linear phonologies have used. Atoms that are bound together in a molecule are arranged geometrically, not in the linear sequence in which chemical formulas are written (Browman and Goldstein 2001). Similarly, gestures are arranged temporally in a complex manner to each other by means of different coordination relationships (see also Goldsmith's 1976 terminology of the "musical score," even though, as noted above, autosegmental association lines provide a domain specification rather than a specification of physical time).

The basis for discrete combinatorial units is provided by those vocal organs or constricting devices of the vocal tract that can perform constrictions relatively independently from each other (cf. relatedly the functional grouping of features in feature geometry: Clements 1985; Sagey 1986; Keyser and Stevens 1994; and the "information channels" in Goldsmith 1976): the tongue tip, tongue blade, tongue root, larynx, velum, and the lips. The parallels and differences to a feature-geometric approach are discussed in detail in Browman and Goldstein (1989; see also Padgett 1994). The six vocal organs are hypothesized to be inherently discrete and provide a small set which we can use as discrete, context-invariant units of phonological contrast (Browman and Goldstein 1989; Goldstein *et al.* 2006). Supporting evidence comes for instance from typological considerations: the typologically most widespread contrasts have been argued to involve distinct vocal organs as opposed to within-organ contrasts (Browman and Goldstein 2001; Goldstein and Fowler 2003). Tongue movement data also support a functional discretization of the parts of the tongue (Stone *et al.* 2004; Iskarous 2005; see further Studdert-Kennedy and Goldstein 2003 for arguments from phonological development).

Gestures can serve as the units of a combinatorial system in three basic ways. First, the presence or absence of a gesture is a categorical distinction. This is assumed to be the most basic type of contrast, since it involves the presence of a constriction of one of the six basic vocal organs. For example, *mad* has a velum opening gesture, while *bad* has no such velum gesture, *pad* has a glottal opening gesture while *bad* has no such gesture, etc. Note that glottal adduction ([+voice] in featural terms) and a closed velum ([−nasal]) are assumed to be default articulatory settings that do not need to be specified at the gestural level (Browman and Goldstein 1992; Goldstein and Browman 1986), so there is no glottal adduction and no velum closing gesture.

A second way in which gestures encode linguistic contrast is by differences in the dynamical parameters that define the spatio-temporal properties of a gesture (within-organ contrasts). Gestures employing the same vocal organ may differ parametrically in constriction location, constriction degree, and stiffness; for example, vowels and consonants differ in stiffness (and in constriction degree). A tongue tip gesture may be specified for a constriction degree value of "closure" (stops) or "critical" (fricatives). It is argued that there are stable ranges in these continuous parameter values that tend to contrast with one another cross-linguistically (cf. Quantal Theory; Stevens 1989). The third way in which gestures express phonological contrast lies in the relative coordination of different gestures to one another.

For example, the opposition between unaspirated, aspirated, and breathy stops is rooted in differences in gestural timing (and in size of the glottal gesture): for an unaspirated stop, the peak glottal opening occurs synchronously with the maximal closure, while for an aspirated stop the peak glottal opening gesture occurs at the consonant release with the consequence of a positive voice onset time. In breathy stops, the onset of the glottal opening gesture is timed to occur at the release (Sawashima and Hirose 1980). Some phonotactic rules can also be expressed as timing relations. English /s/ + stop cluster phonotactics (see CHAPTER 38: THE REPRESENTATION OF SC CLUSTERS) have been described in gestural terms as allowing only a single glottal opening gesture, the peak of which is presumably timed with respect to the maximal constriction of the /s/. As a consequence, the stops in such clusters are unaspirated. Phonological units in general (including segments) are all understood as characteristic patterns of temporal coordination among multiple gestures, as will be discussed in more detail next.

4 Atoms of speech production: Gestures, segments and split-gesture dynamics

It is important to note that in Browman and Goldstein's view (most explicitly stated in Browman and Goldstein 1986) the segment does not have a privileged status (CHAPTER 11: THE PHONEME; CHAPTER 54: THE SKELETON); segments are gestural molecules, just like onset clusters or syllables, for example. Gestures may sometimes be similar in grain size to segments or features, but they differ from both. A gesture corresponds to a feature in *bad vs. mad*. Presence/absence of a gesture corresponds to a segment in *bad vs. add*. For other features, such as [consonantal] or [sonorant] (CHAPTER 13: THE STRICTURE FEATURES), there is no corresponding gesture; these are generalizations over constriction degree, i.e. one of the dimensions of gestural parameterization (Browman and Goldstein 1989). The cluster /sp/ consists of two segments, yet gesturally it is composed of a tongue tip gesture and a labial gesture, as well as a single glottal opening gesture whose maximum opening is timed to occur at mid-frication; but see Saltzman and Munhall (1989) for alternative data and interpretations, and Hoole (2006) for a detailed investigation of the gestural timing of laryngeal and oral gestures in consonant sequences and a discussion of the question of what timepoints should best be considered for alignment.

Scobbie and Pouplier (2010) provide evidence for the gestural constellation view of segments in their investigation of British English /l/ (CHAPTER 31: LATERAL CONSONANTS). English /l/ is known to have two gestures: a tongue tip raising gesture and a tongue dorsum retraction gesture, and the timing and spatial magnitude of the gesture varies systematically with syllable position (Giles and Moll 1975). Scobbie and Pouplier report that the two gestures of English /l/ behave like two elements of a cluster rather than a single segment, in that the two gestures may behave independently of each other in external *l*-sandhi. There is also some evidence from speech errors supporting a gestural rather than a segmental view (Goldstein *et al.* 2007b). Other gestural approaches do recognize segmental levels of organization. Byrd has argued that segmental molecules have a particularly high degree of stability; thus, intra-segmental phase relationships (e.g. the phase relationships binding together the oral and glottal gestures of a voiceless consonant) have a privileged (more stable) status compared to inter-segmental phasing relationships (e.g. the phase relationship binding together the /s/ and /k/ of a /sk/ cluster: Saltzman and Munhall 1989; Byrd 1996; Byrd *et al.* 2009).

Moving a step further away from segmental level of organization, Browman (1994) argues that the release of a gesture needs to be actively controlled; the assumption of the initial gestural model that the release is governed by a return of the articulators to a neutral position (Browman and Goldstein 1985) was shown to be insufficient to predict articulator kinematics. This insight has consequently been implemented in Nam's recent theory of split-gesture dynamics (Nam and Saltzman 2003; Nam 2006, 2007). Nam extends the principles of the molecular composition of clusters to single gestures by modeling each "gesture" on the basis of a separate constriction formation and a separate release gesture. Each consonantal gesture is now conceived of as a gestural molecule or cluster in which formation and release gestures can enter different phonologically specified coupling relations. Singleton consonants and consonant clusters differ only in the number of components and coupling relations that form the molecular structure, but in principle they are both clusters. Empirical underpinnings come from a variety of areas: asymmetries in the kinematics of articulator motion into and out of a constriction have long been noted, and seem to necessitate a separate modeling of release and closure formation (Gracco 1988; Browman 1994). Further, the assumption of split gestures allows the modeling of different types of releases, such as released *vs.* unreleased stops (cf. also Steriade 1993). In addition, differences in articulatory timing and in timing variability in CV *vs.* VC sequences (Byrd 1994) can appropriately be captured by this model. The influence of prosodic factors on articulator kinematics has also previously been modeled as the movement into the constriction being truncated by the movement out of the constriction (Munhall *et al.* 1992; Harrington *et al.* 1995). On the phonological side, several facts about moraic structure and geminate inseparability (CHAPTER 37: GEMINATES) have been mustered as arguments for a separate representation of a release (Nam 2006, 2007). Nam (2006) gives the mora a gestural interpretation, and thus provides an account for the weightlessness of onsets (CHAPTER 55: ONSETS; CHAPTER 57: QUANTITY-SENSITIVITY) in terms of different coordination relations governing prevocalic and post-vocalic positions (cf. §5 on syllables). Nam conceptualizes moraic structure as differences in the complexity of coordination relationships, moraic weight becoming thus a continuous phenomenon, a step that may ultimately increase our understanding of typological timing or rhythm class differences between languages. Note that in the current task dynamic model of Nam *et al.* (2004) only oral consonantal gestures are hypothesized to carry split-gesture dynamics. For glottal, velic, and vocalic gestures, the whole movement cycle into and out of the constriction is modeled on the basis of a single gestural task.

Preceding Nam's work, there had been suggestions to account for the tense–lax opposition in the German vowel system by modeling the movement into and out of the vowel constriction on the basis of split vowel gestures by Hoole, Mooshammer, and colleagues (Kroos *et al.* 1997; Hoole and Mooshammer 2000) in an articulatory interpretation of Vennemann's syllable-cut hypothesis (Vennemann 1991). The authors argue that the difference between tense and lax vowels lies in their dynamic properties, namely the coupling of the opening to the closing gesture (with a relative tighter coupling of the split vowel gestures to each other for the lax vowels) and the kinematics of the opening movement. Lax vowels are more centralized, because the opening gesture into the vowel differs in its dynamics from the tense vowels, such that the lax vowel opening gesture is produced with a bigger amplitude and different acceleration profile. The idea that vowel inventory oppositions such as tense–lax pairs can be modeled on the basis of split-gesture dynamics has not as yet been developed any further (but see Restle 2003).

After reporting these different views on what grain size the atoms of speech production have, we will now consider syllable structure and assimilation as case studies of how gestural approaches model these core phonological phenomena with gestural rather than symbolic atoms.

5 Prosodic structure: Syllables as phonological molecules

In phonological theory, prosodic structure is defined in terms of a symbolic hierarchy of prosodic constituents (Nespor and Vogel 1986; Selkirk 1982; see also CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE; CHAPTER 40: THE FOOT; CHAPTER 51: THE PHONOLOGICAL WORD; CHAPTER 84: CLITICS; CHAPTER 50: TONAL ALIGNMENT). Gestural theories have modeled the syllable hierarchy, and the serial order provided by the different syllable positions on the basis of different intergestural coupling relations, that is, syllables are viewed as characteristic patterns of temporal coordination among multiple gestures (Krakow 1993; Browman and Goldstein 1995b, 2000). By hypothesis, the behavior of a gestural constellation in different syllable positions is governed by the two preferred coordination modes (in-phase, out-of-phase). Syllabic organization is therefore assumed to reflect the preferred coordination modes, typical for skilled action, in speech, whereby the onset carries the intrinsically less marked in-phase relationship.

Empirical evidence for the coordination theory of syllable structure comes from systematic differences in relative timing that can be observed as a function of syllable position. For example, in syllable-initial position, the two gestures of the initial /l/ in *leap* reach their target with a small lag, with the vocalic dorsal gesture being typically closer to the onset. Syllable-finally, e.g. in the word *peal*, the two gestures of /l/ show a different pattern: the tongue body retraction gesture begins early during the vowel (resulting in an audibly darkened vowel), while tongue tip raising lags behind the retraction gesture (Browman and Goldstein 1990b; Sproat and Fujimura 1993), although this may differ between languages (Gick *et al.* 2006). Similarly, for nasal consonants, the oral and velic gestures achieve their targets simultaneously initially, but the velic opening gesture precedes the oral gesture syllable-finally, resulting in a nasalized vowel (Krakow 1993, 1999; Byrd *et al.* 2009). It is assumed that these timing relationships are consequences of the underlying in-phase (onset) and out-of-phase (coda) coordination relationships.

Browman and Goldstein (1988, 2000) argue that in onset clusters, consonants *cannot* all be coordinated synchronously with the vowel; they would not be perceptually recoverable. It is thus assumed that onset consonants are coordinated sequentially with each other, yet each consonant is coordinated synchronously with the vowel, as illustrated in Figure 5.2. These incompatible phasing relationships yield a blended output in which the onset cluster as a whole retains the same stable temporal relationship with the vowel, independently of how many consonants the cluster contains. This stable temporal center of the ensemble of onset consonants has been termed the c-center. According to Browman and Goldstein,

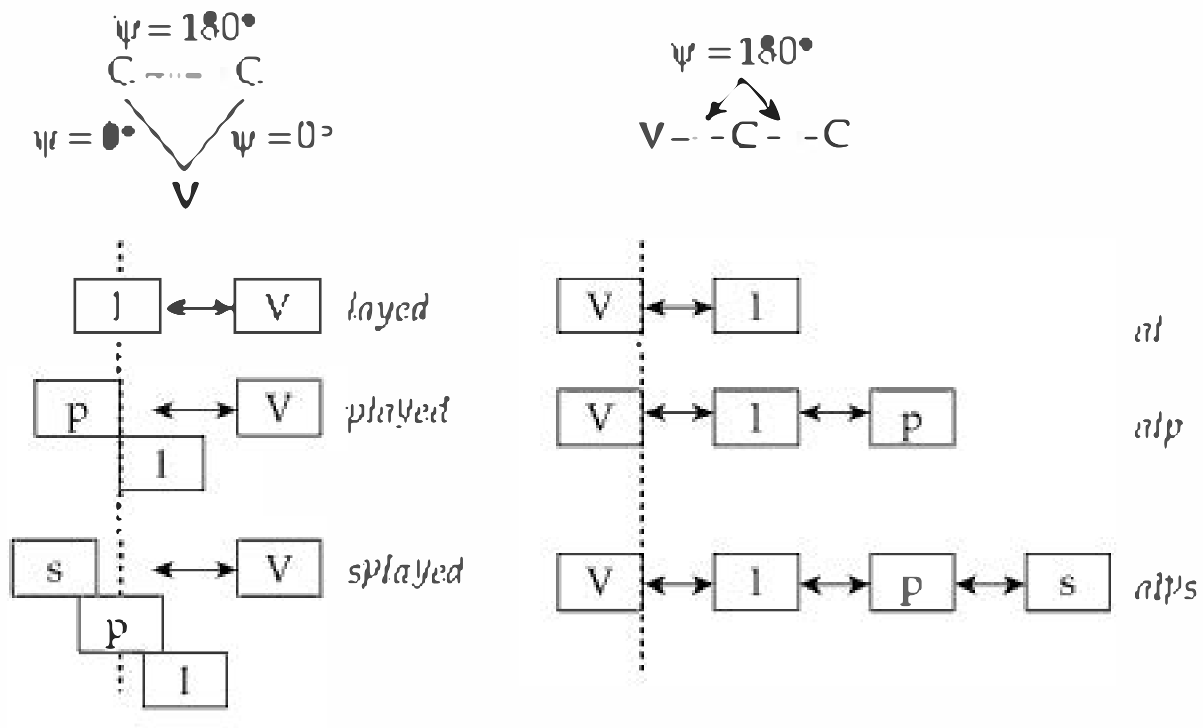


Figure 5.2 Schematic representation of c-centre coordination and the observed timing relationships in onset (left) *vs.* left-edge coordination in coda (right). The dashed vertical line indicates the stable timing point measured in articulatory data that is due to the underlying phase relationships (c-centre in onset; pairwise-sequential coupling in coda)

at least in English, the c-center is observed only in onsets, not in codas. In a $VC_1C_2C_3$ coda, pairwise-sequential phasing, also referred to as left-edge coordination, will be observed: C_1 will be coordinated 180° out of phase with the vowel, C_2 180° out of phase with C_1, etc.

That the onset exhibits predominantly the preferred in-phase coordination mode has been used to explain a variety of phonological phenomena, such as the primacy of onsets in language acquisition, the unmarked status of the CV syllable (but see Tabain *et al.* 2004; Butcher 2006), and the moraic weightlessness of onsets (Nam 2006). The coupled oscillator model of syllable structure further predicts that onsets should cross-linguistically exhibit c-center organization. Studies on English, Italian, and Georgian have confirmed c-center organization for onsets; the c-center has also been used as a diagnostic for syllable structure in Arabic and Berber (Honorof and Browman 1995; Goldstein *et al.* 2007a; Hermes *et al.* 2008; Shaw *et al.* 2009). Fewer studies have investigated coda clusters, and the results are less consistent than for onsets (Byrd 1995; Kochetov 2006; Marin and Pouplier 2010). The coherence of syllables into words and larger constituents has received only a limited amount of attention in gestural approaches so far. Smith (1995) proposed that stress- and mora-timing can be modeled on the basis of different cross-word coupling structures, but much work remains to be done in this area (cf. also Barbosa 2002a, 2002b; Yanagawa 2006; Tilsen 2009). For recent work modeling the prosodic hierarchy based on a hierarchy of coupled oscillators see Byrd and Choi (2010), Byrd and Saltzman (2003), and Saltzman *et al.* (2008).

6 Assimilation, overlap, and reduction

In connected speech we observe a variety of sandhi phenomena, such as assimilation. Describing and predicting patterns of assimilation has been one of the major motivations for non-linear phonologies, and remains a touchstone issue for theories modeling the phonetics–phonology interface. In a non-linear approach to phonology, assimilation has been captured on the basis of feature linking and delinking. For example in fluent speech, the cross-word [kt#m] cluster in the phrase *perfect memory* may be pronounced as [km], with the final coronal being assimilated to the following labial, schematically exemplified in (1):

(1) /t/ /m/ (→ [m])

Place Place

coronal labial

However, articulatory recordings of assimilated sequences revealed that a delinking of features may not accurately capture place assimilation. Articulatory recordings of the phrase *perfect memory* showed that although perceptually the final [t] was completely assimilated to the following nasal, the coronal constriction for the [t] was in fact still produced; it was only perceptually hidden by the temporally overlapping labial articulation (Figure 5.3). This example illustrates the consequences of a particular case of gestural overlap, namely when overlapping gestures control different vocal organs and tract variables. The articulations of coronal and labial

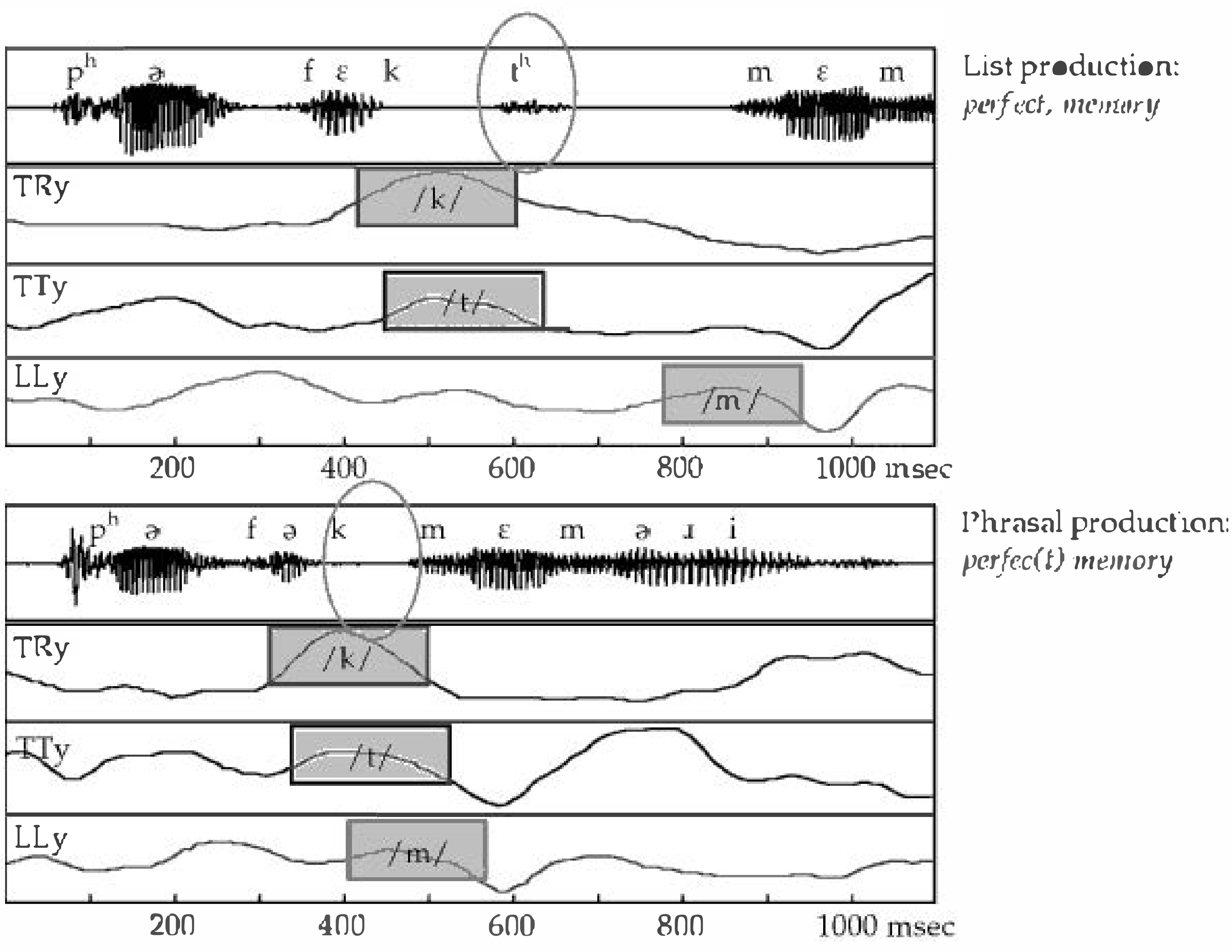


Figure 5.3 Articulatory data showing vertical tongue rear (TRy), tongue tip (TTy), and lower lip (LLy) movement over time. Boxes indicate the kinematic realization of the /k/, /t/, and /m/ gestures. The phrasal production illustrates how increased temporal overlap between gestures may lead to perceptual hiding of a gesture (/t/), although articulatorily the coronal constriction is produced in both cases. From Tiede *et al.* (2001). Used with permission

constrictions are largely independent, and do not affect each other significantly in the case of temporal overlap.

In other cases, articulatory place assimilation at word boundaries could be shown to be gradient, in that the spatial magnitude of the tongue tip gesture is reduced along a continuum of values (Surprenant and Goldstein 1998), yet articulatory assimilation may also be categorical, in that the assimilated gestures may be consistently and categorically absent (i.e. not produced: Honorof 1999; Son *et al.* 2007). For example, the tongue tip gesture for word-final alveolar /n/ in Castilian Spanish (e.g. in *diga*[m] *paja* 'say straw (FORM PL)'; cf. *diga*[n]) is categorically absent, and the lip aperture gesture is temporally extended (Honorof 1999).

Furthermore, if the overlap of gestures involves not stops, but continuants, the temporal overlap may have quite clearly audible consequences, as in *miss you*. In fluent speech, the phrase will be pronounced as [mɪʃju]. The tongue tip gesture for /s/ is overlapped by the tongue body raising gesture for /j/, leading to a production perceptually similar to [ʃ]. Articulatorily and acoustically, the assimilated sequence differs from an underlying [ʃ] (Zsiga 1995). However, for lexically derived forms such as *impression*, she finds no evidence of [ʃ] arising from a co-production of /s#j/ gestures. For lexically derived forms, Zsiga therefore assumes a symbolic (de)linking of features, and a gestural overlap account only for post-lexical assimilation.

The hypothesis that fluent speech phenomena do not involve a symbolic restructuring of phonological units and their affiliations to segmental timing slots, but can be exclusively understood as variation in timing and spatial magnitude (see especially Browman and Goldstein 1992: 173), sparked a heated debate. A vigorous exchange concerned English sibilant assimilation as it occurs in phrases such as *Paris show*. Nolan and colleagues (Holst and Nolan 1995; Nolan *et al.* 1996) investigated phrases like *claps Shawn*, and found evidence for gradient assimilation as well as categorical assimilation, and argued that the latter would more appropriately be described as a symbolic linking and delinking process rather than in terms of gestural overlap. Interestingly, for many tokens Nolan and colleagues also observe intermediate assimilation patterns, as expected for gestural overlap, but they do not offer an account of these assimilation patterns or of how the occurrence of one or the other type of assimilation might be conditioned. In other studies that have been conducted since (for several different languages), we likewise see a wide range of speaker behavior, and assimilation patterns consonant with a symbolic restructuring view (linking–delinking as sketched above) have emerged side-by-side with gradient assimilation and gestural hiding and blending phenomena as first described by Browman and Goldstein (Barry 1991; Farnetani and Busà 1994; Recasens and Pallarès 2001; Ellis and Hardcastle 2002; Kühnert and Hoole 2004; Kochetov and Pouplier 2008).

Overall, it has become clear over the years that a plethora of factors inflluence the degree of overlap found for any given utterance, such as the types of segments and (types of) lexical items, word frequency, and the (hypothesized) casualness of speech; also there are systematic differences in the degree of overlap between languages (Gibbon *et al.* 1993; Zsiga 2000; Chitoran *et al.* 2002; Kochetov 2006; Kochetov *et al.* 2007). There has to date been no deeper account that lets us predict which factors may condition the presence or absence of articulatory place assimilation in the form of a reduction of the final consonant, and under what circumstances any present place assimilation will be categorical or gradient in nature.

In this context, Zsiga (2009) advocates a combined autosegmental–gestural view: in her view gestural coupling relations can be manipulated by phonology; just like autosegmental association lines, they can be linked or delinked as a categorical phonological process. In her approach, the autosegmental association lines specify timing relationships in terms of in-phase or out-of-phase coupling of the associated gestures. Phonetic processes may then influence the timing of gestures within the association lines (in-phase, out-of-phase) provided by phonology. The phonetic fine-tuning of gestural timing leads to gradient assimilation caused by gestural overlap.

Most of the early work done in a gestural approach focused on the consequences of gestural temporal overlap for connected speech processes and on developing a theory of co-articulation as a coproduction of gestures. Gafos (2002) argues that the temporal organization of gestures is also relevant for discrete morphophonological alternations, and proposes a theory in which grammatical constraints govern the temporal coordination of gestures. Gafos's central hypothesis is that grammatical principles and constraints refer to temporal relations between gestures in terms of overlap patterns. He illustrates his argument with template satisfaction in Moroccan Colloquial Arabic. Markedness (CHAPTER 4: MARKEDNESS) is expressed in terms of different gestural coordination relations: for example, Gafos posits a constraint expressing a preference for a specific temporal alignment or phase relationship: CC-Coord = Align(C1, C-Center; C2, onset) requires alignment of the temporal midpoint (timepoint of maximal strength) of the gesture for C1 with the beginning of the activation for gesture C2. However, a second markedness constraint disallows the preferred coordination mode for homorganic consonants; this is a gestural interpretation of Leben's (1973) Obligatory Contour Principle. Motivation for this constraint is recoverability: if two homorganic stops are timed closely together, there will be no release of C1 leading to difficulties with its perceptual recovery. The interaction of these two constraints, together with various repair strategies to avoid their violation, leads to a unified account of template filling processes in Moroccan Colloquial Arabic. Gafos's insightful work was the first to show how new generalizations from seemingly disparate surface phenomena in morphophonological processes (glide insertion, medial C duplication, stability of epenthetic schwa to speaking rate, geminate (in)separability) can be gained under the assumption that all these surface phenomena are the result of constraint interactions that govern the temporal organization of gestures.

Other theoretical approaches seeking a connection between gestural representations and grammatical processes have been presented by Beňuš and Gafos (Gafos and Beňuš 2006; Beňuš and Gafos 2007) in the context of Hungarian vowel harmony (see also Gafos 1999). Davidson's (2006a, 2006b) work on non-native cluster phonotactics and Jun's work on place assimilation (Jun 2004) should be mentioned here. Marin (2007) has presented a gestural model of stress-conditioned vowel alternations in Romanian. There have further been some recent studies on the temporal alignment of intonation to articulatory gestures (Gao 2008; Mücke *et al.* 2009).

7 Concluding remarks

Recent phonological theories have taken an integrated view on the relationship between phonetics and phonology: phonological constraints may express speakers' knowledge of the physics of speech production and perception. Going even further, gestural models have rigorously integrated articulatory timing in both the phonological and phonetic aspects of spoken language by assuming dynamically specified atoms of speech production. These approaches thereby explicitly model the temporal coordination of gestures, and thus allow us to formulate and empirically test hypotheses about the relation of the observable, physical principles of speech to underlying cognitive representations.

ACKNOWLEDGMENTS

The writing of the chapter was supported by the Deutsche Forschungsgemeinschaft (DFG PO 1269/1-1). Thanks to Mark Tiede for permitting the use of Figure 5.3. I am much indebted to Khalil Iskarous and Stefania Marin for many interesting discussions and helpful comments on an earlier version of the chapter.

REFERENCES

Abbs, James H. & Vincent L. Gracco. 1984. Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology* 51. 705–723.

Barbosa, Plínio A. 2002a. Explaining Brazilian Portuguese resistance to stress shift with a coupled-oscillator model of speech rhythm production. *Cadernos de Estudos Lingüísticos* 43. 71–92.

Barbosa, Plínio A. 2002b. Explaining cross-linguistic rhythmic variability via a coupled-oscillator model for rhythm production. *Speech Prosody 2002*, 163–166. Available (August 2010) at www.isca-speech.org/archive/sp2002/sp02_163.html.

Barry, Martin C. 1991. Temporal modeling of gestures in articulatory assimilation. *Proceedings of the 12th International Congress of Phonetic Sciences*, vol. 4, 14–17. Aix-en-Provence: Université de Provence.

Beňuš, Štefan & Adamantios I. Gafos. 2007. Articulatory characteristics of Hungarian "transparent" vowels. *Journal of Phonetics* 35. 271–300.

Bernstein, N. 1967. *The coordination and regulation of movements*. New York: Pergamon.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Browman, Catherine P. 1994. Lip aperture and consonant releases. In Patricia Keating (ed.) *Phonological structure and phonetic form: Papers in laboratory phonology III*, 331–353. Cambridge: Cambridge University Press.

Browman, Catherine P. & Louis Goldstein. 1985. Dynamic modeling of phonetic structure. In Victoria A. Fromkin (ed.) *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 35–53. Orlando: Academic Press.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Browman, Catherine P. & Louis Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45. 140–155.

Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6. 201–251.

Browman, Catherine P. & Louis Goldstein. 1990a. Representation and reality: Physical systems and phonological structure. *Journal of Phonetics* 18. 411–424.

Browman, Catherine P. & Louis Goldstein. 1990b. Tiers in articulatory phonology, with some implications for casual speech. In Kingston & Beckman (1990), 341–376.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180.

Browman, Catherine P. & Louis Goldstein. 1995a. In Robert F. Port & Timothy van Gelder (eds.) *Mind as motion: Explorations in the dynamics of cognition*, 175–194. Cambridge, MA: MIT Press.

Browman, Catherine P. & Louis Goldstein. 1995b. Gestural syllable position effects in American English. In Fredericka Bell-Berti & Lawrence Raphael (eds.) *Producing speech: Contemporary issues. For Katherine Safford Harris*, 19–33. Woodbury, NY: American Institute of Physics Press.

Browman, Catherine P. & Louis Goldstein. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5. 25–34.

Browman, Catherine P. & Louis Goldstein. 2001. Articulatory phonology. Unpublished ms., Yale University.

Butcher, Andrew. 2006. Australian Aboriginal languages: Consonant-salient phonologies and the "place of articulation" imperative. In Jonathan Harrington & Maria Tabain (eds.) *Speech production: Models, phonetic processes, and techniques*, 187–210. New York: Psychology Press.

Byrd, Dani. 1994. Articulatory timing in English consonant sequences. Ph.D. dissertation, University of California, Los Angeles.

Byrd, Dani. 1995. C-centers revisited. *Phonetica* 52. 285–306.

Byrd, Dani. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24. 209–244.

Byrd, Dani & Susie Choi. 2010. At the juncture of prosody, phonology, and phonetics: The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.) *Laboratory phonology 10*, 31–59. Berlin & New York: Mouton de Gruyter.

Byrd, Dani & Elliot L. Saltzman. 1998. Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26. 173–199.

Byrd, Dani & Elliot L. Saltzman. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31. 149–180.

Byrd, Dani, Stephen Tobin, Erik Bresch & Shrikanth Narayanan. 2009. Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics* 37. 97–110.

Chitoran, Ioana, Louis Goldstein & Dani Byrd. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology 7*, 419–447. Berlin & New York: Mouton de Gruyter.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Cole, Jennifer & José Ignacio Hualde (eds.) 2007. *Laboratory phonology* 9, 483–506. Berlin & New York: Mouton de Gruyter.

Connell, Bruce & Amalia Arvaniti (eds.) (1995). *Papers in laboratory phonology IV: Phonology and phonetic evidence*. Cambridge: Cambridge University Press.

Davidson, Lisa. 2006a. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34. 104–137.

Davidson, Lisa. 2006b. Phonotactics and articulatory coordination interact in phonology: Evidence from non-native production. *Cognitive Science* 30. 837–862.

Ellis, Lucy & W. J. Hardcastle. 2002. Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics* 30. 373–396.

Farnetani, Edda & M. Grazia Busà. 1994. Italian clusters in continuous speech. *Proceedings of the 3rd International Conference on Spoken Language Processing*, 359–361. Yokohama: Acoustical Society of Japan.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18. 7–44.

Fowler, Carol A. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8. 113–133.

Fowler, Carol A. 1984. Segmentation of coarticulated speech in perception. *Perception and Psychophysics* 36. 359–368.

Fowler, Carol A. 1986. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14. 3–28.

Fowler, Carol A. 1996. Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America* 99. 1730–1741.

Fowler, Carol A. & Elliot L. Saltzman. 1993. Coordination and coarticulation in speech production. *Language and Speech* 36. 171–195.

Fowler, Carol A. & Mary R. Smith. 1986. Speech perception as "vector analysis": An approach to the problems of invariance and segmentation. In Joseph S. Perkell & Dennis H. Klatt (eds.) *Invariance and variability in speech processes*, 123–139. Mahwah, NJ: Lawrence Erlbaum.

Gafos, Adamantios I. 1999. *The articulatory basis of locality in phonology*. New York: Garland.

Gafos, Adamantios I. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20. 269–337.

Gafos, Adamantios I. & Štefan Beňuš. 2006. Dynamics of phonological cognition. *Cognitive Science* 30. 905–943.

Galantucci, Bruno, Carol A. Fowler & Michael Turvey. 2006. The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review* 13. 361–377.

Gao, Man. 2008. Tonal alignment in Mandarin Chinese: An articulatory phonology account. Ph.D. dissertation, Yale University.

Gibbon, Fiona, W.J. Hardcastle & Katerina Nicolaidis. 1993. Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: A cross-linguistic investigation. *Language and Speech* 36. 261–277.

Gick, Bryan, Fiona Campbell, Sunyoung Oh & Linda Tamburri-Watt. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* 34. 49–72.

Giles, Stephen B. & Kenneth L. Moll. 1975. Cinefluorographic study of selected allophones of English /l/. *Phonetica* 31. 206–227.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Oxford & Cambridge, MA: Blackwell.

Goldstein, Louis & Catherine P. Browman. 1986. Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics* 14. 339–342.

Goldstein, Louis & Carol A. Fowler. 2003. Articulatory Phonology: A phonology for public language use. In Niels O. Schiller & Antje S. Meyer (eds.) *Phonetics and phonology in language comprehension and production: Differences and similarities*, 159–207. Berlin & New York: Mouton de Gruyter.

Goldstein, Louis, Dani Byrd & Elliot L. Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael A. Arbib (ed.) *Action to language via the mirror neuron system*, 215–249. Cambridge: Cambridge University Press.

Goldstein, Louis, Ioana Chitoran & Elisabeth Selkirk. 2007a. Syllable structures as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. In Trouvain & Barry (2007), 241–244.

Goldstein, Louis, Marianne Pouplier, Larissa Chen, Elliot L. Saltzman & Dani Byrd. 2007b. Dynamic action units slip in speech production errors. *Cognition* 103. 386–412.

Gracco, Vincent L. 1988. Timing factors in the coordination of speech movements. *Journal of Neuroscience* 8. 4628–4634.

Guenther, Frank H., Michelle Hampson & David Johnson. 1998. A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review* 105. 611–633.

Hardcastle, W. J. 1976. *Physiology of speech production: An introduction for speech scientists.* London: Academic Press.

Harrington, Jonathan, Janet Fletcher & Corinne Roberts. 1995. Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data. *Journal of Phonetics* 23. 305–322.

Hawkins, Sarah. 1999. Reevaluating assumptions about speech perception: interactive and integrative theories. In J. M. Pickett (ed.) *The acoustics of speech communication: Fundamentals, speech perception theory, and technology*, 232–288. Boston: Allyn & Bacon.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology.* Cambridge: Cambridge University Press.

Hermes, Anne, Martine Grice, Doris Mücke & Henrik Niemann. 2008. Articulatory indicators of syllable affiliation in word initial consonant clusters in Italian. In Rudolph Sock, Susanne Fuchs & Yves Laprie (eds.) *Proceedings of the International Seminar of Speech Production* 8, 433–436.

Holst, Tara & Francis Nolan. 1995. The influence of syntactic structure on [s] to [ʃ] assimilation. In Connell & Arvaniti (1995), 315–333.

Honorof, Douglas. 1999. Articulatory gestures and Spanish nasal assimilation. Ph.D. dissertation, Yale University.

Honorof, Douglas & Catherine P. Browman. 1995. The center or the edge: How are consonant clusters organized with respect to the vowel? In Kjell Elenius & Peter Branderud (eds.) *Proceedings of the 13th International Congress of the Phonetic Sciences*, vol. 4, 552–555. Stockholm: KTH & Stockholm University.

Hoole, Philip. 2006. Experimental studies of laryngeal articulation. Habilitationsschrift, Ludwig-Maximilians-Universität, Munich.

Hoole, Philip & Christine Mooshammer. 2000. Articulatory analysis of the German vowel system. In Peter Auer, Peter Gilles & Helmut Spiekermann (eds.) *Silbenschnitt und Tonakzente*, 129–151. Tübingen: Niemeyer.

Huffman, Marie K. & Rena A. Krakow (eds.) 1993. *Nasals, nasalization, and the velum.* Orlando: Academic Press.

Hume, Elizabeth & Keith Johnson (eds.) 2001. *The role of speech perception in phonology.* San Diego: Academic Press.

Hura, Susan L., Björn Lindblom & Randy L. Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35. 59–72.

Iskarous, Khalil. 2005. Patterns of tongue movement. *Journal of Phonetics* 33. 363–381.

Jun, Jongho. 1995. Perceptual and articulatory factors in place assimilation: An optimality theoretic approach. Ph.D. dissertation, University of California, Los Angeles.

Jun, Jongho. 2004. Place assimilation. In Hayes *et al.* (2004), 58–86.

Jusczyk, Peter W. & Paul A. Luce. 2002. Speech perception and spoken word recognition: Past and present. *Ear and Hearing* 23. 2–40.

Kay, Bruce. 1988. The dimensionality of movement trajectories and the degrees of freedom problem: A tutorial. *Human Movement Science* 7. 343–364.

Keating, Patricia A., Björn Lindblom, James Lubker & Jody Kreiman. 1994. Variability in jaw height for segments in English and Swedish VCVs. *Journal of Phonetics* 22. 407–422.

Keating, Patricia, Taehong Cho, Cécile Fougeron & Chai-Shune Hsu. 2003. Domain-initial articulatory strengthening in four languages. In John Local, Richard Ogden & Rosalind Temple (eds.) *Phonetic interpretation: Papers in laboratory phonology VI*, 145–163. Cambridge: Cambridge University Press.

Kelso, J. A. Scott. 1995. *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Kelso, J. A. Scott & Betty Tuller. 1984. A dynamical basis for action systems. In M. S. Gazzaniga (ed.) *Handbook of cognitive neuroscience*, 321–356. New York: Plenum Press.

Kelso, J. A. Scott, Betty Tuller, Eric Vatikiotis-Bateson & Carol A. Fowler. 1984. Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance* 10. 812–832.

Kelso, J. A. Scott, Elliot L. Saltzman & Betty Tuller. 1986. The dynamical perspective on speech production: Data and theory. *Journal of Phonetics* 14. 29–59.

Keyser, Samuel J. & Kenneth N. Stevens. 1994. Feature geometry and the vocal tract. *Phonology* 11. 207–236.

Kingston, John & Mary E. Beckman (eds.) 1990. *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press.

Kochetov, Alexei. 2006. Syllable position effects and gestural organization: Articulatory evidence from Russian. In Louis Goldstein, Douglas Whalen & Catherine T. Best (eds.) *Laboratory phonology 8*, 565–588. Berlin & New York: Mouton de Gruyter.

Kochetov, Alexei & Marianne Pouplier. 2008. Phonetic variability and grammatical knowledge: An articulatory study of Korean place assimilation. *Phonology* 25. 399–431.

Kochetov, Alexei, Marianne Pouplier & Minjung Son. 2007. Cross-language differences in overlap and assimilation patterns in Korean and Russian. In Trouvain & Barry (2007), 1361–1364.

Krakow, Rena A. 1993. Nonsegmental influences on velum movement patterns: Syllables, sentences, stress and speaking rate. In Huffman & Krakow (1993), 87–116.

Krakow, Rena A. 1999. Physiological organization of syllables: A review. *Journal of Phonetics* 27. 23–54.

Kroos, Christian, Philip Hoole, Barbara Kühnert & Hans Tillmann. 1997. Phonetic evidence for the phonological status of the tense–lax distinction in German. *Forschungsberichte des Instituts für Phonetik und sprachliche Kommunikation der Universität München* 35. 17–25.

Kugler, Peter N. & Michael T. Turvey. 1987. *Information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, NJ: Lawrence Erlbaum.

Kühnert, Barbara & Philip Hoole. 2004. Speaker-specific kinematic properties of alveolar reductions in English and German. *Clinical Linguistics and Phonetics* 18. 559–575.

Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT.

Liberman, Alvin. 1997. *Speech: A special code*. Cambridge, MA: MIT Press.

Liberman, Alvin & Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21. 1–36.

Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35. 1773–1781.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds.) *Speech production and speech modeling*, 403–439. Dordrecht: Kluwer.

Löfqvist, Anders & Vincent L. Gracco. 1999. Interarticulator programming in VCV sequences: Lip and tongue movements. *Journal of the Acoustical Society of America* 105. 1864–1876.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Marin, Stefania. 2007. Vowel to vowel coordination, diphthongs and Articulatory Phonology. Ph.D. dissertation, Yale University.

Marin, Stefania & Marianne Pouplier. 2010. Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control* 14. 380–407.

Mooshammer, Christine, Philip Hoole & Anja Geumann. 2007. Jaw and order. *Language and Speech* 50. 145–176.

Mücke, Doris, Martine Grice, Johannes Becker & Anne Hermes. 2009. Sources of variation in tonal alignment: Evidence from acoustic and kinematic data. *Journal of Phonetics* 37. 321–338.

Munhall, Kevin, Carol A. Fowler, Sarah Hawkins & Elliot L. Saltzman. 1992. "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics* 20. 225–239.

Nam, Hosung. 2006. Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. In Cole & Hualde (2007), 483–506.

Nam, Hosung. 2007. Gestural coupling model of syllable structure. Ph.D. dissertation, Yale University.

Nam, Hosung & Elliot L. Saltzman. 2003. A competitive, coupled oscillator model of syllable structure. In M. J. Solé, D. Recasens & J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*. 2253–2256. Barcelona: Causal Productions.

Nam, Hosung, Louis Goldstein, Elliot L. Saltzman & Dani Byrd. 2004. Tada: An enhanced, portable task dynamics model in Matlab. *Journal of the Acoustical Society of America* 115. 2430.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Nolan, Francis, Tara Holst & Barbara Kühnert. 1996. Modelling [s] to [ʃ] accommodation in English. *Journal of Phonetics* 24. 113–137.

Öhman, Sven E. G. 1966. Numerical model of coarticulation. *Journal of the Acoustical Society of America* 41. 310–320.

Padgett, Jaye. 1994. Stricture and nasal place assimilation. *Natural Language and Linguistic Theory* 12. 465–513.

Perkell, Joseph S. 1969. *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge, MA: MIT Press.

Pierrehumbert, Janet B. 1990. Phonological and phonetic representation. *Journal of Phonetics* 18. 375–394.

Pouplier, Marianne. 2008. The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics* 36. 114–140.

Pouplier, Marianne & Louis Goldstein. 2010. Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production. *Language and Cognitive Processes* 25. 616–649.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Recasens, Daniel & Maria Dolors Pallarès. 2001. Coarticulation, assimilation and blending in Catalan consonant clusters. *Journal of Phonetics* 29. 273–301.

Restle, David. 2003. *Silbenschnitt – Quantität – Kopplung: Zur Geschichte, Charakterisierung und Typologie der Anschlußperiode*. Munich: Willhelm Fink Verlag.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.

Saltzman, Elliot L. 1979. Levels of sensorimotor representation. *Journal of Mathematical Psychology* 20. 91–193.

Saltzman, Elliot L. 1986. Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (eds.) *Generation and modulation of action patterns*, 129–144. New York: Springer.

Saltzman, Elliot L. 1991. The task dynamic model in speech production. In H. F. M. Peters, W. Hulstijn & C. W. Starkweather (eds.) *Speech, motor control and stuttering*, 37–52. Amsterdam: Elsevier Science Publishers.

Saltzman, Elliot L. 1995. Dynamics and coordinate systems in skilled sensorimotor activity. In Robert F. Port & Timothy van Gelder (eds.) *Mind as motion: Dynamics, behavior, and cognition*, 149–173. Cambridge, MA: MIT Press.

Saltzman, Elliot L. & J. A. Scott Kelso. 1987. Skilled actions: A task-dynamic approach. *Psychological Review* 94. 84–106.

Saltzman, Elliot L. & Kevin G. Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1. 333–382.

Saltzman, Elliot L., Hosung Nam, Jelena Krivokapic & Louis Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In Plínio A. Barbosa, Sandra Madureira & Cesar Reis (eds.) *Speech Prosody 2008*, 175–184. Available (August 2010) at www.isca-speech.org/archive/sp2008/sp08_175.html.

Sawashima, Masayuki & Hajime Hirose. 1980. Laryngeal gestures in speech production. *Annual Bulletin of the Research Institute of Logopedics and Phonoiatrics, Faculty of Medicine University of Tokyo* 14. 29–51.

Scobbie, James M. & Marianne Pouplier. 2010. The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction of word-final English /l/. *Journal of Phonetics* 38. 240–259.

Selkirk, Elisabeth. 1982. The syllable. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*. Part II, 337–383. Dordrecht: Foris.

Shaw, Jason, Adamantios I. Gafos, Philip Hoole & Chakir Zeroual. 2009. Temporal evidence for syllabic structure in Moroccan Arabic: Data and model. *Phonology* 26. 187–215.

Smith, Caroline L. 1995. Prosodic patterns in the coordination of vowel and consonant gestures. In Connell & Arvaniti (1995), 205–222.

Son, Minjung, Alexei Kochetov & Marianne Pouplier. 2007. The role of gestural overlap in perceptual place assimilation in Korean. In Cole & Hualde (2007), 507–534.

Sproat, Richard & Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21. 291–311.

Steriade, Donca. 1993. Closure, release, and nasal contours. In Huffman & Krakow (1993), 401–470.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Hume & Johnson (2001), 219–250.

Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Studies in honor of Paul Kiparsky*, 151–179. Cambridge, MA: MIT Press.

Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17. 3–45.

Stone, Maureen, Melissa A. Epstein & Khalil Iskarous. 2004. Functional segments in tongue movement. *Clinical Linguistics and Phonetics* 18. 507–521.

Strogatz, Steven H. & Ian Stewart. 1993. Coupled oscillators and biological synchronization. *Scientific American*. December. 102–109.

Studdert-Kennedy, Michael & Louis Goldstein. 2003. Launching language: The gestural origin of discrete infinity. In Morten Christiansen & Simon Kirby (eds.) *Studies in the evolution of language*, 235–254. New York: Oxford University Press.

Surprenant, Aimée M. & Louis Goldstein. 1998. The perception of speech gestures. *Journal of the Acoustical Society of America* 104. 518–529.

Tabain, Marija, Gavan Breen & Andrew Butcher. 2004. VC vs. CV syllables: A comparison of Aboriginal languages with English. *Journal of the International Phonetic Association* 34. 175–200.

Tiede, Mark K., Joseph S. Perkell, Majid Zandipour & Melanie Matthies. 2001. Gestural timing effects in the "perfect memory" sequence observed under three rates by electromagnetometry. *Journal of the Acoustical Society of America* 110. 2657.

Tilsen, Sam. 2009. Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science* 33. 839–879.

Trouvain, Jürgen & William J. Barry (eds.) 2007. *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken: Saarland University.

Turvey, Michael T. 1977. Preliminaries to a theory of action with reference to vision. In Robert Shaw & John Bransford (eds.) *Perceiving, acting, knowing*, 211–265. Mahwah, NJ: Lawrence Erlbaum.

Turvey, Michael T. 1990. Coordination. *American Psychologist* 45. 938–953.

Turvey, Michael T., Robert E. Shaw & William Mace. 1978. Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions. In Jean Requin (ed.) *Attention and performance VII*, 557–595. Hillsdale, NJ: Lawrence Erlbaum.

Vennemann, Theo. 1991. Syllable structure and syllable cut prosodies in modern standard German. In Pier Marco Bertinetto, Michael Kenstowicz & Michele Loporcaro (eds.) *Certamen Phonologicum II: Papers from the 1990 Cortona Phonology Meeting*, 211–243. Turin: Rosenberg & Sellier.

Yanagawa, Mariko. 2006. Articulatory timing in first and second language: A cross-linguistic study. Ph.D. dissertation, Yale University.

Zsiga, Elizabeth C. 1995. An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. In Connell & Arvaniti (1995), 282–302.

Zsiga, Elizabeth C. 2000. Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics* 28. 69–102.

Zsiga, Elizabeth C. 2009. External sandhi in a second language: The phonetics and phonology of obstruent nasalization in Korean-accented English. Unpublished ms., Georgetown University.

49 Sonority

Steve Parker

1 Introduction

If an interface between phonetics and phonology really exists (*pace* Ohala 1990b), then one topic having a long and controversial history in that domain is sonority. Sonority can be defined as a unique type of relative, *n*-ary (non-binary) feature-like phonological element that potentially categorizes all speech sounds into a hierarchical scale. For example, vowels are more sonorous than liquids, which are higher in sonority than nasals, with obstruents being the least sonorous of all segments. In terms of traditional phonetic systems for categorizing natural classes of sounds, then, the feature encoded by sonority most closely corresponds to the notion *manner of articulation* (see CHAPTER 13: THE STRICTURE FEATURES). In this sense, sonority is like most other features: it demarcates groups of segments that behave similarly in cross-linguistically common processes. At the same time, however, sonority is unlike most features in that it exhaustively encompasses all speech sounds simultaneously, i.e. every type of segment has some inherent incremental value for this feature. Sonority is also unique in that it has never been observed to spread (assimilate), in and of itself.

A major function of sonority is to organize (order) segments within syllables. Specifically, more sonorous sounds, such as vowels, tend to occur in the nucleus, while less sonorous sounds normally appear in the marginal (non-peak) positions – onsets and codas. This concept has engendered several chronic and frequently discussed research questions:

(1) a. What role, if any, does sonority play in Universal Grammar?
    b. How many and what kinds of natural class distinctions need to be made in the sonority hierarchy?
    c. Are its rankings fixed or permutable (reversible)?
    d. Which distinctions in the sonority scale, if any, are universal and which, if any, are language-particular?
    e. Is sonority an abstract phonological mechanism only, or does it also have a consistent, measurable phonetic basis?

To answer (1e) briefly, the main acoustic correlate of sonority is intensity. As Ladefoged (1975: 219) notes, "The sonority of a sound is its loudness relative to that of other sounds with the same length, stress, and pitch." Nevertheless, although much progress has been made in addressing the issues in (1), little consensus has emerged in understanding many of them. This chapter touches on each of the questions in (1), although not necessarily in the same order or to the same degree. The goal is to summarize the debates and document the types of empirical data that have been presented in arguing for the different positions. This chapter is organized as follows: §2 reviews the cross-linguistic phonological evidence for sonority. Thus §2.1 discusses the Sonority Sequencing Principle, §2.2 Minimum Sonority Distance effects, §2.3 the Syllable Contact Law, and §2.4 the contribution of sonority to the relative weight of the rhyme. §3 describes the Sonority Dispersion Principle, while §4 presents several desirable characteristics that a complete sonority hierarchy should ideally display. Finally, §5 examines the physical basis of sonority, as demonstrated experimentally.

2 Phonological evidence for sonority

This section describes various phonological phenomena that demonstrate that sonority is active in many languages. The exposition summarizes works discussing the issues in more depth, such as Parker (2002) and Cser (2003). Phonotactic constraints and morphophonemic alternations provide the most compelling evidence for establishing the divisions in the sonority hierarchy. Consequently, most of the argumentation here relies on these two factors. Several patterns are attested in enough languages to motivate the hypothesis that some notion of sonority should be considered part of Universal Grammar (UG; the innate linguistic faculty shared by all humans; Chomsky and Halle 1968; Kenstowicz 1994). How sonority is best expressed in UG is a separate topic, not discussed in detail here. In many cases, opposing points of view exist, with some linguists denying that sonority is actually involved in these phenomena. See e.g. Ohala (1974, 1990a, 1990b) and Kawasaki (1982) for arguments against appealing to sonority as an explanation for these data.

2.1 The Sonority Sequencing Principle

The domain in which sonority is most often invoked is the syllable, and related notions such as permissible consonant clusters in onset or coda position. This reflects the analogy that the syllable is like a wave of energy (Sievers 1893; Pike 1943). Specifically, syllables universally tend to abide by the following constraint:

(2) Every syllable exhibits exactly one peak of sonority, contained in the nucleus.

This is known as the Sonority Sequencing Principle (SSP) or the Sonority Sequencing Generalization. Key works assuming this principle as a basis for analysis include Hooper (1976), Selkirk (1984), Blevins (1995), and, in Optimality Theory, Cho and King (2003) and Zec (2007). Ohala (1990a, 1990b) and Wright (2004) note that a rudimentary notion of the SSP is observed in the work of de Brosses (1765). For the purpose of formally encoding and testing the SSP, the most frequently cited sonority scale is the following:

Table 49.1 Cross-linguistic variation in syllabic segments based on sonority

	Vowels	*Liquids*	*Nasals*	*Obstruents*
Bulgarian, Hawaiian, Kabardian, Latin, Spanish	✓	–	–	–
Lendu, Sanskrit, Slovak	✓	✓	–	–
English	✓	✓	✓	–
(Central) Carrier, (Imdlawn) Tachelhit (Berber)	✓	✓	✓	✓

(3) *Modal sonority hierarchy* (e.g. Clements 1990; Kenstowicz 1994; Smolensky 1995)

vowels > glides > liquids > nasals > obstruents

⟵⟶

higher in sonority   lower in sonority

In terms of sonority, the five natural classes in (3) are the easiest ones to motivate and the most useful ones to employ. Assuming the SSP and the hierarchy in (3), hypothetical syllables like [ta], [kɾu], [wos], and [pʰlænt] are well formed, since their sonority slope uniformly rises from the beginning of the syllable to the nuclear vowel, and falls from the nucleus to the end of the syllable. Conversely, syllables containing "sonority reversals" such as [lpa] and [odm] violate (2) and are therefore illicit in most languages.

One argument for the SSP is that cross-linguistically the inventory of [+syllabic] segments in particular languages normally forms a continuous range based on a scale like (3). Thus, the propensity for a sound to occur in nuclear position is correlated with how sonorous it is. The typology of permissible syllabic segments across languages is illustrated in Table 49.1, adapted from Blevins (1995) and Zec (2007).[1] The generalization is that if a language permits syllabic segments from a lower sonority class, it also allows nuclei from all higher sonority classes. In Tachelhit even voiceless stops occur in nuclear position. However, glides are omitted here since by definition they are non-nuclear. The following example lists forms containing syllabic consonants from two of these languages, where [.] marks a syllable boundary (Parker 2002; Zec 2007; Ridouane 2008).[2]

(4) a. *Slovak*

[kr̩.vi] 'blood'
[vl̩.ka] 'wolf'

b. *Tachelhit*

[tz̩.dm̩t] 'gather wood'
[tr̩.gl̩t] 'lock'
[tʃ̩.tk̩.ts̩tt] 'you sprained it (FEM)'

Nevertheless, while the pattern in Table 49.1 is a strong tendency, it is not universally obeyed. For example, many languages (especially in Africa) attest syllabic

[1] Language names and genetic affiliations follow the *Ethnologue* (Lewis 2009). In the online version of this chapter, the appendix provides more details about the languages cited here: country, linguistic phylum, primary source of data, etc.

[2] The online version of this chapter contains more illustrative data throughout.

nasals but not syllabic liquids: e.g. Djeebbana, Lele (Chad), and Swahili (Blevins 2006). Thus, factors other than sonority must also be appealed to in some cases.

Besides syllabic consonants, another reason to adopt the SSP is that it accounts for tautosyllabic consonant clusters in most languages. The following example illustrates three languages that strictly follow the SSP in onsets and codas (Blevins 2006):

(5) a. *Cheke Holo*
[kai.ka.fl.i] 'flash on and off'
[kmai.kma.ji] 'eat a varied meal (reduplicated)'

b. *Djeebbana*
[n̩.ka.la] 'fork'
[kalk.bet] 'northern black wallaroo'

c. *Spanish*
[plan] 'plan'
[trans.kri.βir] 'to transcribe'

Again there are exceptions; (6) shows data from two languages in which the underlined consonant clusters apparently violate the SSP (Blevins 2006):

(6) a. *Leti* (Indonesia)
[pni.nu] 'fool'
[sra:t] 'main road'
[rka:.lu] 'they shout'
[rstp.le] 'they sail'

b. *Yir Yoront*
[melt] 'animal, bird'
[patl] 'clean, bald'

Counterexamples to the SSP also occur in some Indo-European languages, such as Czech, Romanian, and Russian. Extreme cases are found in Georgian, a Kartvelian language attesting the word-initial clusters /zrd/, /mkrt/, /msxv/ and /mt͡svrtn/ (Blevins 2006). Such exceptions have led some researchers to conclude that analyses based on sonority are circular in nature (Ohala 1990b: 160). However, no studies exist in which the proportion of languages with sonority reversals is tabulated among a statistically reliable and balanced sample. Therefore, based on available data, it seems safe to conclude that a large percentage of the world's languages do conform to the SSP. Furthermore, purported counterexamples like Georgian /mt͡svrtn/ are dubious if the onset cluster in question occurs only word-initially, but not word-internally. This is crucial since, in a given language, a consonant cluster should appear in a position other than at a word edge in order to count as a canonical syllable type. Otherwise, when a greater number of consonants show up next to a word boundary, it is debatable whether this constitutes a true syllable margin. A more principled explanation is to analyze the SSP-violating segment(s) as a degenerate syllable or an extrasyllabic appendix licensed by the prosodic word. See Cho and King (2003) for further discussion. Finally, some alleged SSP violations cannot withstand further scrutiny. For example, Blevins (2006) lists Leti [rka:lu] in (6) above. However, van Engelenhoven (2004) states that word-initially before another consonant the trilled /r/, nasals, /l/ and /s/ are lengthened and "syllabic." Thus, a more accurate transcription of this word is [r̩.ka:.lu]. Since the /r/ and the /k/

Table 49.2 Typological range of languages containing sC clusters

s +	*Spanish*	*French, Western Keres*	*Greek*	*English*	*Dutch*	*German*	*Russian*
stop	–	✓	✓	✓	✓	✓	✓
fricative	–	–	✓	–	✓	–	✓
nasal	–	–	(–)	✓	✓	✓	✓
lateral	–	–	–	✓	✓	✓	✓
rhotic	–	–	–	–	(–)	✓	✓

are not tautosyllabic, this is not a counterexample to the SSP. Rather, it confirms it.

Cross-linguistically, the most frequent exceptions to the SSP involve initial /s/ followed by a plosive, as in the English words *spill*, *still*, and *skill*. Morelli (2003) and Goad (CHAPTER 38: THE REPRESENTATION OF SC CLUSTERS) focus on this phenomenon. Table 49.2 is adapted from the latter.

Summarizing this table, Goad observes that the lower a consonant is in sonority, the more preferred it is after an initial /s/. She thus posits that s + stop > s + nasal > s + lateral > s + rhotic, where ">" = "is more harmonic than." This scale (minus the s) follows many sonority hierarchies that posit more natural class distinctions than (3), such as the maximally detailed scale in §4.

Morelli (2003) reaches a similar conclusion. She notes that many languages have onset cluster inventories comprising three main types: (1) stop + sonorant, (2) fricative + sonorant and (3) fricative + stop. These first two satisfy the SSP, while the third reverses it (assuming that fricatives are more sonorous than stops; see below). Illustrative languages include Haida, Hindi, Hungarian, Isthmus Zapotec, Italian, Mohave, Swedish, Telugu, Yecuatla Totonac, and Yuchi. To her knowledge, however, no language exists that is analogous to these, yet completely follows the SSP: hypothetically, (1) stop + sonorant, (2) fricative + sonorant, and, crucially, (3) stop + fricative (not counting affricates). Consequently, she posits that among onset clusters consisting of two obstruents, the unmarked type is fricative + stop, where "unmarked" = phonologically default and most common (Kenstowicz and Kisseberth 1973: 3; de Lacy 2006).

Summarizing thus far, the SSP is a strong universal tendency but has exceptions. In some cases the reversals in the sonority slope are of the largest possible degree: an onset consisting of a glide followed by a voiceless stop. To illustrate, Santa María Quiegolani Zapotec exhibits many words like the following (Regnier 1993):

(7) [wkit] 'game'
[wtóːʔ] 'sell (COMPLETIVE)'
[jkǎ] 'buy (POTENTIAL)'

Nevertheless, typological generalizations can still be made. For example, most languages have more consonant cluster types and tokens obeying the SSP than violating it. Furthermore, more languages attest obstruent + liquid (OL) onset clusters, for instance, than the opposite (LO). This can be stated even more forcefully as an implicational universal: if a language allows complex onsets of the type

LO, it must permit OL clusters as well, whereas the inverse is not necessarily true (Greenberg 1978). However, this kind of observation cannot be extended to include all possible natural class combinations. For instance, Texistepec Popoluca does not permit obstruent + nasal (ON) syllable initially, yet it does allow NO clusters: [mbak] 'my bone' (Wichmann 2002).[3] Consequently, absolute claims about the SSP tend to break down given enough languages. Nevertheless, one apparently exceptionless statement is the following:

(8) No language exists in which all tautosyllabic consonant clusters reverse the SSP.

Returning to the five-category sonority hierarchy in (3), many phonologists expand this by making subdivisions within three of the natural classes: vowels, liquids, and obstruents (cf. (27) in §4). For example, fricatives are often claimed to be more sonorous than stops (Hankamer and Aissen 1974; Steriade 1982, 1988; Kager 1999). To illustrate, in Sanskrit reduplication, when a verb base begins with a consonant cluster, the prefix retains the less sonorous of these two sounds. Thus, in (9), when a stop and a continuant are adjacent in either order, the reduplicant invariably surfaces with the stop:

(9) *Sanskrit* (from Whitney 1889)

/praʧʰ/	→	[pa-praʧʰ]	'ask'
/swar/	→	[sa-swar]	'sound'
/tsar/	→	[ta-tsar]	'approach stealthily'
/stʰaː/	→	[ta-stʰaː]	'stand'
/tjaʤ/	→	[ta-tjaʤ]	'forsake'
/ʃratʰ/	→	[ʃa-ʃratʰ]	'slacken'
/druw/	→	[du-druw]	'run'
/mluʧ/	→	[mu-mluʧ]	'set'
/rdʰ/	→	[aːr-di-dʰam][4]	

In (9) the obvious generalization is that the reduplicant copies the less sonorous consonant from the onset of the base, regardless of its relative position within the cluster. Otherwise, if all obstruents are equal in sonority, the analysis of this process is more complicated to express (Benua 1997; Hironymous 1999). For further data and discussion of Sanskrit reduplication, see CHAPTER 119: REDUPLICATION IN SANSKRIT. While the full details are complex, the pattern whereby the least sonorous segment emerges in the prefix is very regular. For a mathematical explanation of this effect, see §3.

When underlying representations juxtapose sounds violating the SSP, these are repaired in four different ways cross-linguistically: (1) vowel epenthesis, (2) deletion, (3) syllabic consonants, and (4) metathesis. First, a vowel can be inserted to rescue the unsyllabifiable consonant, a process called stray epenthesis (Itô 1986). This occurs in Serbo-Croatian (Kenstowicz 1994):

[3] Prenasalized stops (common in African languages) do not violate the SSP, since they are single phonemic units, not true sequences. Syllabic nasals, such as in hypothetical [n̩.da], do not constitute tautosyllabic onsets either.

[4] Whitney does not gloss this root, but notes that the form is aorist.

(10)

		masculine	*neuter*	
	a.	pust	pusto	'empty'
		zelen	zeleno	'green'
	b.	dobar	dobro	'good'
		jasan	jasno	'clear'
	c.	bogat	bogato	'rich'
		kriʒan	kriʒano	'cross'

In the adjective paradigm in (10), the neuter is marked by the suffix /-o/. The masculine column displays no overt morphological marking. In (10b), [a] alternates with Ø. This vowel surfaces phonetically in the final syllable of the masculine forms, between the last two consonants, but is absent in the neuter column. The contrasting forms in (10c) contain an [a] in the second syllable in both columns. This demonstrates that the alternation in (10b) involves epenthesis of [a] in the masculine forms, not syncope of underlying /a/ in the neuter column. The underlying representations of the roots in (10b) are /dobr/ and /jasn/. These underlying representations end with a consonant cluster consisting of an obstruent followed by a sonorant. If these were syllabified directly into a complex coda, they would violate the SSP. In contrast, the root /pust/ in (10a) ends with a cluster in which sonority falls. Therefore stray epenthesis is not needed since the sequence /pust/ can be exhaustively syllabified while respecting the SSP. The sonority profile of two of these contrasting roots is displayed in the following metrical-like grids (Jespersen 1904; Zec 1988; Clements 1990; Kenstowicz 1994):

(11)

vowel		*					*		
glide		*					*		
liquid		*					*		*
nasal		*					*		*
fricative		*	*				*		*
stop	*	*	*	*		*	*	*	*
	p	u	s	t		d	o	b	r

These grids employ the five-category sonority hierarchy in (3), supplemented by obstruents being split between fricatives and stops, motivated by the Sanskrit data in (9). As these figures show, the morpheme /pust/ in isolation (in the masculine column) contains one peak of sonority (the /u/), whereas /dobr/ contains two (the /o/ and the /r/). Consequently, the motivation for inserting a vowel in the second case ([dobar]) is to rescue the /r/, which cannot be incorporated into the same syllable as the /b/ without violating the SSP.

A second process used to fix SSP violations is the deletion of an unlicensed (unsyllabifiable) consonant, known as stray erasure (Itô 1986). This process is illustrated by Ancient Greek. The following data show that complex onset and coda clusters are permitted, including word-medially (Steriade 1982; Kenstowicz 1994):

(12)

	klep^h	'to steal'
	smerd.nos	'power, force'
	am.blus	'dull'

as.tron	'star'
a.elp.tos	'unhoped for'
t^{h}elk.tron	'charm'
pemp.tos	'sent'

The form [t^{h}elk.tron] demonstrates that up to four consonants can be concatenated intervocalically, providing the SSP is respected. However, in the reduplicated form /CV-grap^h-st^{h}ai/ → [gegrap^{h}t^{h}ai] 'to have been written', the underlying /s/ at the beginning of the infinitival suffix occurs between two stops. If this word were assigned a sonority profile as in (11), the /s/ would constitute a peak of sonority. However, this /s/ is not syllabic, nor can it be incorporated into a syllable with the preceding /p^h/ or the following /t^h/ without violating the SSP. Consequently, since it is prosodically unparsable, it is elided. According to Steriade (1982) and Itô (1986), this phenomenon is a default universal mechanism automatically applying at the end of the derivation to clean up any remaining problems (see CHAPTER 68: DELETION).

A third strategy for dealing with SSP violations is to simply retain the offending consonant, in which case it is automatically realized phonetically as syllabic. English illustrates this with unstressed sonorant consonants in word-final clusters: *prism, button, pickle, manner*. Another example is Chamicuro (Parker 1989):

(13)	[w-usm-i]	'I sing'	[w-usm̩-kati]	'I sang'
	1SG-sing-EPENTHETIC		1SG-sing-PAST	

Fourth, and most rarely, SSP violations are resolved by metathesis. The most convincing case of this to date is Western Farsi. When a final vowel is deleted by a general process of apocope, an obstruent or nasal in a potential coda cluster metathesizes with a following liquid. Otherwise (without metathesis), the final liquid would constitute a separate sonority peak, which this language does not allow (Hock 1985):

(14)	ʧaxra	→	ʧarx	'wheel'	
	suxra	→	surx	'red'	(cf. suhr-ab 'ruddy goose')
	vafra	→	barf	'snow, ice'	
	asru	→	ars	'tear'	
	vazra	→	gurz	'club'	
	*namra[5]	→	narm	'soft'	

Hock (1985) attributes this alternation to the SSP. However, from his description this is primarily a historical process, so it may no longer be synchronically active.

2.2 *Minimum Sonority Distance*

While the SSP rules out many of the prohibited syllable types in most languages, it is not the full story. For example, the three syllables [kna], [kla], and [kwa] equally

[5] The form /namra/ is preceded by *, since it is reconstructed.

Table 49.3 Minimal Sonority Distance language types

	Maximal inventory of permissible onset clusters	*Languages*
MSD = 0	OO, ON, OL, OG, NN, NL, NG, LL, LG, GG	Bulgarian, Leti
MSD = 1	ON, OL, OG, NL, NG, LG	Chukchee
MSD = 2	OL, OG, NG	Cizrra, Kurdish, Spanish
MSD = 3	OG	Mono, Panobo, Japanese (?), Mandarin Chinese (?)

satisfy the SSP. Nevertheless, although many languages permit onset clusters such as [kl] and/or [kw], syllables like [kna] are much less common. One explanation for this asymmetry is a language-specific parametric requirement that the members of a tautosyllabic consonant cluster be separated by a minimum number of ranks on the sonority scale (Steriade 1982; Selkirk 1984). For example, /k/ and /l/ are sufficiently distinct in relative sonority and may therefore be combined. However, /k/ and /n/ are too close along this scale, and this is not tolerated in many languages. Conversely, a language like Russian, which permits words like /kniga/ 'book', has a lower threshold on this parameter. This condition is captured by the following principle:

(15) *Minimal Sonority Distance* (MSD)

Given an onset composed of two segments, C_1 and C_2, if a = Sonority Index of C_1 and $b = \mathrm{SI}(C_2)$, then $b - a \geq x$, where $x \in \{0, 1, 2, 3\}$.

Assuming the sonority hierarchy in (3), the typology of possible languages shown in Table 49.3 is generated (cf. Zec 2007). The generalization is that if a language permits clusters with a lower sonority distance, it allows clusters of all higher sonority distances as well, *ceteris paribus*. The inverse of this is not true. The reversed counterparts of these onsets, such as *LO, can be excluded by the independently motivated SSP when necessary (§2.1). The data in (16) illustrate typical consonant sequences from each of the four language types in Table 49.3. Naturally, not every cluster type is fully productive for all phoneme combinations in these languages. Nevertheless, enough representative examples occur to justify the general trends.

(16) a. *Leti* (van Engelenhoven 2004)

[ptu.na]	'star'
[tmu.ra]	'tin'
[kru.ki]	'crab (sp.)'
[m̩.kwo.ri]	'you (SG) lift'

b. *Chukchee* (Kämpfe and Volodin 1995)

[plətkuk]	'end, finish, conclude'
[qlikkin]	'twenty'
[tretʃejwaɣʔe]	'I will go'
[ljur]	'suddenly'

c. *Gizrra* (van Bodegraven and van Bodegraven 2005)
[glɛs] 'dew'
[ta.praz.də] 'on (his) fangs'
[djao] 'palm (sp.)'
[ur.mjao] 'tree (sp.)'

d. *Panobo* (Parker 1992)
[hwĩn.ti] 'heart'
[βwi.ni.kæ̃j] 'they are taking, carrying'
[pja.ka] 'nephew, niece'
[wa.ta.tjan] 'last year'

As Table 49.3 displays, a significant implication of the MSD approach is that the ideal onset cluster consists of an obstruent plus a glide, all else being equal. Thus, if a language allows complex onsets and has glides in its phonemic inventory, it must permit stop + glide clusters. This is the only onset sequence occurring in all four language types in Table 49.3. An explanation for this is that these two natural classes (stops and glides) are maximally separated in terms of their relative sonority, since they occupy the extreme ends of the scale (among consonants). Baertsch (2002) proposes one way to capture MSD effects in Optimality Theory (OT: Prince and Smolensky 1993). The corresponding prediction is that some languages should exist which permit OG but no other onsets. Two such cases are Mono (Democratic Republic of the Congo; Olson 2005) and Panobo. Other possibilities are Japanese (Vance 1987, 2008) and Mandarin Chinese (Yuan 1989). The latter two are listed in Table 49.3, followed by question marks to highlight their controversial status. Also, in Hindi (Ohala 1983) and Koluwawa (Guderian and Guderian 2005), the only initial clusters are OG and NG, but not *OL. There is, however, a problem. In a sequence like [kwa], the [w] is potentially ambiguous since it allows different phonological interpretations. A priori it could pertain to a diphthongal nucleus rather than the onset: [kʊ͡a]. Alternatively, it might be a secondary articulation (labialization) of the preceding /k/: [kʷa]. If so, then there really is no consonant cluster, just a single complex phonemic unit. The third possibility is that [kw] simply constitutes a true onset cluster, as in Panobo. Teasing apart these different conclusions is complicated, and often the language-specific evidence is not compelling either way. Unfortunately, then, when no other canonical onset clusters (such as OL) exist in a language, the argumentation is in danger of circularity regardless of which segmentation is posited. See §3 for an alternative model that claims that the unmarked initial cluster is not OG but OL. Finally, in the MSD approach, the sonority distance between the second onset consonant and the vowel is not crucial, because phonotactic restrictions rarely obtain across onset–nucleus junctures (Blevins 1995). However, see §3 for an approach in which the nature of this sequence (C_2 + V) does matter. See also CHAPTER 15: GLIDES for further discussion of glides, and CHAPTER 55: ONSETS for an expanded treatment of onsets.

2.3 *The Syllable Contact Law*

Another sonority-based principle active in many languages is the Syllable Contact Law (SCL). Some seminal references are Hooper (1976), Murray and Vennemann (1983), and Clements (1990). More recent treatments of the SCL as a

Table 49.4 Alternations motivated by the Syllable Contact Law

Process	*Illustration*	*Language*	*Example*
coda weakening	g.n → w.n	Hausa	/hagni/ → [haw.ni] 'left side'
onset strengthening (desonorization)	k.l → k.t	Kyrgyz	/konok-lar/ → [konok.tar] 'guest-PL'
	l.l → l.d	Kazakh	/kol-lar/ → [kol.dar] 'hand-PL'
	z.l → z.d	Kazakh	/koŋuz-lar/ → [koŋuz.dar] 'bug-PL'
	z.m → z.b	Kazakh	/koŋuz-ma/ → [koŋuz.ba] 'bug-INT'
tautosyllabification	k.l → .kl	Germanic	[tēk-.lıç] 'daily', [e.klıç] 'disgusting'
gemination	b.r → b.br	Latin > Italian	/labrum/ → [lab.bro] 'lip'
epenthesis	n.ɾ → n.dr	Spanish	/benir-a/ → [ben.dra] '(s/he) will come'
regressive assimilation	k.m → ŋ.m	Korean	/kuk-mul/ → [kuŋ.mul] 'broth'
progressive total assimilation	g.n → g.g	Pali	/lag-na/ → [lag.ga] 'attach (PAST PART)'
regressive total assimilation	n.l → l.l	Korean	/non-li/ → [nol.li] 'logic'
anaptyxis	p.r → pV.r	Ho-Chunk	/hipres/ → [hiperes] 'know'
metathesis	d.n → n.d	Sidamo	/gud-nonni/ → [gun.donni] 'they finished'

family of OT constraints include Davis (1998), Gouskova (2004), and Zec (2007). The following are two typical formulations of the SCL:

(17) *Syllable Contact Law*

a. A heterosyllabic juncture of two consonants *A.B* is more harmonic (ideal) the higher the sonority of *A* and the lower the sonority of *B*.
b. In any heterosyllabic sequence of two consonants *A.B*, the sonority of *A* is preferably greater than the sonority of *B*.

By (17), for example, the sequence [l.k] is inherently less marked than [k.l]. Vennemann (1988: 50) provides a list of sample repair strategies that languages employ to improve satisfaction of the SCL. These are summarized in Table 49.4, as annotated by Davis (1998: 183) and supplemented by Seo (CHAPTER 53: SYLLABLE CONTACT), which offers more data and discussion. Based on a survey of 31 languages with SCL effects, her results give a better idea of the range of typological generalizations and their relative robustness. For example, Kazakh tolerates a /j-l/ juncture, as in [mandaj.lar] 'foreheads', since sonority drops slightly from /j/ to /l/. Kyrgyz, nevertheless, requires a greater fall in sonority and maps /aj-lar/ to [aj.dar] 'moons' (Davis 1998).

Examining the details of SCL phenomena in particular languages allows us to establish subtle differences in sonority ranks. For instance, Spanish attests words such as [per.la] 'pearl' and [al.re.ðe.ðor] 'around', yet the hypothetical sequence *[l.ɾ] systematically does not occur. When such a juncture would be created, an intrusive stop appears instead. This happens when the future tense is derived by dropping the infinitival theme vowel: /salir/ 'to leave' → *[sal.ɾa] → [sal.dra] '(s/he) will leave'. These facts motivate the following sonority hierarchy among Spanish liquids, based on the second of the two definitions in (17): flap > lateral > trill (Bonet and Mascaró 1997; Parker 2008).

Nevertheless, there are problems with the SCL too. For example, it predicts that obstruent + sonorant junctures should be "fixed" more often than sonorant + sonorant clusters, *ceteris paribus*. The opposite in fact is true (CHAPTER 53: SYLLABLE CONTACT). Furthermore, in Akan both /O-N/ and /N-O/ sequences result in phonetic [NN]. The latter is a mirror image of Korean nasal assimilation (Table 49.4), even though this makes syllable contact worse: /óń-dú/ → [óń.núʔ] 'he should arrive' (Schachter and Fromkin 1968).

2.4 *Rhyme weight*

It is well known that the heavier a syllable is, the more it tends to attract stress (Hayes 1980; Prince 1990). For example, open syllables are light, but closed syllables are usually bimoraic. Thus, they may be obligatorily stressed (see also CHAPTER 57: QUANTITY-SENSITIVITY). Also, in some languages, rhymes headed by /e/ or /o/ attract stress more than those with /i/ and /u/, indicating that mid vowels have more weight than high vowels in these systems. Furthermore, the propensity for a coda consonant to project a mora is correlated with how sonorous it is (Zec 1988, 1995). An adequate theory of phonology should provide a unified (non-accidental) explanation for these facts. Appealing to a scalar feature like sonority allows us to do that. Based on case studies examining the relationship between segmental quality and syllable weight effects, the following hierarchy of vowel sonority has been posited (Kenstowicz 1997; de Lacy 2002, 2004, 2006, 2007a). Specific languages may choose to exploit different subsets among these natural classes:

(18) *Relative sonority of vowels*

a > e, o > i, u > ə > ɨ

To illustrate, Kobon vowels are divided into four groups in terms of stress assignment: /a/ > /e o/ > /i u/ > /ə ɨ/. In this case the potential distinction between /ɨ/ and /ə/ is underexploited. In unaffixed Kobon words, stress predictably falls on the most sonorous nucleus within a disyllabic window at the right edge (Davies 1980, 1981):

(19)	a > e	[haŋ'gaβe]	'blood'
	a > i	[kʰi.'a]	'tree (sp.)'
	a > ə	[kʰəβə'ja]	'rat'
	a > i	['aɲ'ɨm-'aɲɨm]	'lightning'
	o > u	['mo.u]	'thus'
	o > i	[si.'oŋkʰ]	'bird (sp.)'
	o > i	[gɨ'ɾo-gɨ'ɾo]	'talk (mother pig to piglet)'
	i > ə	[ga'ɽinəŋ]	'bird (sp.)'
	i > ɨ	['jimbɨɽ]	'very'

The generalization in (18) and (19) is that vowels which are more peripheral in the acoustic space are more sonorous than central ones. Furthermore, within these two sets, segments involving a lower jaw configuration outrank their higher counterparts (Coetzee 2006). At the bottom of this hierarchy, /ə/ is higher in sonority than /ɨ/. This is due to languages such as Lushootseed, in which /ə/ can be stressed (unlike English) when it is the only nucleus in a root. Stress falls on the first "full" vowel of the stem; otherwise on the first schwa (Urbanczyk 2006):

(20) *Stressable /ə/ in Lushootseed*

['ʔitut]	'sleep'
[dzə'ɬixʷ]	'creek'
['ʧuqʷu-d]	'to whittle something'
[ʧə'gʷas]	'wife'
[k'ə'daju]	'rat'
['ʤəsəd]	'foot'
['bəʧ]	'fall down'

Other languages in which sonority is crucial to stress assignment include Pichis Ashéninka (Payne 1990; Hayes 1995), Komi (Hayes 1995; de Lacy 1997), and Finnish (Anttila 1995; Hayes 1995). However, a reviewer notes that contrastive /ə/ in languages like Lushootseed may be different in quality from phonetic [ə] resulting from reduction in English and analogous languages. For example, the former is probably longer in duration than the latter. This is a valid point that must be controled for in cross-linguistic comparisons of this sort. For Lushootseed, Urbanczyk posits a phonemic /ə/ in underlying forms. In motivating a constraint against stressed schwas she writes:

> *ə́ has the distributional hallmarks of a markedness constraint because there are languages which never stress schwa, languages which avoid stressing schwa, and languages which permit schwa, along with other vowels, to be stressed, but no language enforces the stressing of schwa in preference to other vowels. (Urbanczyk 2006: 210)

She then lists other Salishan languages in which this prohibition is active (2006: 211, fn. 24). See CHAPTER 26: SCHWA for further discussion of schwa in general.

Concerning the relative weight of coda consonants, the typological range of languages is also dependent on sonority. Table 49.5 is adapted from Zec (1995, 2007). The generalization is that if a lower-sonority class is moraic in a particular language, then all higher-sonority categories are also moraic in syllable-final position. Zec (2007) knows of no language in which coda liquids count as heavy but nasals do not; she considers this an accidental gap. In addition to stress attraction, other diagnostics for consonant moraicity are: (1) the ability to bear a contrastive tone (Tiv; Zec 1995), (2) prosodic minimality (Fijian; Dixon 1988), and

Table 49.5 Inventories of moraic segments

Natural classes contributing to syllable weight	*Languages*
vowels only	Fijian, Halh Mongolian, Lardil, Yidiny
vowels and liquids	?
vowels, liquids, and nasals	Gonja, Kwakiutl, Lithuanian, Tiv
vowels, liquids, nasals, and obstruents	Egyptian Arabic, English, Latin, Maithili

(3) blocking of processes such as vowel reduction (Maithili; Hayes 1995) (see also CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE). (21) gives sample data from the three attested language types in Table 49.5:

(21) a. *Fijian* (Dixon 1988)
Stress the syllable containing the penultimate mora (closed syllables do not occur).

[ˈsiŋa]	'day'
[ᵐbuˈtaʔo]	'steal'
[ᵐbutaˈʔoða]	'steal-TRANS'
[ʔiˈlaa]	'know-TRANS'
[ˈraiða]	'see-TRANS'
[ˈlu.a]	'vomit (VB)'

b. *Tiv* (Zec 1995)
Only sonorant consonants occur in codas, where they bear tone.

[báŕ]	'salt'
[fáˈḿ]	'rainy season'
[rùmùr̀]	'agreed, confessed'

c. *Egyptian (Cairene) Arabic* (Hayes 1995)
Stress the ultima if superheavy (trimoraic), otherwise the penult if heavy, otherwise the antepenult.

[kaˈtabt]	'I wrote'
[hadʒˈdʒaːt]	'pilgrimages'
[ˈbeːtak]	'your (MASC SG) house'
[kaˈtab.ta]	'you (MASC SG) wrote'
[muˈdar.ris]	'teacher'
[ʔinˈkasara]	'it got broken'
[ˈkataba]	'he wrote'

If fricatives are more sonorous than stops (§2.1), by implication some languages should exist in which fricatives occur in coda position but stops do not. This is exemplified by Panobo. Syllable-final consonants include glides, nasals, and fricatives. However, this is complicated by the fact that the flap /ɾ/ occurs in onsets, yet not in codas. Evidence that coda consonants are moraic in Panobo is that in word-final position they attract stress. Otherwise, the default quantity-sensitive foot type, a moraic trochee, assigns stress to the penultimate syllable (Parker 1992):

(22) *Heavy final syllables in Panobo*

[ˈatsa]	'manioc'
[kaˈnoti]	'bow (weapon)'
[jaˈwiʃ]	'opossum'
[tahˈpõŋ]	'root'
[pihˈkæ̃j]	'(they) will eat'

3 The Sonority Dispersion Principle

Clements (1990) proposes an approach to syllable phonotactics that is also based on sonority. In his model, syllables are divided into two parts. The initial demisyllable consists of any onset consonants (if present) plus the nucleus, and the final demisyllable contains the nucleus plus the coda, i.e. the rhyme. The term demisyllable is borrowed from Fujimura and Lovins (1978). The nucleus crucially resides in both demisyllables simultaneously. For example, the word /plom/ contains the demisyllables /plo/ and /om/. The essence of the Sonority Dispersion Principle (SDP) is that initial demisyllables are preferred when their constituents are maximally and evenly dispersed in sonority; e.g. /ta/. The same tendency is inverted for final demisyllables, favoring open rhymes (those ending with a vowel). More precisely, initial demisyllables of the same length (number of segments) are more harmonic to the degree that they minimize *D* in (23) below. Conversely, final demisyllables are more harmonic to the degree that they maximize *D*, all else being equal. The formula for *D* comes from the realms of physics and geometry, where it governs the distribution of mutually repelling forces in potential fields (like electrons). Its linguistic use originates with the work of Liljencrants and Lindblom (1972) on perceptual distance between segments in the acoustic vowel space. Hooper (1976) and Vennemann (1988: 13–14) anticipate its application to sonority and syllable structure.

(23) *Sonority Dispersion Principle*

$$D = \sum_{i=1}^{m} \frac{1}{d_i^2}$$

where d = distance between the sonority indices of each pair of segments, and m = number of pairs of segments (including non-adjacent ones), where $m = n(n-1)/2$, and where n = number of segments.

Clements (1990: 304) paraphrases (23) as follows: "D . . . varies according to the sum of the inverse of the squared values of the sonority distances between the members of each pair of segments within" a demisyllable. *D*, then, is the reciprocal of dispersion. To illustrate the application of (23), Clements assumes a sonority scale with the five categories from (3):

(24)

		sonority index
vowels	(V)	5
glides	(G)	4
liquids	(L)	3
nasals	(N)	2
obstruents	(O)	1

When *D* is computed for demisyllables containing exactly one or two consonants, it yields the following values (ignoring types that violate the SSP):

(25) *Sonority Dispersion demisyllable values*

a.			b.
OV = .06	most natural onset		OLV = .56
NV = .11	↑		OGV = 1.17
LV = .25			ONV = 1.17
GV = 1.00			NGV = 1.36
	↓		NLV = 1.36
	least natural onset		LGV = 2.25

c.	VO = .06	least natural coda	d.	VLO = .56
	VN = .11	↑		VGO = 1.17
	VL = .25			VNO= 1.17
	VG = 1.00			VGN = 1.36
		↓		VLN = 1.36
		most natural coda		VGL = 2.25

In (25a) the SDP favors a single syllable-initial consonant that maximizes the sonority slope between the onset and the vowel. Therefore, it correctly predicts that the preferred syllable of type CV has an onset consonant as low in sonority as possible ($D = .06$). This results in the following scale of relative unmarkedness: /ta/ > /na/ > /la/ > /ja/. This is foreshadowed in the discussion of Sanskrit reduplication (§2.1). Recall that the pattern there demonstrates that fricatives can outrank stops in sonority. Therefore, not all obstruents are necessarily equal in sonority (see §4).

Summarizing thus far, a unique consequence of the formula for *D* is that in an initial demisyllable of two segments (CV), the segments should differ from each other in sonority as much as possible. This is somewhat analogous to the MSD approach for onset clusters (§2.2). However, when the initial demisyllable contains three segments (CCV), what matters (given *D*) is that the aggregate total of the sonority distances between all of these together be maximized. This is accomplished by spacing apart the segments as evenly as possible (in sonority). This results in the best evaluation for OLV in (25b), since it has the lowest obtained value (.56). This is because liquids fall precisely midway between obstruents and vowels in terms of their sonority indices in Clements's five-category scale in (24). As evidence for the SDP, Clements notes that underlyingly French permits complex demisyllables of the type OLV only. However, in surface forms, some instances of OGV also exist. This raises an important typological question: which onset cluster is universally unmarked, OL or OG? On the one hand (as noted in §2.2), the MSD principle predicts that some languages should permit OG but not *OL. This seems to be correct, but see the caveats in §2.2. On the other hand, the SDP claims that OL is preferred. One piece of evidence that could help resolve this would be an alternation mapping underlying OLV to OGV, or vice versa. Unfortunately, no such process has yet been observed. A cross-linguistic survey documenting the number of languages with one type of cluster but not the other would also be enlightening. It may be that both kinds of onsets (OG and OL) need to be optimal simultaneously, i.e. in the grammars of different languages. Rod Casali and Ken Olson (personal communication) note that apparent OG-only languages are especially common in Africa.

Finally, the SDP assigns an equal evaluation to the two demisyllable types OGV and ONV in (25b). This may be problematic, since many languages exhibit the initial sequence OG, but not *ON. For example, Table 49.3 mentions Gizrra, Kurdish, Mono, Panobo, and Spanish. However, Clements does not claim that demisyllables of the same rank necessarily co-occur in any language containing one of them. At the same time, no languages appear to allow ONV but not *OGV. If this is a systematic gap, it is troubling for the predictions of the SDP.

4 The complete sonority hierarchy

Perhaps no issue in phonological theory has led to more competing proposals than the internal structure of the sonority hierarchy, i.e. the numbers and types of natural classes, and their corresponding ranks. Parker (2002) notes that more than 100 distinct sonority scales are found in the literature. The purpose of this section is to lay out several desirable characteristics that a full and final sonority hierarchy should possess, and then present one specific model that arguably comes closest to fulfilling those goals. Briefly, an adequate sonority scale should display the a priori traits in (26). In principle these criteria apply not just to sonority, but to all phonological features, that is, classic binary features like [±voice], [±round], etc.

(26) All else being equal, an ideal sonority scale would have these characteristics:

a. *Universal:* It potentially applies to all languages.
b. *Exhaustive:* It encompasses all categories of speech sounds.
c. *Impermutable:* Its rankings cannot be reversed (although they may be collapsed or ignored).
d. *Phonetically grounded:* It corresponds to some consistent, measurable physical parameter shared by all languages.

Each of the points in (26) will now be discussed. First, ideally we can establish a single, unique sonority hierarchy to analyze all known languages. This is not to say that any particular language actually exploits every one of the natural class rankings in the sonority scale. On the contrary, it would be quite amazing (although fortuitous) to discover such a case. Nevertheless, the explanatory power of sonority is maximized if we ascribe it to UG, making it equally available to all humans.

Second, an adequate theory of sonority should include every known type of phonological segment. Many hierarchies omit recalcitrant natural classes such as glottal consonants (/h/ and /ʔ/), affricates, etc., perhaps because of their inherent complications. Such scales then cannot apply to all languages. This undermines their universality.

Third, the rankings in the sonority scale should be impermutable. This is a beneficial characteristic since it is the most restrictive hypothesis possible, i.e. it severely limits the types of processes directly attributable to sonority. In addition to avoiding overgeneration of non-attested language types, impermutability makes claims about sonority easier to falsify. This in turn reduces the danger of circular argumentation. For example, once it is established that laterals, for instance, are more sonorous than nasals, the entailment is that there is no language in which nasals pattern as more sonorous than laterals by the same criteria. At the same time, however, potential divisions between sonority ranks are frequently underexploited in many languages. See de Lacy (2002, 2004, 2006, 2007a) for a formal approach to "underspecification" of sonority classes in OT.

Fourth, in an ideal world we can show that sonority is based on concrete articulatory gestures and/or their acoustic counterparts. This is the topic of the next section.

Although no sonority scale to date perfectly fulfills every desideratum sketched above, one that perhaps comes closest is that of Parker (2008), reproduced below:

(27) *Final hierarchy of relative sonority*

Natural class	*Sonority index*
low vowels	17
mid peripheral vowels (not [ə])	16
high peripheral vowels (not [ɨ])	15
mid interior vowels ([ə])	14
high interior vowels ([ɨ])	13
glides	12
rhotic approximants ([ɹ])	11
flaps	10
laterals	9
trills	8
nasals	7
voiced fricatives	6
voiced affricates	5
voiced stops	4
voiceless fricatives (including [h])	3
voiceless affricates	2
voiceless stops (including [ʔ])	1

Space does not permit a detailed justification of (27). Nevertheless, to highlight a few positive aspects of this scale, the evidence for most of the natural classes is fairly robust and secure. Parker (2002, 2008) summarizes the debates and provides at least one argument to motivate every ranking in this hierarchy (every pair of adjacent categories). Much of this is reviewed in §2 of this chapter. To give another example, Koine (Ancient) Greek permits the clusters /pn/ and /kn/ (/pniktos/ '(things) strangled', /kneθo/ 'have itching, tickled'; Mounce 1993), but proscribes */bn/ and */dn/. This can be explained as an MSD effect if voiced stops are closer to nasals (in sonority) than voiceless stops are. Furthermore, flaps are higher in sonority than trills in Spanish, as established by the SCL in §2.3. Three other facts confirm this. First, in word-initial position the contrast between /r/ and /ɾ/ is neutralized to /r/ ([rana] 'frog'). Second, in codas it is neutralized in favor of /ɾ/: [ʧaɾlaɾ] 'to chat'. These two points follow from the SDP (Bonet and Mascaró 1997). Third, /ɾ/ and /l/ appear as the second member of complex onsets, yet /r/ does not (see (5)). This is another MSD effect if /r/ is less sonorous than /ɾ/ and /l/ (Baković 1994). These distributional facts indicate that liquids do not always pattern as a monolithic class in Spanish. Finally, the rhotic approximant /ɹ/ is higher in sonority than /l/ in English: (1) the contrast between *Carl* (one syllable) *vs. caller* (two syllables) follows from the SSP if /ɹ/ outranks /l/ (Borowsky 1986); (2) /ɹ/ is the default epenthetic coda in Eastern Massachusetts speech (McCarthy 1993). This follows from the SDP if /ɹ/ is the most sonorous consonant available in this position. And (3) syllabic /ɹ/ may bear stress (*bird, curtain*), but /l/ never does (Zec 2003).

Another strength of the scale in (27) is its breadth, i.e. the number of different types of segment classes it encompasses. Nevertheless, it is still not exhaustive, because it leaves out a few rarer kinds of sounds, such as clicks and implosives. Since no known sonority hierarchy includes these, more research in this area would be welcome. A third advantage of (27) is that it has been rigorously tested and found to provide a good fit with all phonetic segments in seven specific languages. This is discussed in §5.

However, a few problems persist. For example, the placement of affricates between stops and fricatives is a controversial issue, remaining open to disagreement. Many scales either leave affricates out entirely or group them with plosives, using a term such as *stops*. Such proposals may simply assume that affricates behave phonologically like stops, as they do in many languages. For instance, Kang (CHAPTER 95: LOANWORD PHONOLOGY) concludes that affricates are just stops with a special place specification, e.g. [strident]. See CHAPTER 16: AFFRICATES for more discussion of affricates in general. Similarly, the ranking of voiced stops over voiceless fricatives is harder to justify than most aspects of this hierarchy. A major reason for this is that many languages require consonant clusters to agree in voicing. Therefore, crucial diagnostic examples are rare, but one such token is the English word *midst*. Since this form is monosyllabic, /d/ is higher in sonority than /s/ by the SSP since /d/ is closer to the nucleus. Finally, the question of whether glottal consonants are sonorants or obstruents is also contested. Clearly /h/ and /ʔ/ pattern phonologically with prototypical sonorants in some languages, yet behave like obstruents in others. In (27) they are classified as obstruents. One piece of evidence supporting this is that in Panobo, /h/ groups with /β p t/ in exclusively obstruent + glide clusters (see (16)). Also, in many languages [ʔ] is inserted as a default onset, where segments of low sonority are preferred by the SDP (Lombardi 2002). Finally, in the P-base sample of 549 languages, there are 65 distinct phonological processes in which /h/ and/or /ʔ/ pattern solely with consonants that are unambiguously obstruents. In 21 other cases they group with sonorants (Mielke 2008).[6]

In the scale in (27) the tendency is obviously to "split" rather than to "lump together" natural classes. The motivation for this is as follows. There is ample evidence that fine-grade distinctions in sonority need to be made in some languages, such as between fricatives and stops in Sanskrit (§2.1). UG then must allow for these options, and hence the potential exists for other languages to exploit them as well. If we start with a hierarchy that assigns a unique rank to every distinct manner of articulation (like (27)), it is a trivial matter to formally compress (conflate) ranks together in order to analyze languages not invoking those splits. This procedure applies to every language in one way or another. However, if we assume a maximal sonority hierarchy with just the five groups in (3), no mechanism exists to "decompose" these, making more narrow distinctions in the scale when necessary. Consequently, only a fully detailed hierarchy such as (27) is flexible enough to generate the range of variation attested among the languages of the world with respect to processes involving sonority.

Nevertheless, based on acoustic studies of many languages, Zhang (2001) and Gordon (2006) deny that a universal sonority scale is theoretically the most parsimonious option. Rather, they reject the existence of invariant sound classes. Gordon, for instance, concludes that syllable weight effects are not a unified

[6] Thanks to Jeff Mielke (personal communication) for these counts.

phenomenon. He claims that nasals, for example, display slightly different phonetic behavior from one language to another. This can then influence their phonological patterning in terms of sonority.

5 The physical substance of sonority

As summarized in §2, the function of sonority in phonological systems is fairly well understood. Nevertheless, these phenomena raise an important, related question that has provoked much contention and speculation: is there any coherent notion of sonority grounded in evidence external to the phonotactic facts that sonority is assumed to account for? In other words, what is the articulatory, acoustic, and/or perceptual source of sonority in the speech signal? To date at least 98 different correlates of sonority have been posited, documented in Parker (2002). The most frequently proposed phonetic definition of sonority is probably openness (of the vocal tract) or (supralaryngeal) aperture (Bloomfield 1914; Jespersen 1922; Goldsmith 1990; Kirchner 1998), or its inverse, (supraglottal) stricture, closure, impedance, etc. (Halle and Clements 1983; Kenstowicz 1994; Hume and Odden 1996). However, notions such as impedance are difficult to quantify. A more promising correlate of sonority is amplitude/intensity, or its perceptual counterpart, loudness (Bloomfield 1914; Laver 1994; cf. §1 and see also CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). Recently a major instrumental study was carried out measuring relative sound levels or RMS (root mean square) intensity of all phonemes of Peruvian Spanish, Cusco Quechua, and Midwestern American English (Parker 2008). See Jany *et al.* (2007) for a similar investigation of four other languages (Egyptian Arabic, Hindi, Mongolian, and Malayalam). In Parker (2008), the obtained intensity values for all segments yield an overall mean Spearman's correlation of .91 with the sonority indices proposed in (27). Given those results, it is proposed there that the best way to characterize the physical basis of sonority is via a linear regression equation such as (28) below. This is calculated from the mean intensity measurements of all English coda consonants pertaining to nine of the natural classes from (27). These were pronounced five times each by five male native speakers of English.

(28) $sonority = 13.9 + .48 \times dB$ (dB = decibel; $r^2 = .95$)

This formula predicts an estimated sonority index based on a hypothetical intensity value. It characterizes the best-fitting line corresponding to the relative sonority of English coda consonants in phrasally stressed words. The obtained intensity value of each of these segments was compared to that of a stressed, utterance-initial low vowel (/ɑ/) in a fixed carrier sentence. In (28) the Y intercept is 13.9. This is the projected value of Y (sonority) when X (dB) equals 0. Here it is significantly higher than the theoretical null value. This is because the obtained intensity of the reference vowel /ɑ/, whose sonority index is 17, was subtracted from that of the target consonant for each utterance measured. This is a type of normalization procedure often performed to control for random fluctuations in loudness across speakers and tokens. The slope in (28), .48, indicates the rate of change in the dependent variable Y (sonority) per unit change in the independent variable X (dB). Its obtained value allows us to approximate the mathematical nature of the relationship between intensity and sonority. Specifically, for every decibel by which the relative sound level is increased, the corresponding sonority rank increases by about .48 units (for this sample of five English speakers). Also in (28), $r^2 = .95$. This is the coefficient of determination. It indicates the proportional reduction of error gained by using the linear regression model. Given this r^2 value, we can conclude that the single factor *sonority* accounts for (predicts) about 95 percent of the systematic variability in the intensity measurements of that dataset.

Compared with previous accounts of sonority, the definition in (28) has several advantages (Parker 2008): (1) it is precise; (2) it is non-arbitrary; (3) it is phonetically grounded; (4) it is empirically verifiable and replicable; (5) it can be calculated for other speakers and languages; and (6) the underlying methodology (regression) is compatible with different (competing) sonority scales. However, while studies of this type represent progress, some problems remain. For example, in Parker (2008) the majority of the mismatches between sonority ranks and segmental intensity values (in all three languages) involve the sonorant consonants, particularly the approximants (laterals and glides). The reason for this is not clear at this point and merits further investigation.

Finally, other researchers appeal to more functional aspects of the speech signal to avoid invoking sonority altogether. For example, building on the phonetically based work of Mattingly (1981) and Silverman (1995), Wright (2004: 35) reformulates the SSP as "a perceptually motivated and scalar constraint in which an optimal ordering of segments is one that maximises robustness of encoding of perceptual cues to the segmental make-up of the utterance." Similarly, Ohala (1990a) claims that what drives the phonological phenomena discussed here is not really sonority but simply a need for adequate modulation in the acoustic wave.

6 Conclusion

Despite its problems, sonority makes sense. If it did not exist, it would be invented (Parker 2008). In this chapter a number of important issues have been examined. Nevertheless, certain topics need to be left for future work. For example, in §3 a possible contradiction between the claims of the MSD approach and the SSP is noted, involving the relative unmarkedness of OGV *vs.* OLV demisyllables. An in-depth typological study of these clusters would be helpful. Also, more attention should be given to the phonetic and/or functional bases of principles such as the SSP, the SCL, and the MSD. The question of why these hold true is potentially intriguing. Finally, another interesting point not discussed here is whether sonority scales are necessarily the same across different domains, such as phonotactics *vs.* the calculation of syllable weight.

ACKNOWLEDGMENTS

Thanks to the following people for helpful comments: Marc van Oostendorp, Keren Rice, Michael Boutin, Mike Cahill, Ken Olson, two anonymous reviewers, and the students in my Frontiers in Phonology course taught at the Graduate Institute of Applied Linguistics in the fall of 2009. This work was partially funded by NSF grant #0003947.

REFERENCES

Anttila, Arto. 1995. Deriving variation from grammar: A study of Finnish genitives. Unpublished ms., Stanford University (ROA-63).

Baertsch, Karen. 2002. An optimality theoretic approach to syllable structure: The split margin hierarchy. Ph.D. dissertation, Indiana University.

Baković, Eric. 1994. Strong onsets and Spanish fortition. *MIT Working Papers in Linguistics* 23. 21–39.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Ph.D. dissertation, University of Massachusetts, Amherst (ROA-259).

Blevins, Juliette. 1995. The syllable in phonological theory. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 206–244. Cambridge, MA & Oxford: Blackwell.

Blevins, Juliette. 2006. Syllable: Typology. In Keith Brown (ed.) *Encyclopedia of language and linguistics*. 2nd edn. vol. 12, 333–337. Amsterdam: Elsevier.

Bloomfield, Leonard. 1914. *An introduction to the study of language*. New York: Henry Holt & Co.

Bodegraven, Nico van & Elly van Bodegraven. 2005. Phonology essentials: Gizrra language. In Parker (2005), 191–210.

Bonet, Eulalia & Joan Mascaró. 1997. On the representation of contrasting rhotics. In Fernando Martínez-Gil & Alfonso Morales-Front (eds.) *Issues in the phonology and morphology of the major Iberian languages*, 103–126. Washington, DC: Georgetown University Press.

Borowsky, Toni. 1986. Topics in the lexical phonology of English. Ph.D. dissertation, University of Massachusetts, Amherst.

Brosses, Charles de. 1765. *Traité de la formation méchanique des langues, et des principes physiques de l'étymologie*, vol. 1. 130–133. Paris: Saillant, Vincent, Desaint.

Cho, Young-mee Yu & Tracy Holloway King. 2003. Semisyllables and universal syllabification. In Féry & van de Vijver (2003), 183–212.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 283–333. Cambridge: Cambridge University Press.

Coetzee, Andries W. 2006. Variation as accessing "non-optimal" candidates. *Phonology* 23. 337–385.

Cser, András. 2003. *The typology and modelling of obstruent lenition and fortition processes*. Budapest: Akadémiai Kiadó.

Davies, John. 1980. *Kobon phonology*. (*Pacific Linguistics* B68.) Canberra: Australian National University.

Davies, John. 1981. *Kobon*. Amsterdam: North-Holland.

Davis, Stuart. 1998. Syllable contact in Optimality Theory. *Korean Journal of Linguistics* 23. 181–211.

de Lacy, Paul. 1997. Prosodic categorisation. M.A. thesis, University of Auckland (ROA-236).

de Lacy, Paul. 2002. The formal expression of markedness. Ph.D. dissertation, University of Massachusetts, Amherst.

de Lacy, Paul. 2004. Markedness conflation in Optimality Theory. *Phonology* 21. 145–199.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.

de Lacy, Paul. 2007a. The interaction of tone, sonority, and prosodic structure. In de Lacy (2007b), 281–307.

de Lacy, Paul (ed.) 2007b. *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press.

Dixon, R. M. W. 1988. *A grammar of Boumaa Fijian*. Chicago: University of Chicago Press.

Engelenhoven, Aone van. 2004. *Leti, a language of Southwest Maluku*. Leiden: KITLV Press.

Féry, Caroline & Ruben van de Vijver (eds.) 2003. *The syllable in Optimality Theory*. Cambridge: Cambridge University Press.

Fujimura, Osamu & Julie B. Lovins. 1978. Syllables as concatenative phonetic units. In Alan Bell & Joan B. Hooper (eds.) *Syllables and segments*, 107–120. Amsterdam: North-Holland.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Oxford & Cambridge, MA: Blackwell.

Gordon, Matthew. 2006. *Syllable weight: Phonetics, phonology, typology*. London: Routledge.

Gouskova, Maria. 2004. Relational hierarchies in Optimality Theory: The case of syllable contact. *Phonology* 21. 201–250.

Greenberg, Joseph H. 1978. Some generalizations concerning initial and final consonant clusters. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 2: *Phonology*, 243–279. Stanford: Stanford University Press.

Guderian, Brad & Toni Guderian. 2005. Organised phonology data supplement, Koluwawa language. In Parker (2005), 75–86.

Halle, Morris & G. N. Clements. 1983. *Problem book in phonology*. Cambridge, MA: MIT Press.

Hankamer, Jorge & Judith Aissen. 1974. The sonority hierarchy. *Papers from the Annual Regional Meeting, Chicago Linguistic Society: Parasession on natural phonology*, 131–145.

Hayes, Bruce. 1980. A metrical theory of stress rules. Ph.D. dissertation, MIT. Published 1985, New York: Garland.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hironymous, Patricia. 1999. Selection of the optimal syllable in an alignment-based theory of sonority. Ph.D. dissertation, University of Maryland at College Park.

Hock, Hans Henrich. 1985. Regular metathesis. *Linguistics* 23. 529–546.

Hooper, Joan B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.

Hume, Elizabeth & David Odden. 1996. Reconsidering [consonantal]. *Phonology* 13. 345–376.

Itô, Junko. 1986. Syllable theory in prosodic phonology. Ph.D. dissertation, University of Massachusetts, Amherst. Published 1988, New York: Garland.

Jany, Carmen, Matthew Gordon, Carlos M. Nash & Nobutaka Takara. 2007. How universal is the sonority hierarchy? A cross-linguistic acoustic study. Paper presented at the 16th International Congress of Phonetic Sciences, Saarbrücken.

Jespersen, Otto. 1904. *Lehrbuch der Phonetik*. Leipzig & Berlin: Teubner.

Jespersen, Otto. 1922. *A Modern English grammar on historical principles*. Part 1: *Sounds and spelling*. 3rd edn. Heidelberg: Carl Winter's Universitätsbuchhandlung.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kawasaki, Haruko. 1982. An acoustical basis for universal constraints on sound sequences. Ph.D. dissertation, University of California, Berkeley.

Kämpfe, Hans-Rainer & Alexander P. Volodin. 1995. *Abriß der tschuktschischen Grammatik auf der Basis der Schriftsprache*. Wiesbaden: Harrassowitz.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kenstowicz, Michael. 1997. Quality-sensitive stress. *Rivista di Linguistica* 9. 157–187.

Kenstowicz, Michael & Charles W. Kisseberth. 1973. Unmarked bleeding orders. *Studies in the Linguistic Sciences* 1.(1), 8–28. Reprinted in Charles W. Kisseberth (ed.) 1973. *Studies in generative phonology*, 1–12. Champaign, IL: Linguistic Research.

Kirchner, Robert. 1998. An effort-based approach to consonant lenition. Ph.D. dissertation, UCLA (ROA-276).

Ladefoged, Peter. 1975. *A course in phonetics*. New York: Harcourt Brace Jovanovich.
Laver, John. 1994. *Principles of phonetics*. Cambridge: Cambridge University Press.
Lewis, M. Paul (ed.) 2009. *Ethnologue: Languages of the world*. 16th edn. Dallas: SIL International.
Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48. 839–862.
Lombardi, Linda. 2002. Coronal epenthesis and markedness. *Phonology* 19. 219–251.
Mattingly, Ignatius G. 1981. Phonetic representation and speech synthesis by rule. In Terry Myers, John Laver & John Anderson (eds.) *The cognitive representation of speech*, 415–420. Amsterdam: North Holland.
McCarthy, John J. 1993. A case of surface constraint violation. *Canadian Journal of Linguistics* 38. 169–195.
Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.
Morelli, Frida. 2003. The relative harmony of /s+stop/ onsets: Obstruent clusters and the Sonority Sequencing Principle. In Féry & van de Vijver (2003), 356–371.
Mounce, William D. 1993. *The analytical lexicon to the Greek New Testament*. Grand Rapids, MI: Zondervan Publishing House.
Murray, Robert W. & Theo Vennemann. 1983. Sound change and syllable structure in Germanic phonology. *Language* 59. 514–528.
Ohala, John J. 1974. Phonetic explanation in phonology. *Papers from the Annual Regional Meeting, Chicago Linguistic Society: Parasession on natural phonology*, 251–274.
Ohala, John J. 1990a. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26(2). 319–338.
Ohala, John J. 1990b. There is no interface between phonology and phonetics: A personal view. *Journal of Phonetics* 18. 153–171.
Ohala, Manjari. 1983. *Aspects of Hindi phonology*. New Delhi: Motilal Banarsidass.
Olson, Kenneth S. 2005. *The phonology of Mono*. Dallas: SIL International & University of Texas at Arlington.
Parker, Steve. 1989. The sonority grid in Chamicuro phonology. *Linguistic Analysis* 19. 3–58.
Parker, Steve. 1992. *Datos del idioma huariapano*. Yarinacocha, Pucallpa, Peru: Ministerio de Educación & Instituto Lingüístico de Verano.
Parker, Steve. 2002. Quantifying the sonority hierarchy. Ph.D. dissertation, University of Massachusetts, Amherst.
Parker, Steve (ed.) 2005. *Data papers on Papua New Guinea languages, vol. 47: Phonological descriptions of PNG languages*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
Parker, Steve. 2008. Sound level protrusions as physical correlates of sonority. *Journal of Phonetics* 36. 55–90.
Payne, Judith. 1990. Asheninca stress patterns. In Doris L. Payne (ed.) *Amazonian linguistics: Studies in lowland South American languages*, 185–209. Austin: University of Texas Press.
Pike, Kenneth L. 1943. *Phonetics: A critical analysis of phonetic theory and a technic for the practical description of sounds*. Ann Arbor: University of Michigan Press.
Prince, Alan. 1990. Quantitative consequences of rhythmic organization. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26(2). 355–398.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.
Regnier, Sue. 1993. Quiegolani Zapotec phonology. *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 37. 37–63.
Ridouane, Rachid. 2008. Syllables without vowels: Phonetic and phonological evidence from Tashlhiyt Berber. *Phonology* 25. 321–359.

Schachter, Paul & Victoria Fromkin. 1968. A phonology of Akan: Akuapem, Asante, Fante. *UCLA Working Papers in Phonetics* 9.
Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*, 107–136. Cambridge, MA: MIT Press.
Sievers, Eduard. 1893. *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen.* Leipzig: Breitkopf & Härtel. Originally published 1881.
Silverman, Daniel. 1995. Phasing and recoverability. Ph.D. dissertation, University of California, Los Angeles. Published 1997, New York: Garland.
Smolensky, Paul. 1995. On the structure of the constraint component *Con* of UG (ROA-86).
Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT. Published 1990, New York: Garland.
Steriade, Donca. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5. 73–155.
Urbanczyk, Suzanne. 2006. Reduplicative form and the root-affix asymmetry. *Natural Language and Linguistic Theory* 24. 179–240.
Vance, Timothy J. 1987. *An introduction to Japanese phonology.* Albany: State University of New York Press.
Vance, Timothy J. 2008. *The sounds of Japanese.* Cambridge: Cambridge University Press.
Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin.* Berlin: Mouton de Gruyter.
Whitney, William Dwight. 1889. *Sanskrit grammar, including both the classical language, and the older dialects, of Veda and Brahmana.* 2nd edn. Cambridge, MA: Harvard University Press.
Wichmann, Søren. 2002. *Diccionario analítico del popoluca de Texistepec.* México, D.F.: Universidad Nacional Autónoma de México.
Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 34–57. Cambridge: Cambridge University Press.
Yuan, Jiahua. 1989. *Hanyu fangyan gaiyao* [*An introduction to Chinese dialects*]. 2nd edn. Beijing: Wenzi Gaige Chubanshe.
Zec, Draga. 1988. Sonority constraints on prosodic structure. Ph.D. dissertation, Stanford University. Published 1994, New York: Garland.
Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12. 85–129.
Zec, Draga. 2003. Prosodic weight. In Féry & van de Vijver (2003), 123–143.
Zec, Draga. 2007. The syllable. In de Lacy (2007b), 161–194.
Zhang, Jie. 2001. The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis. Ph.D. dissertation, University of California, Los Angeles. Published 2002, New York: Routledge.

50 Tonal Alignment

PILAR PRIETO

1 Introduction

In recent decades, the issue of *tonal alignment* has been at the forefront of several phonological and phonetic debates in the analysis of intonation. Since the groundbreaking work of Bruce (1977), the autosegmental metrical approach to intonation proposed that intonational patterns were to be represented as autosegmental tone melodies (Pierrehumbert 1980; Beckman and Pierrehumbert 1986; Ladd 1996; and others). Given that melodies are independent from the segments which realize them in this theory (CHAPTER 45: THE REPRESENTATION OF TONE; CHAPTER 14: AUTOSEGMENTS), and since the tones are realized potentially over quite long strings, it is a central research issue to find a set of principles for mapping tones to segments. The term tonal alignment thus refers to the temporal implementation of fundamental frequency (F0) movements with respect to the segmental string. Tonal alignment has not only been used in crucial ways as an argument in favor of a given phonological framework, but has also been the focus of debate in itself. This notion has played an important role in current theories of intonational phonology, since relative alignment of tones with the segmentals has been shown to be a crucial piece of information when describing the phonological make-up of the melodic contour. This chapter reviews four important topics in the recent history of phonology in the discussion of which tonal alignment has been a crucial component.

One of the important issues in intonational phonology is the investigation of the acoustic correlates that encode *intonational categories*. Since the beginning of the autosegmental metrical approach to intonation, tonal alignment has been claimed to play a central role in encoding intonational contrasts. Pierrehumbert (1980) and Pierrehumbert and Steele (1989) showed that the timing of F0 peaks or valleys with segments functions contrastively in English, and that early-aligned pitch accents are phonologically distinct from late-aligned pitch accents. In the decades since the publication of these studies, a body of experimental research has shown that tonal alignment cues semantic distinctions in a number of languages and that it can be perceived in a near-categorical fashion (e.g. Kohler 1987; Niebuhr 2007 for German; D'Imperio and House 1997 and D'Imperio 2000 for Neapolitan Italian; Gili-Fivela 2009 for Pisa Italian; Pierrehumbert and Steele

1989 and Dilley 2007 for English). In §2 we will review recent experimental evidence that elucidates the role of tonal alignment in encoding intonational distinctions in a number of languages.

The relationship between *tonal association* and *tonal alignment* has been a central issue in the tonal representation debates within the autosegmental metrical theory of intonation. Though the autosegmental metrical representational proposal has met with considerable success in accounting for melodic patterns in a variety of languages, the literature on tonal representation has identified a few phenomena that resist transparent analysis. Two such phenomena have to do with the metrical part of the model and the standard interpretation of the relationship between phonological association and phonetic alignment. It has recently been claimed that the theoretical concept of starredness is somewhat unclear and that its definition cannot be based solely on phonetic alignment (Arvaniti *et al.* 2000; Prieto *et al.* 2005). In §3 we describe the standard view of the relationship between phonological association of tones and phonetic alignment and then review some recent proposals on the topic.

Another important goal of several models of intonation has been to develop a *phonetic model of tonal alignment*. Within these models, it is a central issue to determine what part of the variation in the realization of the tune-to-text mapping is due to phonetic implementation and what part is phonological and is accounted for in a phonological representation (either of the tone melodies or of prosodic or segmental anchors for tones). A body of work on tune–text alignment has shown that, apart from phonological distinctions in alignment, a variety of phonetic factors, such as tonal crowding, speech rate and syllable structure influence the fine-grained patterns of F0 location in predictable ways. For example, it has been demonstrated that time pressure from the right-hand prosodic context (i.e. the proximity of an upcoming accent or boundary tones) is crucial in determining the location of H peaks (see e.g. Silverman and Pierrehumbert 1990 for English and Prieto *et al.* 1995 for Spanish). Recent work has shown that when such right-hand prosodic effects are excluded (i.e. when the tonal features under investigation are not in the vicinity of pitch accents or boundary tones), the alignment of F0 peak targets is consistently governed by *segmental anchoring* (Arvaniti *et al.* 1998 for Greek; Ladd *et al.* 1999 for English). Similarly, other work on production and perception supports the hypothesis that prosodic structure must play an essential part in our understanding of the coordination of pitch gestures with the segmentals and that listeners are able to employ these fine details of H tonal alignment due to syllable structure or within-word position to identify lexical items (D'Imperio *et al.* 2007b; Prieto *et al.*, forthcoming). In §4 we review recent proposals regarding phonetic models of tonal alignment and the role of prosodic structure in the implementation of F0 tonal alignment patterns.

Finally, tonal alignment studies have also been used to test specific predictions by different *phonological models* of prosody and intonation. Arvaniti and Ladd (2009) provide a useful example of how a production study on alignment can be used to test specific predictions by *target-based vs. configuration-based* models of intonation (CHAPTER 32: THE REPRESENTATION OF INTONATION). As we will see below, Arvaniti and Ladd undertook a very detailed phonetic study of the intonation of Greek wh-questions and tested different predictions about tonal implementation. The F0 alignment data showed predictable adjustments in alignment depending on the location of adjacent tonal targets. The authors conclude that models that specify the F0 of all syllables, and models that specify F0 by superposing contour shapes for shorter and longer domains, cannot account for predictable variation without resorting to ad hoc tonal specifications, which, in turn, do not allow for phonological generalizations about contours applying to utterances of different lengths. In §5 we review the evidence coming from a variety of tonal alignment studies that test specific predictions from different phonological models of intonation.

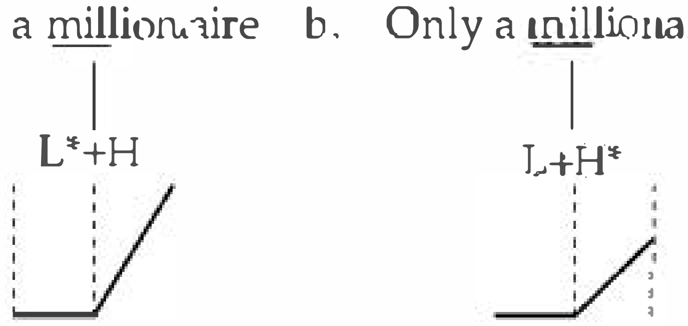
In the following sections, we present and discuss each of these four topics, providing the relevant data and highlighting some of the unresolved issues.

2 The role of tonal alignment in distinguishing intonational categories

One of the key discoveries within work on intonation is the fact that tones in intonational languages are associated with either metrically prominent syllables (*pitch accents*) or prosodic edges (*boundary tones*). Many theories of intonational phonology thus draw a clear distinction between the two sorts of tonal units, i.e. tonal entities associated with prominent or metrically strong syllables and tonal entities associated with edges of prosodic domains. Within the autosegmental metrical (AM) approach to intonation initially developed by Pierrehumbert (1980), she argues that the English intonation system consists of an inventory of tonal units, each consisting of either one or two tones, which can be High (H) or Low (L) (see CHAPTER 14: AUTOSEGMENTS; CHAPTER 116: SENTENTIAL PROMINENCE IN ENGLISH). These tones can either be associated with metrically strong syllables (and represented with a *, i.e. H* and L*) or be associated with prosodic edges (and represented with a %, i.e. H% and L%).

Tonal units can be monotonal or bitonal. In the case of tonal units associated with prominent syllables, or pitch accents, Pierrehumbert proposed a phonological inventory of six pitch-accent shapes for English (H*, L*, H*+L, H+L*, L*+H, L+H*), some of them encoding alignment differences. Crucially, the AM model started to make use of the star notation (*) in bitonal pitch accents to indicate tonal association with metrically strong syllables and relative alignment – see §3 for a review of the starredness concept. The autosegmental representations in (1) capture the fact that the LH shape is aligned differently in the two contrastive pitch accents exemplified in figure 50.1. While L*+H has a low tone (L) on the stressed syllable and a high tone (H) trailing it, L+H* has a high tone on the stressed syllable with a low tone leading it:

(1) a. Only a <u>mill</u>ionaire b. Only a <u>mill</u>ionaire



In sum, an important proposal of the AM model of intonation, based on Bruce's (1977) analysis of the tonal alignment contrast between Accent I and Accent II in

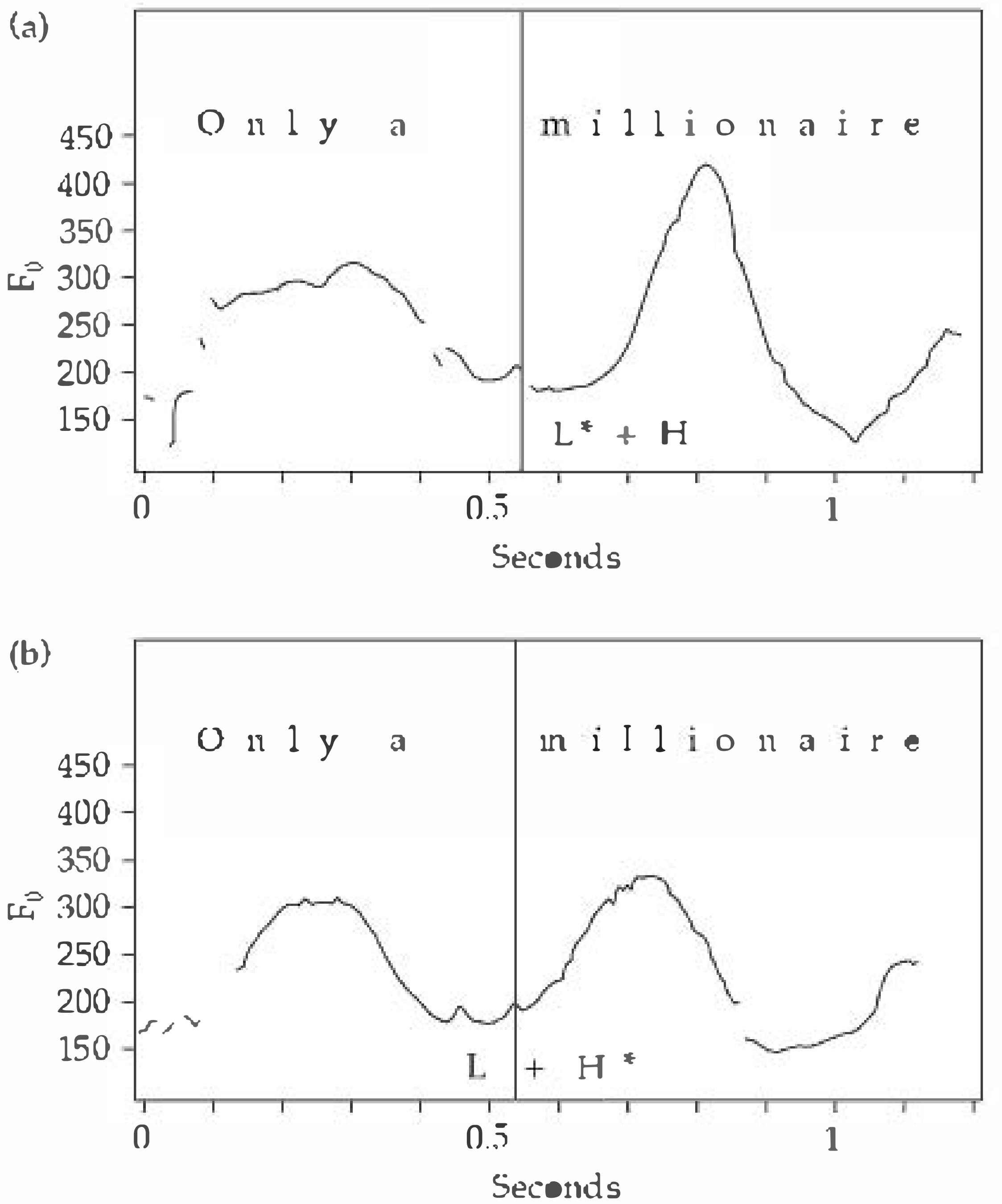


Figure 50.1 The fundamental frequency contour of the utterance *Only a millionaire* spoken with two different pitch accents on *millionaire*: the late-aligned pitch accent, which indicates incredulity or uncertainty (a), and the early-aligned pitch pattern, which indicates assertion (b). The vertical cursor is placed at the [m] release in *millionaire*. Figure reproduced from Pierrehumbert and Steele (1989: 182)

Swedish, is that pitch accent types can be phonologically distinguished by their *relative alignment* with the metrically prominent syllable. Pierrehumbert (1980) shows that tonal alignment functions contrastively in English and that early-aligned pitch accents are phonologically distinct from late-aligned pitch accents. Figure 50.1 shows two intonation patterns of the utterance *Only a millionaire* spoken with two different pitch accents on *millionaire*: the late-aligned pitch accent, which indicates incredulity or uncertainty (a), and the early-aligned pitch pattern, which indicates assertion (b).

In their seminal paper, Pierrehumbert and Steele (1989) performed an *imitation task* with the two intonation patterns of the abovementioned utterance *Only a millionaire* (see Figure 50.1). They created a synthesized continuum of several steps of alignment between the two, and asked subjects to imitate the utterance. The results of the imitation task revealed the existence of two separate phonological categories. The authors argued that if the subjects had been able to reproduce the full range of the continuum in their imitation, peak alignment differences could be regarded as gradient. However, since they found that by and large the distribution of peak alignment was bimodal in the imitation data, they therefore concluded that the distinction between early and late peak alignment was categorically distinct.

Pierrehumbert and Steele's paper represented an important first step in a series of experimental investigations on the perception of tonal alignment (see CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). Since then, a body of experimental research has demonstrated that tonal alignment cues *intonational meaning distinctions* in a number of languages (e.g. Kohler 1987 and Niebuhr 2007 for German; D'Imperio and House 1997 and D'Imperio 2000 for Neapolitan Italian; Gili-Fivela 2009 for Pisa Italian; Dilley 2007 for English). The issue of whether a certain pair of intonational contrasts can be accompanied by categorical differences in meaning and whether these contrasts are perceived in a discrete or a gradient fashion has been an important research question in the field of intonation. A number of experimental methods have been used to study what is categorical or linguistic in intonation and what is paralinguistic and gradient (see the reviews in Gussenhoven 2004, 2006; also CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY). In what follows we review recent studies that have provided evidence from a number of languages on the role of tonal alignment in encoding intonational distinctions. All in all, these articles provide robust experimental evidence for the claim that changes in F0 alignment of peaks and valleys are especially salient and cue phonological distinctions across languages. This evidence has been generally interpreted as direct support for AM theory, as tonal alignment differences in this model are encoded phonologically at the pitch accent level.

Kohler's (1987) paper was the first to apply the *Categorical Perception* paradigm to alignment data and to show that alignment contrasts can be perceived categorically. The Categorical Perception paradigm involves firstly an identification/classification task in which the listeners have to categorize stimuli taken from a continuum, and secondly a discrimination task in which listeners are asked to judge pairs of stimuli as being either the same or different. For perception to be considered categorical, a peak of discrimination is expected at the point in the acoustic domain that separates the two categories (for a review, see Dilley 2007). Kohler (1987) employed the complete paradigm to investigate the perception of a set of F0 contours in German involving rises with a continuum created between early and medial peaks. He found that the early peak was associated with finality ("knowing," "coming to the end of an argument"), and the medial peak with openness ("observing," "starting a new argument"). The results of both tasks of the paradigm revealed categorical changes in the identification of early *vs.* medial peaks, with a discrimination maximum across the category boundary. More recently, Niebuhr (2007) carried out a series of perception experiments with the same German alignment contrasts and showed that the function-based identification of the peak categories is influenced not only by peak synchronization, but also by peak shape and height. In general, though, his findings corroborate the existence of the two categories in German intonation and support the idea that the timing of the peak movements with regard to the accented vowel is important for their perceptual differentiation.

Similar results have been obtained for American English tonal alignment contrasts. Following Pierrehumbert and Steele's (1989) investigation, a number of studies have examined the distinction between an early-aligned pitch accent (L+H*) and a late-aligned pitch accent (L*+H) in American English. In the most comprehensive study, Dilley (2007) conducted a series of perception experiments with the two pairs of accents attested in American English (H* and H+L*, and L* and L+H*), an identification task, two types of discrimination tasks and an imitation task. Evidence of discrimination maxima that aligned well with identification crossover points in the identification task demonstrated categorical perception for intonation and provided converging evidence with earlier results by Kohler (1987). Moreover, converging evidence for the categorical perception of intonation categories was obtained from the imitation study.

Though Kohler (1987) and Dilley (2007) are advocates of the application of the Categorical Perception paradigm to intonation, few other studies have shown clear evidence of categorical perception, i.e. with a clear discrimination peak in the expected position. The discrimination functions observed differ between studies, and in the majority of cases no discrimination peaks appear at the category crossover point revealed by the identification test. One such example is described in Gili-Fivela's (2009) article. She investigated the contrast between narrow focus and narrow contrastive focus in Pisa Italian, represented as H* and H*+L. In Pisa Italian, as in other languages, narrow contrastive focus is expressed through the use of retracted pitch peaks and an increase in pitch height. Gili-Fivela applied the Categorical Perception paradigm to the data, with both identification and discrimination tasks being performed, and also an imitation task. She manipulated both the alignment and scaling patterns of a rising pitch accent in narrow focus and a rising-falling pitch accent in contrastive narrow focus. The results showed that while there is a clear difference between a narrow focus pattern and a contrastive focus pattern in production, the contrast might not be categorically perceived, as the identification and discrimination functions do not correspond to an abrupt shift in identification aligned with a discrimination peak.

Other studies have shown that the slope of the rise and the shape of the peak also contribute to tonal contrast identification. D'Imperio and House (1997) and D'Imperio (2000) investigated the distinction between questions and statements in Neapolitan Italian. In Neapolitan Italian, questions and statements are characterized by a rise in pitch that occurs in the vicinity of the accented syllable. The materials in D'Imperio and House (1997) consisted of a series of stimuli in which the F0 peak of a rising-falling pitch accent was shifted forward and backwards within the accented syllable. Neapolitan listeners performed an identification task in which they listened to the stimuli and then classified each of them as either a question or a statement. The results showed that questions and statements are primarily distinguished by the relative alignment of the rise in a rise-fall pattern in the accented syllable. In subsequent experiments using this same contrast, D'Imperio (2000) showed that both details of the temporal alignment of target tones and the shape of the peak contribute to the identification of the contrast between questions and statements in this language. Moreover, she found that syllable structure detail modifies acoustic target alignment but does not modify the crossover point between the two categories (for more details, see §4).

New experimental paradigms have been recently applied to study the role of tonal alignment in spoken language processing. Chen *et al.* (2007) adopted the *eye-tracking paradigm* to investigate the role of pitch accent type and deaccentuation in online processing of information status in British English.[1] It was found that

[1] For a review of the eye-tracking paradigm applied to prosody research, see Watson *et al.* (2006) and Watson *et al.* (2008).

two types of pitch accents (H*L and L*HL) create a strong bias toward newness, whereas deaccentuation and the L*H pitch accent create a strong bias toward givenness. Watson *et al.* (2008) also used the eye-tracking paradigm to investigate whether the presence of a pitch accent difference between L+H* and H* in English biases listeners toward interpreting a temporarily ambiguous noun as referring to either a discourse-given or a discourse-new entity. Participants had to perform a word-recognition task (for example, *candle vs. candy*) and pick up one of the two competing objects, while their eye movements were being monitored. They found that although listeners interpreted these accents differently, their interpretive domains overlapped. L+H* created a strong bias toward contrast referents, whereas H* was compatible with both new and contrast referents.

The electro-encephalography (EEG) technique, a procedure which measures electrical activity of the brain and which allows for the non-invasive measuring of brain activity during cognitive processing, has also been used to study pitch processing. For example, Fournier *et al.* (2010) used this technique to investigate the tonal and intonational pitch processing of some tonal contrasts (some of them alignment contrasts) by native speakers of the tonal dialect of Roermond Dutch as compared to a control group of speakers of Standard Dutch, a non-tone language. A set of words with identical phoneme sequences but distinct pitch contours, which represented different lexical meanings or discourse meanings (e.g. statement *vs.* question), were presented to both groups. The stimuli were arranged in a mismatch paradigm, under several experimental conditions: in the first condition (lexical), the pitch contour differences between stimuli reflected differences between lexical meanings; in the second condition (intonational), the stimuli differed in their discourse meaning. In these two conditions, both native and non-native responses showed a clear magnetic mismatch negativity in a time window from 150 to 250 msecs after the divergence point of standard and deviant pitch contours. In the lexical condition, a stronger response was found over the left temporal cortex of speakers of standard as well as non-standard Dutch. Crucially, in the intonational condition, the same activation pattern was observed in the control group, but not in the group of Roermond Dutch speakers, who showed a right-hemisphere dominance instead. Thus the lateralization of pitch processing was condition-dependent in the Roermond Dutch group only, suggesting that processes are distributed over both temporal cortices according to the functions available in the grammar.

Finally, researchers have begun to consider the role of potential *articulatory landmarks* and the coordination or alignment between *tonal gestures* (measured as F0 turning points) and *oral constriction gestures*. Recent work by Mücke *et al.* (2006), D'Imperio *et al.* (2007a), and Mücke *et al.* (2009) has investigated alignment patterns for three different languages (Italian, German, and Catalan) by using electromagnetic mid-sagittal articulography (EMMA) for capturing oral constriction gestures alongside acoustic recordings. The end of pitch movements in bitonal pitch accents co-occurs with the minima and maxima of the closing gesture of C_2 in $C_1V.C_2$ and C_1VC_2 sequences. In all these studies, such pitch targets were seen to be more closely aligned in time with articulatory landmarks than with acoustic ones. However, there was some variation as to the articulatory landmark which served as an anchor for the tonal target. For example, in German nuclear LH accents, the H peaks co-occurred with the intervocalic C target, whereas in pre-nuclear accents peaks co-occurred with the target for the following vowel (what is called "accent shift"; Mücke *et al.* 2009). In Catalan it was the consonantal peak velocity rather than the consonantal target which served as the landmark. Such an apparently small alignment difference in the articulatory anchor type may be used by speakers to make (or contribute toward making) phonological distinctions, as in Neapolitan, where H in L*+H (questions) aligns with the maximum constriction, and H in L+H* (statements) aligns with peak velocity (see D'Imperio *et al.* 2007a).

3 Phonological encoding: Tonal association and tonal alignment

The topic of this section is the relation between phonological association and phonetic alignment of tones and how it is encoded in a representational system. The starting point is provided by the autosegmental metrical approach to intonation, which has developed an explicit phonological representational approach that has been applied to a variety of languages (Pierrehumbert 1980; Pierrehumbert and Beckman 1988; Ladd 1996; Gussenhoven 2004; among others). Though the AM representational proposal can account for melodic patterns in a variety of languages, there are a number of areas that remain unresolved. Two of these issues relate to how to interpret the relationship between tones and metrically strong syllables in the AM model, namely the concept of starredness on the one hand and the interpretation of the relationship between phonological association and phonetic alignment on the other.

The AM phonological representation of pitch accents encodes "autosegmental" information (or pitch accent shapes, LH or HL) and "metrical" information, i.e. information about the association of tones with metrical constituents and the relative alignment of tones with the metrically prominent syllable. The surface alignment of tones is basically derived from the use of the star notation (*). The star notation encodes two complementary things: (i) phonological association between pitch accent shapes and stressed syllables – in other words, a tone gets a star when it is associated to a metrically strong position; (ii) relative alignment in bitonal pitch accents – i.e. the tone that gets the star is the one that is directly linked to the metrically strong position. In bitonal accents, the question of which tone in LH or HL accent shapes should be assigned a star is not completely straightforward. On this issue, Pierrehumbert's original definition states that "a strength relationship is defined on the two tones of bitonal accents and that it is the stronger tone which lines up with the accented syllable" (Pierrehumbert 1980: 76–77). According to this definition, it is ambiguous whether the star notation * indicates phonetic alignment between the tonal unit and the stressed syllable or just a "looser" phonological association. Similarly, Pierrehumbert and Beckman (1988: 234) note that "the * diacritic marks which tone of a bitonal accent is aligned with stress." Arvaniti *et al.* (2000: 120) state that "phonetically this use of the star is to be interpreted as signifying that the starred tone is aligned in time with the stressed syllable." In subsequent work, one of the most common interpretations of the star notation is that the starred tone is phonetically aligned with the stressed syllable, and thus a strict temporal alignment between the tone and its tone-bearing unit is expected.

Recently, attention has been drawn to the various problems created by the representational ambiguity of the star notation. One of them is that it can be difficult to decide between competing AM analyses of bitonal accents, because the same contours can be transcribed in different ways (Prieto *et al.* 2005). For example, let us compare the surface alignment of the tones described by the English and Spanish L+H* – L*+H contrasts according to, respectively, Pierrehumbert (1980) and Sosa (1999). Even though the two phonological units capture the two-way phonological contrast present in both languages, the same labels L+H* and L*+H refer to different phonetic realizations (or alignment patterns) in the two languages. In fact, English L+H* corresponds to Spanish L*+H. This difference between the notational systems is caused by different interpretations of the star notion: while in the English notation the star is interpreted as an indication of phonological association between the tone and the prominent syllable, in Spanish it is interpreted as phonetic alignment, that is, the star is indicating whether the H peak is aligned (H*) or not aligned (L*) with the stressed syllable.

(2) *Schematic representation of* L+H* *and* L*+H

a. *English* (after Pierrehumbert 1980) b. *Spanish* (after Sosa 1999)

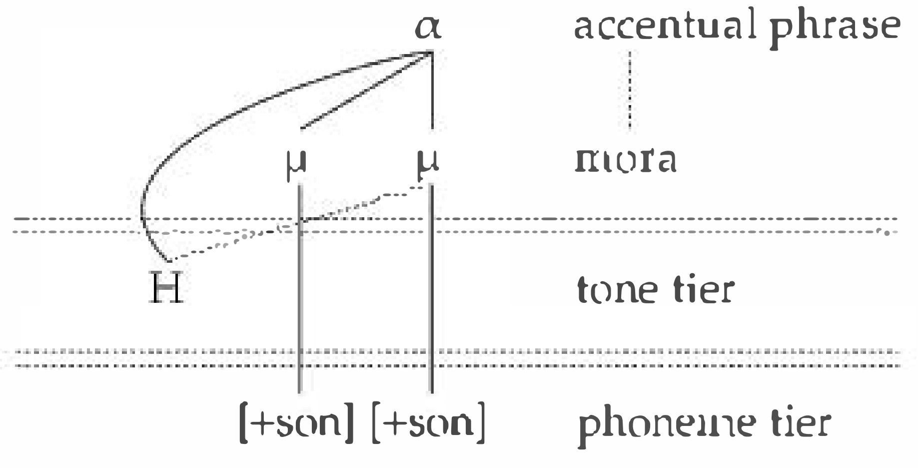
In addition, some authors have pointed out that the theoretical concept of starredness is ill-defined and cannot be based solely on phonetic alignment (Arvaniti *et al.* 2000). Arvaniti *et al.* present evidence from Greek of the types of problems that arise when phonetic alignment to the accented syllable is taken to be the exponent of association of tones with segments. As they note:

> we show that there exist pitch accents that are clearly bitonal but neither tone is, strictly speaking, aligned with the accented syllable. We argue from this fact that association cannot be based on phonetic alignment in any straightforward way and that a more abstract and rigorously defined notion of starredness is required.

In Greek rising pitch accents in pre-nuclear position, typically, neither L nor H is phonetically aligned with the stressed syllable: in most cases, the L is consistently aligned before the beginning of the accented syllable (5 msecs on average before the onset), and H displays more variability and is typically located in the post-tonic. Thus, these authors conclude that "if alignment is the sole exponent of the association of tones to segments, phonetic variability in this domain becomes a crucial issue when the phonological structure of a bitonal accent is in question" (Arvaniti *et al.* 2000: 121). We take it as essentially correct that a one-to-one relationship between phonological association and phonetic alignment is difficult to maintain in the current AM model.

In a recent proposal, Prieto *et al.* (2005) describe the contrastive possibilities of alignment of rising accents in three Romance languages, Central Catalan, Neapolitan Italian, and Pisa Italian (see also CHAPTER 2: CONTRAST). According to these authors, these Romance languages provide evidence that small differences in alignment in rising accents must be encoded phonologically. To account for such facts within the AM model, they develop the notion of "phonological anchoring" as an extension of the concept of secondary association originally proposed by Pierrehumbert and Beckman (1988). They propose that the phonological representation of pitch accents needs to include two independent mechanisms to encode alignment properties with metrical structure: (i) encoding of the primary phonological association (or affiliation) between the tone and its tone-bearing unit; and (ii), for some specific cases, encoding of the secondary phonological anchoring of tones to prosodic edges (i.e. moras, syllables, and prosodic words). (3) shows the schematic representation of the primary and secondary associations of a phrasal H within the accentual phrase in Japanese (Pierrehumbert and Beckman 1988: 129). The solid line indicates primary association to the accentual phrase α and the dashed line secondary association to the second sonorant mora μ within the accentual phrase.

(3) *Japanese* (after Pierrehumbert and Beckman 1988: 129)



The Romance data provide crucial evidence of mora-edge, syllable-edge, and word-edge H tonal associations and suggest that not only peripheral edge tones seek secondary associations. In this way, the specification of metrical anchoring points in the phonological representation offers a more transparent analysis of the alignment contrasts found in Romance languages and, ultimately, can help in the task of defining a more transparent pitch accent typology. Finally, Prieto *et al.* (2005) argue that such an approach makes the mapping from phonological representation to surface alignment patterns more explicit, and that it thus allows for more straightforward cross-linguistic comparisons.

The evidence described above shows that even though AM representations are adequate when it comes to characterizing the minimal contrasts in pitch accent types found in different languages, the proper procedures by which to map phonological representations and the surface alignment of tones (through the use of the star notation) are still somewhat unclear. This is because the specific details of the coordination between tones and the segments that are linked to the structural unit are not part of the phonological representation itself. We thus agree with Arvaniti *et al.*'s (2000: 130) suggestion "that the task for the future is to refine the notion of the phonological association of tones in intonational systems." In the near future, the contrastive possibilities of alignment found cross-linguistically need to be explored. This will provide firm ground from which to advocate a further refinement of the metrical side of the AM model.

4 Phonetic models of tonal alignment

Apart from changes in tonal alignment which have phonological effects, i.e. which encode a difference in meaning (see §2 and §3), tonal alignment is influenced by a variety of phonetic factors, such as tonal crowding, speech rate, segmental composition, and syllable structure composition. These fine-grained F0 alignment differences do not affect meaning or representation, and are instead considered to arise from differences in phonetic implementation rather than phonological representation. In this section we review some of the production studies that have investigated the influence of such factors on tonal alignment patterns and the perception studies that have demonstrated that some of these effects are employed by native speakers in lexical access tasks.

Cross-linguistically, the location of fundamental frequency peaks (or H values) has been shown to be greatly affected by the right-hand prosodic context, in such a way that the peak is retracted before upcoming pitch accents and boundary tones (see Silverman and Pierrehumbert 1990 for English and Prieto *et al.* 1995 for Spanish, for example). Prieto *et al.* (1995) examined the peak placement patterns in rising accents in Spanish and found the following: (i) the location of the start of the F0 rise is fairly constant (generally at the onset of the accented syllable); (ii) as in English, the duration of the rising gesture is highly correlated with syllable duration. These results show that the slope and/or duration of a speech F0 movement are not constant, as claimed by the *fixed rise-time hypothesis* (Fujisaki 1983; 't Hart *et al.* 1990; and others), but are instead governed by the coordination of the movement with the segmental string. Both studies demonstrated that a successful quantitative model of peak placement must contain at least two factors, namely the duration of the accented syllable and the distance in syllables to upcoming pitch accents or boundary tones.

The *Segmental Anchoring Hypothesis* (henceforth SAH), as articulated by Ladd *et al.* (1999) on the basis of work by Prieto *et al.* (1995) and Arvaniti *et al.* (1998), refers to the idea that the slope of tonal movements is not invariant, but rather is specifically related to segmental anchors. Arvaniti *et al.* (1998) found an unexpected and consistent stability effect when little or no tonal pressure was exerted on the pitch accent. In a Greek word such as [pa'ranoma] 'illegal', the H target in the LH pitch accent associated with the test stressed syllable ['ra] was consistently aligned over – or "anchored to" – the frontier between the post-accentual onset and the following vowel ([n] and [o]). This clearly contradicts the traditional "constant slope" and "constant duration" hypotheses (i.e. the fixed rise-time hypothesis: Fujisaki 1983; 't Hart *et al.* 1990; and others). The SAH says that both the beginning and the end of a rising or falling F0 movement are anchored to specific points in the segmental string, such as the beginning of the stressed syllable or the following unstressed vowel, and consequently the duration of the F0 movement is strongly dependent on the duration of the segmental interval between the anchor points. As we will see below, work on the effects of lower prosodic structure levels such as the syllable or the prosodic word on tonal alignment shows that we need to refine the SAH to incorporate these findings.

Recent work on tonal alignment in different languages has shown that the position of the peak tends to change across syllable structure types (e.g. Rietveld and Gussenhoven 1995 for Dutch; D'Imperio 2000 for Neapolitan Italian; Prieto and Torreira 2007 for Peninsular Spanish; Prieto 2009 for Catalan). For example, D'Imperio (2000) found that the peak was located closer to the vowel offset in closed syllables in Neapolitan Italian. While in open syllables the peak was aligned with the end of the accented vowel, in closed syllables the peak was somewhat retracted and located within the coda consonant. This same effect of coda consonants on alignment has been detected in both rising accents in various languages (see citations above) and falling nuclear accents in Catalan (Prieto 2009). The results indicate that while the beginning of the falling accent gesture (H) is tightly synchronized with the onset of the accented syllable, the end of the falling gesture (L) is more variable and is affected by syllable structure: in general, while in open syllables the end of the fall is aligned roughly with the end of the accented syllable, in closed syllables it is aligned well before the coda consonant.

D'Imperio *et al.* (2007b) hypothesized that Neapolitan listeners might capitalize on the alignment regularity for the perception of lexical contrast. Specifically, their hypothesis was that listeners of Neapolitan Italian might identify more closed syllable items when tonal alignment details are congruent with those for this type of syllable structure (see also Petrone 2008). In order to test this hypothesis, two natural productions of the words *nono* 'ninth' and *nonno* 'grandfather', both carrying a yes/no question nuclear accent, were manipulated in two ways. First, the researchers modified the length of the stressed vowel and the following consonant in five steps, in order to shift the perception of each item from *nono* to *nonno* and *vice versa*. Then, tonal alignment was shifted earlier, in four steps, without changing the percept of the question to that of a statement but merely creating question patterns that would be more or less congruent with the syllabic structure of the base. Thirteen Neapolitan listeners identified the stimuli as either *nono* or *nonno*. Significantly, the results showed that the alignment manipulation produced a category boundary shift in the *nonno* base stimulus series, but no effect in the open syllable series, supporting the hypothesis that fine detail of tonal alignment not only is employed to signal pragmatic contrast but may also be stored as part of the phonological specification of lexical items.

Similarly, acoustic work on a variety of languages has shown that H peaks are consistently affected by the position of the accented syllable within the word (for English, see Silverman and Pierrehumbert 1990, and for Spanish Prieto *et al.* 1995). In general, peaks tend to shift backwards as their associated syllables approach the end of the word: in other words, the distance from the beginning of the accented syllable to the peak is longer in words with antepenultimate stress than in words with penultimate stress, which in turn show a longer distance than in words with final stress. In order to correct for the potentially confounding effects of stress clash (or distance to the next accented syllable), Prieto *et al.* (1995) analyzed a subset of the data obtained from test syllables in different positions in the word (e.g. *<u>nú</u>mero* 'number', *nu<u>me</u>ro* 'I number', *nume<u>ró</u>* '(s)he numbered'). Their materials consisted of word sequences in which there was a distance of two unstressed syllables between one accented syllable and the next (e.g. *<u>nú</u>mero <u>rá</u>pido* 'quick number', *nu<u>me</u>ro ner<u>vio</u>so* 'I number nervously', and *nume<u>ró</u> regu<u>lar</u>* '(s)he numbered in a regular way'). The three diagrams in Figure 50.2 show a schematic representation of the difference in F0 timing patterns in the three conditions, *número rápido*, *numero nervioso*, and *numeró regular*. A significant effect of word position on different measures of peak alignment was found in all the comparisons. Similarly, in Silverman and Pierrehumbert's (1990) model of F0 peak location, the dropping of the variable "Word-Boundary" (while leaving the variable "Stress Clash" as a main predictor) significantly worsened the fit of the model.

Prosodic word effects seem to suggest the possibility that the end of the word (and not only the presence of upcoming accents or boundary tones) is acting as a kind of prosodic boundary that exerts prosodic pressure on H tonal targets and

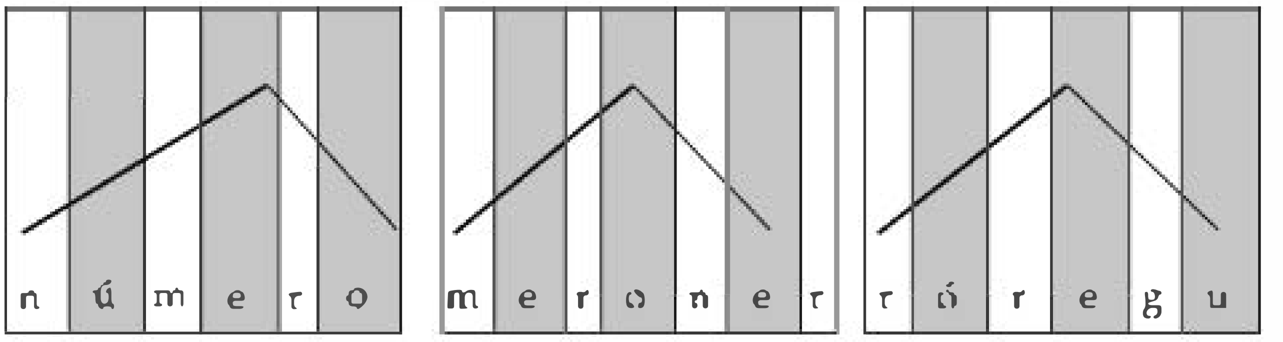


Figure 50.2 A schematic representation of the difference in F0 timing patterns in the three conditions, *número rápido*, *numero nervioso*, and *numeró regular*

that this effect can be exploited in word boundary identification tasks. Prieto *et al.* (forthcoming) performed a set of production and perception experiments that dealt with potentially ambiguous utterances distinguished by word boundary location in Catalan and Spanish (e.g. Catalan *mirà batalles* '(s)he looked at battles' *vs.* *mirava talles* 'I/(s)he used to look at carvings'; Spanish *da balazos* '(s)he fires shots' *vs.* *daba lazos* 'I/(s)he gave ribbons'). For the perception experiments, they hypothesized that relative peak location would help Catalan and Spanish listeners in terms of lexical access. The results of the production experiments showed that the prosodic word domain has a significant shifting effect on F0 peak location, and the results of the perception experiments showed that these alignment patterns are actively used by listeners in word identification tasks.

In general, the results of studies on lexical access (D'Imperio *et al.* 2007b; Prieto *et al.*, forthcoming) support the hypothesis that listeners are able to employ fine allophonic details of H tonal alignment due to syllable structure or within-word position to identify lexical items. This empirical evidence demonstrates that prosodic structure must play an essential role in our understanding of the coordination of pitch gestures with the segmentals and argues in favor of a view supported by other work that prosodic structure is manifested in details of articulation.

5 Tonal alignment: Evidence for target- *vs.* configuration-based theories of intonation

As pointed out in §2, work on tonal alignment has provided robust experimental evidence that changes in the synchronization of peaks and valleys with segmental landmarks are key perceptual cues for phonological distinctions across languages. This evidence has been interpreted as direct support for AM theory, which is widely held to afford a number of advantages over other discrete tone theories, as tonal alignment differences in this model are encoded phonologically in pitch accent units.

Alignment studies have also been used to test specific predictions about different phonological models of prosody and intonation. For example, one of the old controversies in intonation studies surrounds the relative merits of the *target-based vs. configuration-based* theories of intonational primitives (see Ladd 1996: §1.2 for a review; also Arvaniti and Ladd 2009). The target-based model (also called *target-and-interpolation model* by Arvaniti and Ladd 2009) is the phonetic basis of AM intonational phonology, which has become the dominant phonological framework for analyzing intonation. This model assumes that certain points in the contour (e.g. local targets or F0 maxima and F0 minima) reflect phonologically specified targets and thus derive the intonational contour by defining the tonal targets and then connecting those through an interpolating F0 curve that goes from one target to the next. In recent years there has been accumulating evidence from tonal alignment studies that L and H tones behave as static targets and that they align with the segmental string in extremely consistent ways. Typically, in a variety of languages, the L valley of pre-nuclear rises is precisely aligned with the beginning of the accented syllable (see Prieto *et al.* 1995 for Spanish, Arvaniti *et al.* 1998 for Greek, and Ladd *et al.* 1999 for English, for example). Moreover, some studies have shown that this precise L intonational alignment with word or syllable boundaries is used by listeners in lexical identification tasks. For example, Ladd and Schepman (2003) showed that the different alignment of L in minimal pairs like *Norman Elson/Norma Nelson* is a useful cue to the word-boundary distinction between them. If L alignment was modified experimentally in such ambiguous phrases, this affected the listeners' judgments in the identification task. Similarly, a recent study on the tonal marking of the French Accentual Phrase (AP) by Welby (2003) showed that the L tone associated with the left edge of the first content word of the AP is aligned at the boundary between the last function word and the first syllable of the first content word. Welby's results for perception showed that French listeners use the alignment of the L tone as a cue for lexical access (in pairs such as *mes galops* 'my gallops' and *mégalo* 'megalomaniac'). All in all, these alignment results, as well as many scaling results, have been interpreted in favor of the target-based hypothesis (for a review, see Ladd 1996).

On the other hand, *configuration-based theories* (also called *concatenation models* by Arvaniti and Ladd 2009) treat the contour as the result of stringing together entire tonal sequences (not necessarily straight lines) of various lengths. Traditional intonational descriptions of the so-called "British school" (e.g. Crystal 1969; O'Connor and Arnold 1973) and the approach adopted by the Eindhoven-based *Instituut voor Perceptie Onderzoek* (IPO; e.g. 't Hart *et al.* 1990) are of this sort, as is the more recent syllable-concatenation model proposed by Xu and colleagues (e.g. Xu and Wang 2001; Xu and Xu 2005). There have been several results reported in the literature that provide support for a configuration-based theory of intonation. For example, as mentioned above, D'Imperio and House (1997) undertook a perception experiment that investigated the contrast between questions and statements in Neapolitan Italian. They wanted to determine whether the major perceptual cue to this category distinction involved only the temporal alignment of the high-level target with the syllable or if instead the category percept also depended on the presence of a rising or falling melodic movement within the syllable nucleus. The results showed that the primary perceptual cue for questions is a rise through the vowel, while the primary cue for statements is a fall through the vowel. D'Imperio and House claimed that their results confirmed the second hypothesis, in that perceptually a rise in the vowel was the most important cue for the question, while a fall in the vowel was the most important cue for the statement, thus supporting the notion that pitch movements through areas of stability are perceptually important for identifying tonal categories.

Contrasting results were obtained by Arvaniti and Ladd (2009), who carried out a production study in which they used acoustic alignment measures to test specific predictions about different phonological models of intonation. This involved

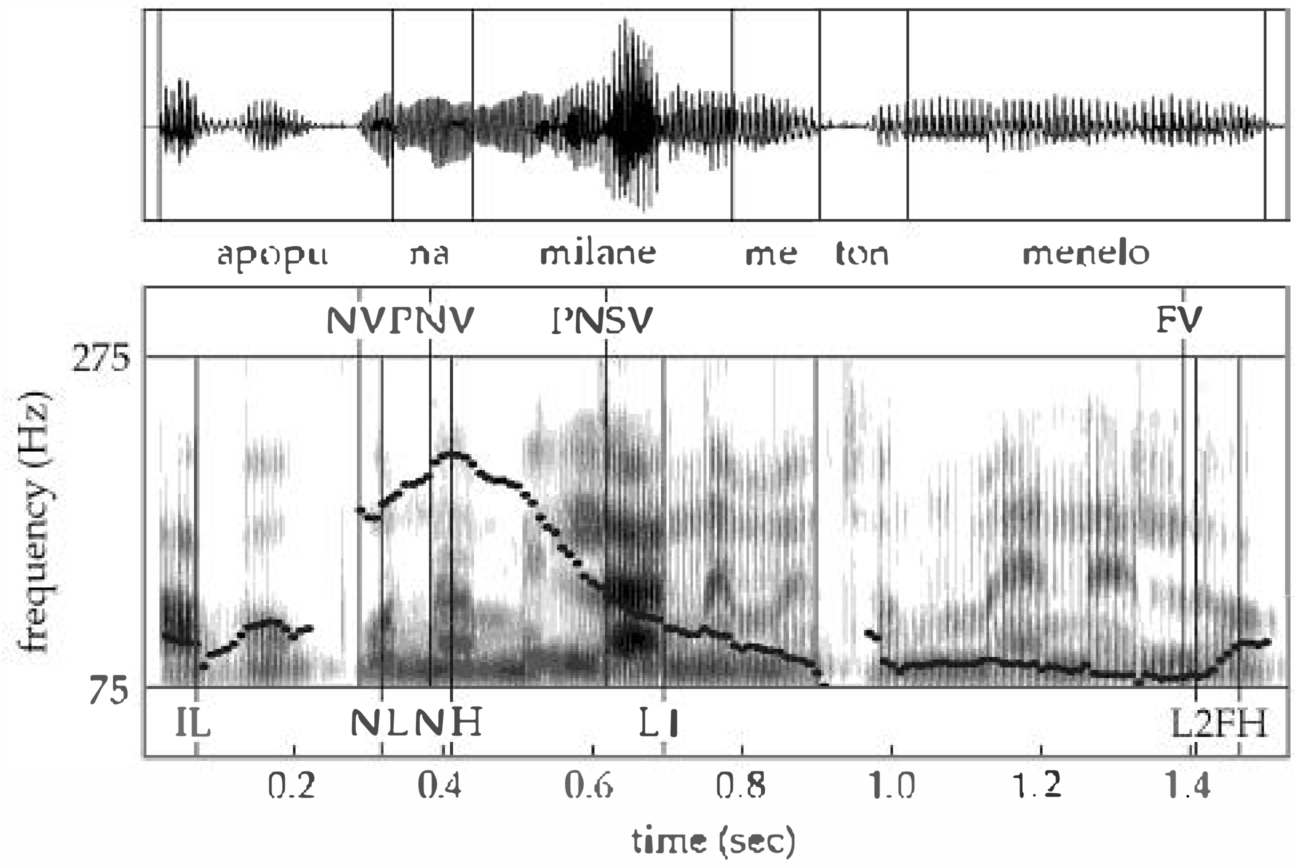


Figure 50.3 The waveform, spectrogram, and F0 contour of [apo'pu na mi'lane me ton 'menelo] 'Where could they be speaking to Menelos from?' (speaker KP), illustrating the measurements taken on the F0 contour and relevant segmental onsets. Figure reproduced from Arvaniti and Ladd (2009: 55)

undertaking a very detailed phonetic study of the Greek wh-question melody. According to their results, certain points in the Greek wh-question melody show little variability in scaling and predictable variability in alignment. A close analysis of the F0 alignment data showed that (i) the exact contour shape depended on the length of the question, and (ii) the position of the first peak and the low plateau depended on the position of the prominent anchor syllables. The study also showed predictable adjustments in alignment depending on the proximity of adjacent tonal targets. Figure 50.3 shows the F0 contour of a long wh-question. In long wh-questions, the contour starts with a rise from a low F0 point, the fall from the peak is relatively shallow, and the following low F0 stretch is long. By contrast, short wh-questions consist of a high tone associated in time with the stressed syllable of the wh-word, followed by a rapid fall to a stretch of low F0, followed by a small rise.

Arvaniti and Ladd (2009) argue that the Greek wh-question data strongly argue in favor of a *target-based model* of intonational phonology like that proposed by the autosegmental metrical framework of intonational phonology, and in particular in favor of the notion of sparse tonal specification. This is because one key assumption of the autosegmental metrical framework is that there is not necessarily any role for the syllable in modeling utterance contours. Rather, F0 targets can be temporally anchored to the segmental string in a variety of ways. This is exactly what we find in the wh-contour data in Greek, as the alignment and scaling adjustments observed in the contour are totally predictable, and depend on the length and tonal crowding manipulations in the target utterance. Arvaniti and Ladd claim that these predictable effects cannot be explained by superposition models of intonation, such as Fujisaki's (1983) command–response model, or by configuration-based models that specify F0 by superposing contour shapes for shorter and longer domains, since both of them lack the mechanisms to account for effects such as the truncation of targets or asymmetrical adjustments to the larger tonal domains. Similarly, models that specify the F0 of all syllables (like Xu and colleagues' model), and thus assume that all syllables are specified for tone, cannot account for lawful variation except by using ad hoc tonal specifications, which, in turn, do not allow for phonological generalizations about contours applying to utterances of different lengths.

6 Conclusion

In recent decades, the issue of tonal alignment has been a key focus of phonological research in intonational phonology. We now have solid evidence coming from different languages that F0 alignment differences can convey intonational contrasts, and that these alignment differences can be perceived in a near-categorical way. In this chapter, we have reviewed this work, and the use of several techniques in the investigation of tonal alignment processing (§2). As we have seen, a wide range of methodological paradigms have been applied to alignment research, including acoustic and articulatory analyses of speech productions, judgments and reaction times obtained during identification and discrimination tasks, measurements of brain activity, and eye movements.

A recent debate within the autosegmental metrical approach to intonation has been how to represent these phonological contrasts in tonal alignment. As has been reported before, this theory does an especially good job of accounting for why tone alignment differences can convey intonational contrasts. In the AM framework, the star notation encodes both phonological association of the tones with a stressed syllable and the relative alignment in bitonal pitch accents. However, though the AM representations can adequately characterize the minimal contrasts in pitch accent types found in different languages, the procedures for mapping the surface alignment of tones through the use of the star notation onto phonological representations are still somewhat unclear. This chapter has reviewed some recent proposals regarding this issue which highlight the need to further investigate the contrastive possibilities of alignment found cross-linguistically.

Apart from the phonological contrasts induced by tonal alignment, F0 tonal patterns are influenced by a variety of phonetic factors, such as prosodic crowding, speech rate, segmental composition, upcoming syllable structure, and prosodic word boundaries. In this case these fine-grained F0 alignment differences do not affect intonational meaning. This chapter has reviewed some of the production and perception studies that have informed the current phonetic models of tonal alignment. This work has highlighted principles of stability and also of adaptation to neighboring prosodic structure as basic pillars of phonetic models of tonal alignment. Importantly, some of these alignment patterns have been shown to be actively used by listeners in word identification tasks and lexical access.

Finally, tonal alignment issues have historically been used as arguments to test the predictions of phonological models of intonation and to bear upon current theories of intonational phonology. The last section of this chapter has offered a selection of the arguments put forth in favor of the target-based model of intonation. As a final note, we believe that the full exploitation of recent methodological advances will provide important answers to the role of tonal alignment in phonological and phonetic models of intonation.

ACKNOWLEDGMENTS

Preparation of this chapter was supported by two grants awarded by the Spanish Ministerio de Ciencia e Innovación: projects FFI2009-07648/FILO and CONSOLIDER-INGENIO 2010 "Bilingüismo y Neurociencia Cognitiva CSD2007-00012." I am grateful to Rafèu Sichel, two anonymous reviewers, and the editors of this volume, Colin Ewen, Beth Hume, Marc van Oostendorp, and Keren Rice, for all their help and very useful comments on an earlier version of the manuscript.

REFERENCES

Arvaniti, Amalia & D. Robert Ladd. 2009. Greek wh-questions and the phonology of intonation. *Phonology* 26. 43–74.

Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen. 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics* 26. 3–25.

Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen. 2000. What is a starred tone? Evidence from Greek. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, 119–131. Cambridge: Cambridge University Press.

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.

Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.

Chen, Aoju, Els den Os & Jan Peter de Ruiter. 2007. Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. *Linguistic Review* 24. 317–344.

Crystal, David. 1969. *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

Dilley, Laura C. 2007. The role of F0 alignment in distinguishing categories in American English intonation. Unpublished ms., Bowling Green State University.

D'Imperio, Mariapaola. 2000. On defining tonal targets from a perception perspective. Ph.D. dissertation, Ohio State University.

D'Imperio, Mariapaola & David House. 1997. Perception of questions and statements in Neapolitan Italian. *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech 1997)*, vol. 1, 251–254. Rhodes, Greece.

D'Imperio, Mariapaola, Robert Espesser, Hélène Lœvenbruck, Caroline Menezes, Noël Nguyen & Pauline Welby. 2007a. Are tones aligned with articulatory events? Evidence from Italian and French. In Jennifer Cole & José Ignacio Hualde (eds.) *Laboratory phonology* 9, 577–608. Berlin & New York: Mouton de Gruyter.

D'Imperio, Mariapaola, Caterina Petrone & Noël Nguyen. 2007b. Effects of tonal alignment on lexical identification in Italian. In Carlos Gussenhoven & Tomas Riad (eds.) *Tones and tunes*, vol. 2: *Experimental studies in word and sentence prosody*, 79–106. Berlin & New York: Mouton de Gruyter.

Fournier, R., C. Gussenhoven, O. Jensen & P. Hagoort. 2010. Lateralization of tonal and intonational pitch processing: An MEG study. *Brain Research* 1328. 79–88.

Fujisaki, Hiroya. 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. In Peter F. MacNeilage (ed.) *The production of speech*, 39–55. New York: Springer.

Gili-Fivela, Barbara. 2009. From production to perception and back: An analysis of two pitch accents. In Susanne Fuchs, Hélène Lœvenbruck, Daniel Pape & Pascal Perrier (eds.) *Some aspects of speech and the brain*, 363–405. Frankfurt am Main: Peter Lang.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos. 2006. Experimental approaches to establishing discreteness of intonational contrasts. In Sudhoff *et al.* (2006), 321–334.

Hart, Johan 't, René Collier & Antonie Cohen. 1990. *A perceptual study of intonation: An experimental-phonetic approach.* Cambridge: Cambridge University Press.

Kohler, Klaus J. 1987. Categorical pitch perception. In Thomas V. Gamkrelidze (ed.) *Proceedings of the 11th International Congress of Phonetic Sciences*, vol. 5, 331–333. Tallinn: Academy of Sciences of the Estonian SSR.

Ladd, D. Robert. 1996. *Intonational phonology.* Cambridge: Cambridge University Press.

Ladd, D. Robert & Astrid Schepman. 2003. Sagging transitions between high pitch accents in English: Experimental evidence. *Journal of Phonetics* 31. 81–112.

Ladd, D. Robert, D. Faulkner, H. Faulkner & A. Schepman. 1999. Constant "segmental" anchoring of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106. 1543–1554.

Mücke, Doris, Martine Grice, Johannes Becker, Anne Hermes & Stefan Baumann. 2006. Articulatory and acoustic correlates of prenuclear and nuclear accents. In Rüdiger Hoffmann & Hansjörg Mixdorff (eds.) *Proceedings of Speech Prosody* 2006, 297–300. Dresden: TUDpress.

Mücke, Doris, Martine Grice, Johannes Becker & Anne Hermes. 2009. Sources of variation in tonal alignment: Evidence from acoustic and kinematic data. *Journal of Phonetics* 37. 321–338.

Niebuhr, Oliver. 2007. The signalling of German rising-falling intonation categories: The interplay of synchronization, shape, and height. *Phonetica* 64. 174–193.

O'Connor, J. D. & G. F. Arnold. 1973. *Intonation of colloquial English.* London: Longman.

Petrone, Caterina. 2008. From targets to tunes: Nuclear and prenuclear contribution in the identification of intonation contours in Italian. Ph.D. dissertation, Université de Provence.

Pierrehumbert, Janet B. 1980. The phonetics and phonology of English intonation. Ph.D. dissertation, MIT.

Pierrehumbert, Janet B. & Mary E. Beckman. 1988. *Japanese tone structure.* Cambridge, MA: MIT Press.

Pierrehumbert, Janet B. & Shirley Steele. 1989. Categories of tonal alignment in English. *Phonetica* 46. 181–196.

Prieto, Pilar. 2009. Tonal alignment patterns in Catalan nuclear falls. *Lingua* 119. 865–880.

Prieto, Pilar & Francisco Torreira. 2007. The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35. 473–500.

Prieto, Pilar, Jan van Santen & Julia Hirschberg. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23. 429–451.

Prieto, Pilar, Mariapaola D'Imperio & Barbara Gili-Fivela. 2005. Pitch accent alignment in Romance: Primary and secondary associations with metrical structure. *Language and Speech* 48. 359–396.

Prieto, Pilar, Eva Estebas-Vilaplana & Maria del Mar Vanrell. Forthcoming. The relevance of prosodic structure in tonal articulation: Edge effects at the prosodic word level in Catalan and Spanish. *Journal of Phonetics.*

Rietveld, Toni & Carlos Gussenhoven. 1995. Aligning pitch targets in speech synthesis: Effects of syllable structure. *Journal of Phonetics* 23. 375–385.

Silverman, Kim E. A. & Janet B. Pierrehumbert. 1990. The timing of prenuclear high accents in English. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 72–106. Cambridge: Cambridge University Press.

Sosa, Juan Manuel. 1999. *La entonación del español: Su estructura fónica, variabilidad y dialectología.* Madrid: Cátedra.

Sudhoff, Stefan, Denisa Lenertová, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter & Johannes Schließer (eds.) 2006. *Methods in empirical prosody research*. Berlin & New York: Mouton de Gruyter.

Watson, Duane G., Christine A. Gunlogson & Michael K. Tanenhaus. 2006. Online methods for the investigation of prosody. In Sudhoff *et al*. (2006), 259–282.

Watson, Duane G., Michael K. Tanenhaus & Christine A. Gunlogson. 2008. Interpreting pitch accents in online comprehension: H* *vs*. L+H*. *Cognitive Science* 32. 1232–1244.

Welby, Pauline. 2003. The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation. Ph.D. dissertation, Ohio State University.

Xu, Yi & Emily Wang. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33. 319–337.

Xu, Yi & Ching X. Xu. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33. 159–197.

51 The Phonological Word

ANTHI REVITHIADOU

1 Introduction

In the past few decades, the field of phonology has witnessed the development of an assortment of phonological theories and their offshoots. Seminal among them is the theory of *Prosodic Phonology*, which explores how prosodic structure is built in relation to morphosyntactic structure. Prosodic Phonology employs mapping rules that aim at organizing chunks of structure (e.g. strings smaller or larger than the grammatical word) into hierarchically ordered layers of prosodic units which, in turn, form the domains within which phonological rules apply. Such *phonological domains* need not be isomorphic to morphosyntactic constituents. More importantly, the existence of a mapping mechanism entails that rules of phonology proper (i.e. rules inducing changes in the phonological shape and pattern of a string of elements) do not make direct reference to morphosyntactic constituents.[1] In general, the basic tenet of Prosodic Phonology is that phonological rules cannot see nor refer to any structure other than the units of the *Prosodic Hierarchy* (Selkirk 1978b, 1980, 1981a, 1981b, 1984, 1986, 1995; Nespor and Vogel 1982, 1986; Hayes 1989; see also CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE, CHAPTER 40: THE FOOT, CHAPTER 48: STRESS-TIMED *VS.* SYLLABLE-TIMED LANGUAGES, CHAPTER 54: THE SKELETON, CHAPTER 104: ROOT–AFFIX ASYMMETRIES, and CHAPTER 56: SIGN SYLLABLES for some other aspects of the Prosodic Hierarchy):

[1] In this sense, therefore, Prosodic Phonology sharply contrasts with *direct reference* approaches to the interface, most articulately expressed in the work of Kaisse (1983, 1985) and Odden (1987, 1990).

(1) *Prosodic Hierarchy*

Utterance (U)
|
Intonational Phrase (IP)
|
Phonological Phrase (PPh)
|
<Clitic Group> (CG)[2]
|
Phonological Word (PW)
|
Foot (F)
|
Syllable (σ)

The *Phonological Word* (PW) is one of the best-established constituents of the Prosodic Hierarchy, and has received a great deal of attention. Despite its wide acceptance, certain aspects of the PW are still under debate in the literature. Part of the controversy surrounding its name pertains to the fact that the PW has broadly been established as the constituent that mediates the interface of phonology with morphology (i.e. lexical component) and syntax (i.e. post-lexical component), even though originally it was not intended to encapsulate both aspects of the interface. Moreover, it is the prosodic constituent which, roughly, corresponds to a grammatical word, and hence, naturally, cannot escape the conflicts and ambiguities associated with this notion.

This chapter, therefore, aims to address all the major issues pertaining to the PW. More specifically, I will first explore the exact nature of the mapping mechanism(s) involved in the construction of this prosodic constituent. The pivotal questions center on the ways the mapping rules operate to group a specific chunk of morphosyntactic structure into a PW and the way this is formally expressed. It will become apparent from the discussion that, despite the thirty or so years of research in this area, certain aspects of the mapping mechanism are still poorly understood. Second, I will present the methodology and, in particular, the diagnostic criteria employed by researchers in identifying the domain of the PW. This survey will provide the opportunity to investigate and, more importantly, assess the amount of knowledge that has been accumulated over the years regarding the nature of rules that identify this prosodic constituent. Third, I will review the empirical situation and, more specifically, whether cross-linguistic evidence renders the PW universally viable or not. To this end, alternative proposals – which range from the formation of extended versions of the PW to the introduction of smaller or larger reincarnations of it (e.g. the small word, etc.) – will also be reviewed. Such enriched versions of the Prosodic Hierarchy have been proposed to accommodate

[2] The CG is the most debatable prosodic constituent, and is usually not included in the Prosodic Hierarchy. It was originally introduced by Hayes (1989) and later adopted and further established by Nespor and Vogel (1986) (see also CHAPTER 84: CLITICS).

complex constructions involving function elements (e.g. particles, clitics, etc.), compounds, and complex predicates, which yield somewhat "looser" versions of the notion "word" and hence raise challenging questions for the mode in which the mapping is performed.

The remainder of this chapter is organized as follows: §2 sketches out the main advancements that led to the development of a hierarchical model of prosodic constituency, an integral part of which is the PW. §3 examines how prosodic units relate to constituents of morphosyntactic structure and the places where this mapping is performed. §4 sets out the diagnostic criteria for identifying phonological wordhood. §5 focuses on the properties of an extended version of the PW, and §6 concludes this chapter.

2 The birth of the Prosodic Hierarchy: From boundaries to prosodic domains

In the *Sound Pattern of English* (*SPE*; Chomsky and Halle 1968) two concepts of syntactic surface structure are acknowledged: (a) output of the syntactic component, and (b) input to the phonological component. Re-adjustment rules are employed to handle discrepancies between these two types of structure. Their main task is to convert the syntactic string into a form that can later be read off and interpreted by phonology. More specifically, syntactic information is encoded in phonology by rules that insert *boundary* symbols at the edges of syntactic constituents. Such boundaries are considered to be segment-like elements that lack any phonetic manifestation. Two are relevant for the discussion here: (a) the syntactic boundary, #, used to indicate edges of major syntactic categories (N, V, A, etc.), phrasal categories (e.g. NP, VP, etc.) and stress-neutral affixes (Chomsky and Halle 1968: 12, 366), and (b) the morphological boundary, +, which indicates lexically assigned morphological boundaries (Chomsky and Halle 1968: 94). Essentially, the introduction of boundaries initiates an indirect mode of interaction between the components of grammar: phonological rules can either refer to boundaries in their structural description or be blocked by them, but they can never refer directly to syntactic edges.

Much of the post-*SPE* era has been devoted to defining the exact number of boundaries and their relative strength prominence (e.g. McCawley 1968; Selkirk 1972). Gradually, however, the focus of attention shifted from linearly ordered boundary-defined domains to hierarchically organized prosodic ones.[3] Based on the seminal work of Liberman (1975) and Liberman and Prince (1977), Selkirk (1978b, 1980, 1981a) proposes that phonological representations, like syntactic ones, are hierarchical in nature. She provides compelling arguments in favor of a "suprasegmental, hierarchically arranged organization" of the utterance (Selkirk 1981a: 111). Selkirk aptly points out the "problem of nestedness": a phonological rule that applies across a "stronger" boundary can also apply across all "weaker" ones. Analogously, a phonological rule that applies before or after a certain

[3] The increasing debate against boundaries was also stimulated by their clearly diacritic character. Boundaries do not constitute linguistic objects and, as such, lack any formal existence in mental representations (see e.g. Pyle 1972; Rotenberg 1978).

boundary can do so with all stronger ones. For instance, in Sanskrit a rule of voicing affects stops in intervocalic position (Selkirk 1980: 115). Crucially, the rule applies at the domain of the Utterance and affects only consonants at # boundaries (2b), ignoring the ones residing at a + boundary (2a).

(2) a. #marut+i# → maruti
  b. #parivraṭ# #ajam# → parivraḍ ajam

In the boundary theory, the typological predictions in the domain of application of phonological rules described above are a mere stipulation. However, they come for free in a theory that assumes a hierarchically organized set of *prosodic categories*, i.e. sub-units of prosodic structure such as the syllable, the foot, the phonological word, the phonological phrase, the intonational phrase, and the utterance. This hierarchical constellation, the Prosodic Hierarchy, signals the birth of *Prosodic Phonology*. The advantage of the new theory is that the principle in (3) (Selkirk 1984; Hayes 1989) can easily capture the "nested" effect in the application of phonological rules mentioned above.

(3) *Strict Layer Hypothesis (SLH)* (Hayes 1989: 204)

The categories of the Prosodic Hierarchy may be ranked in a sequence C_1, C_2, . . . C_n, such that
a. all segmental material is directly dominated by the category C_n, and
b. for all categories C_i, $i \neq n$, C_i directly dominates all and only constituents of the category C_{i+1}.

Selkirk strengthens the argumentation in favor of prosodic constituency by showing that prosodic domains have independent motivation, besides the interface (Selkirk 1980: 110, 126–129; 1981a: 125; 1984: 8ff.). For instance, the PW in English is a category with an internal organization that yields patterns of relative (i.e. strong *vs.* weak) stress prominence, e.g. (ˌɪrɪ)$_{F_w}$ (ˈspɛktɪv)$_{F_s}$, while at the same time constituting the domain of application of various phonological rules (e.g. non-low vowels are tensed in PW-final position).

After an exploratory period (Nespor and Vogel 1982, 1983; Nespor 1985, 1986; Vogel 1985, 1986), Nespor and Vogel (1986) extended Selkirk's work and further enriched it with novel data from a wide array of languages (e.g. Hungarian, Greek, Turkish, and Italian). The focus of research in the following years was on defining the basic premises of Prosodic Phonology (e.g. the nature of the mapping algorithm, the way it is performed, etc.) and lending further support to the theory with empirical evidence from cross-linguistic research (e.g. Booij 1983, 1985a, 1985b; Booij and Rubach 1984; Hayes 1989; Itô and Mester 2003).

In the Prosodic Phonology literature, however, the main motivation for prosodic constituency is non-isomorphism, most commonly expressed as a mismatch between phonological and morphosyntactic boundaries (e.g. Selkirk 1981a; Nespor and Vogel 1982). Nespor and Vogel (1986) argue that in Hungarian, for instance, vowel harmony takes place within a stem and a string of suffixes – (4a) and (4b) – but fails to apply in strings that contain sequences of stems (4c) or a prefix plus a stem (4d) (see CHAPTER 123: HUNGARIAN VOWEL HARMONY). The reason for the disparity in the application of the process relies on the nature of the phonological domain formed by the elements involved. Vowel harmony applies within the domain of the PW, and a stem + suffix string is mapped into one PW. Significantly, compounds and prefixed constructions form two separate PWs, as shown in (4c) and (4d), despite the fact that they constitute single grammatical words.

(4) *Hungarian vowel harmony* (Nespor and Vogel 1986)

a.	*ölelés*	[ölelés]$_{PW}$	'embracement'
	hajó	[hajó]$_{PW}$	'ship'
b.	*ölelés-nek*	[ölelésnek]$_{PW}$	'embracement-DAT SG'
	hajó-nak	[hajónak]$_{PW}$	'ship-DAT SG'
c.	*Buda-Pest*	[Buda]$_{PW}$ [Pest]$_{PW}$	'Budapest'
	könyv-tar book-collection	[könyv]$_{PW}$[tar]$_{PW}$	'library'
d.	*be-utazni* in-commute	[be]$_{PW}$ [utazni]$_{PW}$	'to commute in'
	oda-menni there-go	[oda]$_{PW}$ [menni]$_{PW}$	'to go there'

The reference to prosodic constituents such as the PW allows us to describe the rule of vowel harmony in a unified way. The issue of (non-)isomorphism is central in defining and exploring the nature of phonological wordhood and is addressed in detail in subsequent sections of this chapter.

To summarize, each one of the prosodic categories in the Prosodic Hierarchy is governed by its own principles of internal constituency, and each one forms a domain for the application of phonological rules. Moreover, certain prosodic domains are introduced in order to capture the non-isomorphic character of the interface. Boundaries, on the other hand, are entities that "assist" the interface but have a strong diacritic flavor. The ensuing sections explore the mode in which a particular prosodic constituent, namely the PW, relates to constituents of the morphosyntactic structure and the place this mapping occurs.

3 The Phonological Word at the interface

The classical Prosodic Phonology theory (Selkirk 1981a, 1981b, 1984, 1986, 1995; Nespor and Vogel 1982, 1986) assumes the organization of grammar in (5). Thus, the rules that undertake the mapping of morphosyntactic structure into a PW are part of the *post-lexical* phonology. Selkirk (1984: 82) dubs this a *syntax-first* model.

(5) syntactic surface structure → mapping rules → prosodic representations → phonological rules → phonetic representations

This view, however, has been challenged by several researchers (e.g. Booij 1983, 1988; van der Hulst 1984; Inkelas 1989; Booij and Lieber 1993). Nowadays it is widely accepted that constituents up to the level of the PW are built lexically, and that phonological and morphological structure is constructed at the same time, from bottom up. This issue is addressed in detail in §3.3.

3.1 *Mapping rules under the Strict Layer Hypothesis*

Nespor and Vogel (1986) recognize that there are several options available for the definition of PW. They attribute the attested variation to an assortment of morphological notions that mapping rules can be sensitive to (e.g. stems, roots, sequences of stems, and suffixes, etc.). There are languages, like Greek for instance, in which a lexical word, i.e. the terminal element of the syntactic tree, whether a compound or simple word, is mapped into a PW. On the other hand, there are languages like Hungarian, in which mapping rules are sensitive to smaller elements such as stems, prefixes, or sequences of stems and suffixes. Finally, there are languages such as Dutch, in which certain elements (e.g. suffixes or prefixes) are endowed with a diacritic that grants them independent PW status regardless of the general dictates of the mapping mechanism. For instance, in Italian vowel-final prefixes are assigned independent PW status due to language-specific syllable well-formedness conditions, a restriction that the mapping rule must read and comply with. In Nespor and Vogel's (1986) model of mapping there is yet another source of phonological wordhood, namely the SLH in (3). The principle of proper nesting requires all "unparsed" elements to prosodify into a PW in the absence of a neighboring host. The definition of PW in (6) encapsulates the aforementioned mapping possibilities.

(6) *PW domain* (Nespor and Vogel 1986: 141)

A. The domain of PW is Q (= any terminal element of the syntactic tree).

or

B. I. The domain of PW consists of
   a. a stem;
   b. any element identified by specific phonological and /or morphological criteria;
   c. any element marked with the diacritic [+W].

II. Any unattached elements within Q form part of the adjacent PW closest to the stem; if no such PW exists, they form a PW on their own.

The above set of statements on the application of mapping rules makes some predictions. First, PWs cannot be larger than a terminal element in the syntactic tree. Second, no single stem can be mapped into more than one PW, and, third, affixes will always be part of the PW of their base unless they bear a diacritic. Interestingly, the definition in (6) encompasses diacritic information as well as special restrictions imposed by the phonological and/or morphological components of the grammar. Such a mapping algorithm, therefore, is far too powerful to be insightfully implemented in an interface theory which aspires at advancing the understanding of the factors involved in PW formation.

Selkirk (1986), based on Chen (1987), proposes an *end-based* mapping theory, which operates on the edges of X-bar constituents. The basic idea is that a prosodic constituent is demarcated by the right or the left edge of selected syntactic constituents (i.e. X^0, X', X'') (Selkirk and Shen 1990: 319):

(7) *The syntax–phonology mapping*

For each category C^n of the prosodic structure of a language there is a two-part parameter of the form

C^n: {*Right/Left*; X^m}, where X^m is a category type in the X-bar hierarchy.

It is a parametric choice of a language as to which edge of the X^0 (= word), for instance, will serve as the beginning or the end point of the PW domain. An advantage of the model is that the mapping algorithm is cross-categorical: the end rules can apply at some level of the X-bar hierarchy in order to form the appropriate prosodic domains. For instance, they apply at the X^0 to derive PWs and at the X′ or X″ to derive PPhs.

A comparison between the two algorithms reveals a significant difference in their descriptive and predictive power. The prosodic patterns established between a PW and a *morphological word* (MW)[4] by Nespor and Vogel's (1986) algorithm are given in (8). Recall that this algorithm grants independent PW status to sub-minimal elements such as function words (fnc), prefixes (prf), etc., either as a result of a diacritic or due to the SLH. Crucially, a PW can never be larger than an MW.

(8) a. PW = MW or *x* where *x* = fnc/prf
b. PW < MW
c. *PW > MW

Selkirk's (1986) end-based mapping algorithm, on the other hand, assigns a PW status to MWs but, importantly, permits a PW larger than an MW as a prosodic output. This domain can be derived when function elements are trapped in between two MWs, as shown in (9). Selkirk (1986, 1995) explicitly states that only lexical categories – not functional ones – and their projections are visible to the mapping rules. This entails, therefore, that, depending on the end rule parameter setting, the function words will prosodify together with the preceding or the following MW (= X^0). In the following abstract example, the PW will extend from the left (9b) or the right (9c) edge of one lexical item to the left or the right edge, respectively, of the next, incorporating in the process any intervening function words.

(9) a. X^0 fnc fnc X^0
b. $[X^0$ fnc fnc$]_{PW}$ $[X^0]_{PW}$ end rule: left
c. $[X^0]_{PW}$ [fnc fnc $X^0]_{PW}$ end rule: right

Importantly, the original formulation of the end-based mapping rule does not allow elements that are smaller than the MW to constitute independent PWs. As a consequence, clitics, prefixes, and suffixes are deprived of this prosodic possibility. Given the SLH, only one option is available: incorporation of the sub-minimal element into the PW of a neighboring MW. The parsing options for PWs derived by the end-based algorithm are summarized in (10):

(10) a. PW = MW
b. *PW < MW
c. PW > MW

[4] In linguistics, the notion "morphological or grammatical word" is not uncontroversial. Dixon and Aikhenvald (2002: 18ff.) provide different types of criteria to define it and also distinguish it from PW. The most important three are that the elements a grammatical word consists of must always occur together, in a fixed order, and have a conventionalized coherence and meaning.

Interestingly, the two models diverge in the cases of non-isomorphism: Nespor and Vogel's algorithm allows a PW to be smaller but not larger than an MW, whereas the end-based mapping sanctions the opposite. Furthermore, each theory makes different predictions with respect to the prosodization of sub-minimal elements. Due to the SLH, the end-based algorithm will force such elements to incorporate into an adjacent PW, whereas Nespor and Vogel's algorithm will allow them to form an independent PW. The end-based algorithm is silent regarding the prosodization of non-terminal syntactic elements such as stems, prefixes, and suffixes.

The empirical facts provide only partial support for each model. Let us start with the pattern PW < MW, predicted only by Nespor and Vogel's algorithm. In northern Italian, prefixes and the stems of compounds, but not suffixes, form separate PW-domains for the rule of intervocalic s-voicing (Nespor and Vogel 1986: 125). As evidenced by the examples in (11), this rule applies in monomorphemic and inflected words, (11a) and (11b), which constitute one PW. It is blocked, however, between a prefix and stem (11c) or between the stems of a compound (11d), suggesting that such constructions are parsed into two PWs.

(11) *Northern Italian: Intervocalic* s-*voicing*

a.	*a/s/ola*	*a*[z]*ola*	$[\text{azola}]_{PW}$	'button hole'
b.	*ca/s/-e*	*ca*[z]*e*	$[\text{caze}]_{PW}$	'house-PL'
c.	*a-/s/ociale*	*a*[s]*ociale*	$[\text{a}]_{PW}$ $[\text{sociale}]_{PW}$	'asocial'
d.	*tocca-/s/ana*	*tocca*[s]*ana*	$[\text{tocca}]_{PW}$ $[\text{sana}]_{PW}$	'cure all'

During the exploratory period of Prosodic Phonology, a growing body of cross-linguistic evidence revealed that affixes may form independent PWs (e.g. Booij and Rubach 1984 for Polish and English; Hannahs 1991, 1995a, 1995b for French; amongst many others). Cross-linguistic evidence, therefore, gives an empirical advantage to the Nespor and Vogel mapping algorithm compared to the end-based one. However, several researchers have shown that the latter can easily handle these facts if end rules are appropriately modified so that they can read off edges of elements smaller than the word, such as stems (Kang 1992), or even edges of functional categories, e.g. agreement (Rice 1993).

Furthermore, the end-based algorithm seems to have the advantage on the empirical side in cases where the PW is larger than the MW. Booij (1988) draws attention to clitic constructions from Latin and Dutch where clitics – even though they are independent grammatical words – are phonologically dependent on an adjacent word. This mapping is predicted by the end-based algorithm, but not by the Nespor and Vogel one. It is important to emphasize here that Nespor and Vogel (1986) posit a different prosodic constituent for the prosodic organization of such clitic constructions, namely the Clitic Group (CG). Within this larger constituent, however, the Nespor and Vogel mapping algorithm, under the dictates of SLH, is forced to elevate each independent element, i.e. clitic, into a PW. Below I exemplify from Greek the results of each model of mapping.

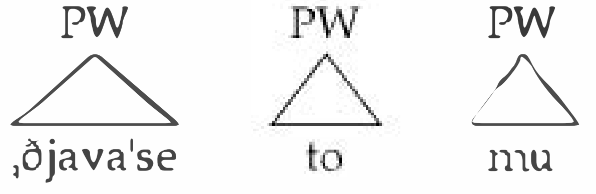
In Greek, weak object pronouns precede the non-imperative verb form (12a) but follow the imperative (12b) (Revithiadou and Spyropoulos 2008, and references cited therein).

(12) *Greek object clitics*

a. o 'petr-os to 'ðjava-se
the Peter-NOM.SG CLT.3N.SG.ACC read-PST.3SG
'Peter read it.'

b. ˌðjava-'se to (< 'ðjava-se)
read-2SG.IMP CLT.3N.SG.ACC
'Read it!'

Under a Nespor and Vogel-type mapping, each clitic will be granted PW status, as shown in (13). The problem with this structure, however, is that the PW of the clitic exhibits a different behavior from that of the PW of the host element. A clitic may "trigger" stress on the final syllable of the preceding PW (14a) or carry stress itself (14b). More importantly, examples like (14a) pose a serious threat to the SLH. The addition of the clitic causes the form to be further footed so that the three-syllable restriction imposed by the language can be salvaged: (ˌðjava)('se to) mu. The problem now is that the newly formed foot cuts across two PW boundaries: [(ˌðjava)('se]$_{PW}$ [to)]$_{PW}$ [mu]$_{PW}$, in total disrespect of proper containment.[5]

(13)


(14) a. ˌðjava-'se to mu
read-2SG.IMP CLT.3N.SG.ACC CLT.1SG.GEN
'Read it to me!'

b. ˌpar-e 'mu to
take-2SG.IMP CLT.1SG.GEN CLT.3N.SG.ACC
'Take it for me!'

Furthermore, there is compelling evidence that postverbal clitics are incorporated into the PW of the verb (see footnote 5 and Revithiadou and Spyropoulos 2008 for more evidence), as correctly predicted by the end-based algorithm PW: {*Left*; X^0}. Thus, the comparison so far gives a descriptive advantage to the end-based model. A more careful examination of the facts, however, reveals that this algorithm is not unproblematic either. Given the left-end orientation of the rule, pre-verbal clitics are expected to encliticize to the preceding PW, but they do not, as shown in (15). If the clitics incorporated into the PW of the NP /o 'θoðoros/ 'the Theodore-NOM.SG', the window restriction would have forced the development of a new stress on the last syllable of the noun, yielding [o θoðo'ros tu to]$_{PW}$ 'the Theodore-NOM.SG CLT.3SG.GEN CLT.3N.SG.ACC.' Clearly, this expectation is not confirmed by the data.

[5] Nespor and Vogel (1986: 154–155) escape this problem by employing a (grid-based) *Stress Readjustment* rule. There is independent evidence, however, that clitics in Greek are footed. For instance, imperatives opt to incorporate the enclitic by deleting their final vowel, e.g. /ɣrapse ton/ (*'ɣrapston*)$_F$ 'write-IMP.2SG CLT.3M.SG.ACC'. Alternatively, the clitic may be augmented so that it can form its own foot, e.g. (*ˌɣrapse*)$_F$ (*'tone*)$_F$. Similar augmentation phenomena are independently enforced by foot well-formedness conditions, e.g. /ro'ta-n/ (with inherent accent) 'ask-3PL' *ro('tane)* ~ *ro('tan)*.

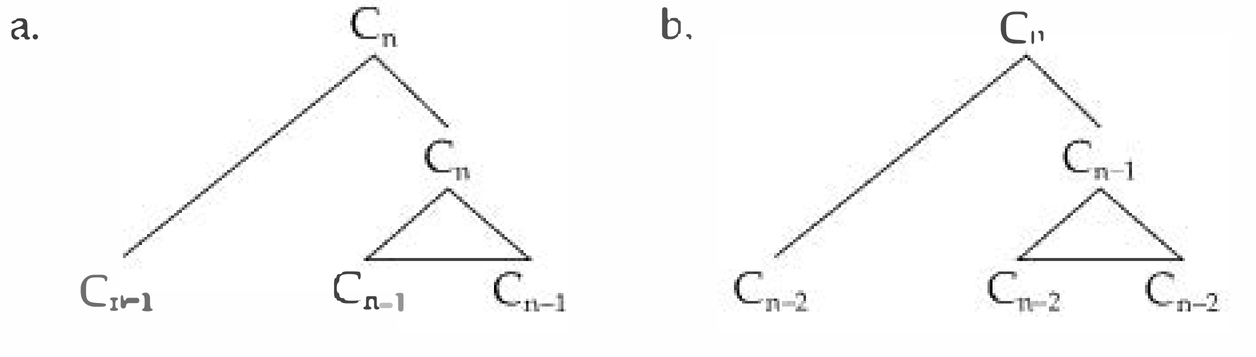
(15) o 'θoðor-os tu to ðja'vaz-i
the Theodore-NOM.SG CLT.3SG.GEN CLT.3N.SG.ACC read-3SG
'Theodore reads it to him.'

Similar cases have been found in a variety of languages, underlining, among other things, the asymmetric behavior often witnessed in the prosodic organization of constructions with function elements (e.g. particles, clitics, etc.), compounds, and other complex expressions (e.g. Inkelas 1989; Booij 1996; Leben and Ahoua 1997; Peperkamp 1997; Vigário 1999, 2003).[6] These findings therefore challenged the descriptive and explanatory efficiency of both mapping algorithms, and opened up new directions for research in this area.

3.2 *The Weak Layer Hypothesis: Mapping as constraint interaction*

A number of researchers have sought a solution to the problems mentioned above in the architecture of the Prosodic Hierarchy. In particular, they questioned the necessity of the SLH as a well-formedness principle of the arboreal prosodic structure. The proposed modifications involve the introduction of (a) *recursive* structures, in which a prosodic constituent of a certain type dominates another constituent of the same type (16a), and (b) *non-exhaustively parsed* structures, in which a constituent of one type is permitted to skip intermediate levels and dominate a constituent of more than one level lower in the Prosodic Hierarchy (16b). The relaxed version of the SLH is referred to as the *Weak Layer Hypothesis* (e.g. Booij 1988, 1995, 1996, 1999; Itô and Mester 2003).

(16) *Weak layering: Recursion and non-exhaustivity*



Under the influence of Optimality Theory (Prince and Smolensky 1993), Selkirk (1995) goes one step further and proposes that the SLH should be decomposed into its primitive components, which take the form of the *prosodic domination* constraints in (17).

(17) *Constraints on prosodic domination* (Selkirk 1995: 443)
(where C^n = some prosodic category)

a. LAYEREDNESS: No C^i dominates a C^j, $j > i$.
b. HEADEDNESS: Any C^i must dominate a C^{i-1}.
c. EXHAUSTIVITY: No C^i immediately dominates C^j, $j < i-1$.
d. NON-RECURSIVITY: No C^i dominates C^j, $j = i$.

6 The asymmetry in the prosodization of weak elements constitutes one of the main arguments against the CG as a prosodic constituent.

Layeredness and Headedness are argued to be universally inviolable and therefore undominated in the constraint ranking of all languages. The structures in (16) result from the relative ranking of Exhaustivity and Non-Recursivity to the other constraints of the system. A significant role in the construction of prosodic constituents has been played by the *alignment* family of constraints, which basically undertakes the mapping of morphosyntactic constituents to prosodic structure (McCarthy and Prince 1993a, 1993b):

(18) *Alignment constraints*

a. MW-Constraint (WCon): Align(MW, L/R; PW, L/R)
b. PW-Constraint (PCon): Align(PW, L/R; MW, L/R)

These interface constraints are translations of Selkirk's (1986) parameterized end-based theory of mapping into Optimality Theory (OT) constraint-based terminology. They have a uniform general scheme, which can easily account for cross-categorical mappings, since units from different levels of the Prosodic Hierarchy can coincide with morphosyntactic units. For instance, the constraint in (18a) requires the left or right edge of every MW to coincide with the left or right edge of some PW.

Differences in the ranking of the relevant constraints and/or the morphosyntactic structure of an input string are taken to be responsible for the attested variation in the prosodization of weak elements (e.g. function words, prefixes, etc.) cross-linguistically. The interaction of the above constraints yields the following prosodic patterns for the abstract string /*x x* V/ (where *x* is a weak element):

(19)

	structures	*typology*	*rankings*
a.	$[[\acute{x}\ x]_{PW}\ [V]_{PW}]_{PPh}$	PW	Exh, NonRec, MCon >> PCon
b.	$[x\,x\ [V]_{PW}]_{PPh}$	free	MCon, PWCon >> NonRec >> Exh
c.	$[x\ x\ [V]_{PW}]_{PW}$	recursive	Exh, MCon >> NonR, PCon
d.	$[x\,x\ V]_{PW}$	internal	Exh, NonRec >> MCon, PWCon

It remains an open question whether all predicted typologies receive empirical support or whether they lead to vast overgeneration of PW patterns. There is, however, ample cross-linguistic evidence in support of the typology of function words in (19). In a cross-dialectal study of Italian clitics, for example, Peperkamp (1997) shows that the patterns in (19b) and (19d) correspond to specific Italian dialects. Similarly, Revithiadou (2008) provides evidence from a cross-dialectal survey of object clitics in Greek that all four prosodic patterns are empirically attested. Moreover, she demonstrates on the basis of diachronic evidence that there is a transition from "looser" types of constructions – e.g. patterns (19a) and (19b) – to nested ones – pattern (19c) – and from there to total integration of the clitic to its host, pattern (19d) (see also chapter 84: clitics).

3.3 *Split between two worlds*

After discussing various approaches as to how a mapping rule assigns a PW make-up to a portion of the morphosyntactic string, I now turn to addressing where and when the mapping takes place. Such questions are of course meaningful only within a theory that assumes a division of labor between the components of grammar and a procedural view of the interface.[7]

Selkirk takes a clear stance on this issue and argues that both word structure and sentence structure are sensitive to the same mechanism, and must therefore be treated alike, i.e. after syntax at the post-lexical component (Selkirk 1984: 415). Any additional tools would simply lead to an unnecessary proliferation of the machinery phonology has at its disposal. Nespor and Vogel (1986), on the other hand, take a more moderate approach. They draw a distinction between two sets of rules: (a) those that refer to phonological domains only and (b) those that make *direct reference* to morphological information. By doing so, they implicitly adopt a compartmentalized view of phonology. First, there is a section of phonology that may access directly "morphological structure and/or specific morphological elements" (Nespor and Vogel 1986: 18). Its rules are handled by a different mechanism, possibly *Lexical Phonology* (Kiparsky 1982a, 1982b; Mohanan 1982; CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME). Second, there is another section that accesses the interface indirectly, via the constituents of the Prosodic Hierarchy, and operates strictly on Prosodic Phonology rules proper. The question as to how these blocks of rules are ordered is left open to further research (see also CHAPTER 103: PHONOLOGICAL SENSITIVITY TO MORPHOLOGICAL STRUCTURE).

Inkelas (1989) attempts to salvage the indirect nature of the interface alluded to by Nespor and Vogel (1986) with the introduction of a radical move: she eliminates the σ and the F from the Prosodic Hierarchy and introduces prosodic constituency below the PW in the Lexicon. The ranks below the PW in (1), which roughly correspond to levels in Lexical Phonology, can now accommodate strings of structure into prosodic units that are smaller than the PW. They basically provide the appropriate domain within which phonological rules can apply without having to directly refer to morphological constituents.

Booij and Rubach (1984) and Booij (1988) concur that a lexically constructed PW is necessitated on empirical grounds, e.g. the prefix *a-* in Italian; see example (11). They further argue that it can offer solutions to bracketing paradoxes, special rules of allomorphy, and so on. For instance, the word $[[un[grammatical]_A]_A ity]_N$ poses a problem, because the prefix *un-* is stress-neutral (i.e. level 2 in Lexical Phonology terminology), and hence should be added *after* the stress-shifting suffix *-ity* (i.e. level 1). This implies, however, that the prefix *un-* should attach to *grammaticality* and not to *grammatical*, which is also problematic, because *un-* attaches only to adjectival bases. At the phonological level, the suffix *-ity* shifts the stress of the base *ungrammatical*, which nevertheless contains a stress-neutral prefix. This is problematic too, because only stress-neutral suffixes can be added after that prefix (i.e. a level 1 affix may not follow a level 2 affix). This bracketing paradox receives a straightforward explanation once a prefix is assigned an independent PW status: $[un]_{PW}$ $[grammaticality]_{PW}$. Each PW constitutes an independent domain of stress assignment. As a result, the shifting property of the suffix can never interfere with the prefix, simply because the latter belongs to a different PW. Elements that impose such prosodic restrictions are often called *non-cohering*.

[7] In non-serial theories of phonology such as OT, where mapping is the job of interface constraints that operate simultaneously on both modules, such questions are almost redundant, unless one embraces the distinction between a lexical and a post-lexical component in phonology.

If this property is not derivable from the size and the segmental shape of the morpheme in question, it must be specified in the subcategorization frame of the relevant lexical item (Booij and Lieber 1993). Clitics may have similar prosodic selection requirements. For instance, *ie* 'he' in Dutch subcategorizes for a right PW boundary: $]_{PW}$ __ (Booij and Lieber 1993: 39).

However, there is a major drawback in assigning PW status to morphological elements in the lexicon: it reintroduces via the back door the problem of the PW boundary as a diacritic, thus posing a serious threat to the very nature of the Prosodic Hierarchy. Another major consequence of moving the construction of PWs into the lexicon is that the mapping now applies twice: once in the lexicon and again after syntax, i.e. post-lexically. This suggests that there are two types of PWs, lexical and post-lexical, and, consequently, the interface between phonology and morphology is of a different nature than that between phonology and syntax. If we admit this, we need to jettison the idea of a unified phonology that Prosodic Phonology – especially through the work of Selkirk – originally pursued. Yet another problem with implementing prosodic rules in the lexicon is that it blurs the division between Prosodic Phonology and Lexical Phonology (see CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME). The former theory is designed to deal with the interface of phonology and morphosyntax post-lexically; the latter relies on a procedural mode of interaction of phonology and morphology in order to account for "idiosyncratic," morphology-dependent phonological behavior. If the two models meet in the lexicon, then one may wonder which type of phenomenon each model targets and, moreover, on which grounds the division of labor is decided. Do their rules apply simultaneously or in an ordered fashion?

Even though the model of Lexical Phonology has gradually lost headway, the distinction between lexical and post-lexical PW has endured over time (Booij and Lieber 1993; see also Booij 1999 for Dutch; Nespor 1990, Vogel 1991, Peperkamp 1997 for Italian; amongst others). The lasting nature of this division emphasizes all the more the truly interface nature of PW and further establishes it as the prosodic constituent that intersects the interface of phonology with morphology and syntax. However, more work needs to be done in order to acquire a better understanding of the factors that dictate the formation of PWs, and the workspace where this is done.

4 Diagnostics for the Phonological Word as a prosodic constituent

In the previous sections, we examined ways and places within which PWs are built. Here, the focus will be on what type of phonological evidence has been put forward to substantiate the PW as an integral part of the Prosodic Hierarchy. Nespor and Vogel (1986: 58ff.) identify certain diagnostics for establishing the concept of *constituent* in phonology. Phonological rules must refer to a string of elements in their formulation and have this string as their domain of application. The same piece of structure may also serve as the domain of phonotactic restrictions and stress prominence relations. Vogel (2009) aptly remarks that in establishing a prosodic constituent a number of phenomena must cluster together in using a particular string of elements as their domain. In the almost thirty years of research on Prosodic Phonology, various diagnostic criteria have been put forward to identify the domain of the PW. In the ensuing paragraphs, we will closely examine the most important ones, at the same time drawing attention to the existence of conflicting evidence and its repercussions for the status of PW as part of Universal Grammar.

4.1 *Segmental rules*

In previous sections, we have seen that it has been proposed that both vowel harmony in Hungarian and intervocalic *s*-voicing in Italian apply within the domain of the PW. The Prosodic Phonology literature is replete with examples of segmental rules used as diagnostics for the definition of the PW domain in a variety of languages (see e.g. Hannahs 1991, 1995a, 1995b; Kang 1992; Peperkamp 1997; Kleinhenz 1998; Hall and Kleinhenz 1999; Vigário 1999, 2003). To illustrate with an example from French, Hannahs (1995a, 1995b) argues that glide formation and vowel nasalization can be safely used as diagnostics of the PW domain. For instance, underlying high vowels such as /i y/ semi-vocalize intervocalically to [j] and [ɥ], respectively. Crucially, this rule applies only in stem + suffix strings (20a) but never when the vowel in question occurs at the end of a prefix (20b) or the first element of a compound (20c) (Hannahs 1995b: 1131). The blocking of the rule in the latter environments is taken as evidence that the stem plus suffix string constitutes a different PW from the remainder of the word. This is a typical example of non-isomorphism between prosodic and morphological structure.

(20)	a.	*colonie*	[kɔlɔni]	'colony'
		colonial	[kɔlɔnjal]	'colonial'
	b.	*anti-alcoolique*	[ãtialkɔlik]	'anti-alcohol'
			*[ãtjalkɔlik]	
	c.	*tissue-éponge*	[tisyepɔ̃ʒ]	'terry-cloth'
			*[tisɥepɔ̃ʒ]	

Similarly, Vigário (2003) employs a great variety of both lexical and post-lexical rules to substantiate the existence of the PW in European Portuguese. One such rule is vowel reduction, which affects all stressless vowels of a word except for the word-initial one, e.g. *promo'ver* → *promu'ver* '(to) promote'. A more careful examination of the data nevertheless reveals that the vowel is protected even when it is not in absolute word-initial position (21b). For Vigário (1999: 272–273), this constitutes evidence for the presence of a PW boundary at the left of the base *ocu'par*. The exact prosodic structure of such prefixed formations will be discussed in §5.

(21)	a.	ocu'par	'(to) occupy'
	b.	desocu'par	'(to) not occupy'

4.2 *Stress and tone*

Stress is prototypically considered to be an infallible diagnostic for the PW domain: a PW must bear only one primary word stress (see CHAPTER 41: THE REPRESENTATION

OF WORD STRESS). In Greek, for example, the phonological wordhood of lexical words and stem–word compounds is signaled by the presence of one primary stress:

(22) a. /anθrop-os/ ['anθropos] 'man'
man-NOM.SG
b. /pali-o-maɣaz-o/ [paljo'maɣazo] 'lousy shop'
bad-LINKV-shop-NOM.SG

Dixon (2002) reports that in Jarawara, a Madi dialect of the Arawá family of southern Amazonian, primary stress is on the penultimate syllable, and rhythmic stress occurs on every other syllable to the left of the main-stressed foot (23a). Curiously, compounds and reduplicative formations display two stress peaks, which, crucially, are not on the second and the fourth mora from the end of the complex word but rather on the penultimate syllable of each of their constituents; (23b) and (23c). The attested stress patterns, therefore, suggest that the relevant formations form two PWs, e.g. ['bani]$_{PW}$ [ka'sako]$_{PW}$ (Dixon 2002: 128).

(23) a. to-'wa-ka-'tima-'maro
away-APPLIC-in.motion-upstream-FPEF 'took upstream'
b. 'bani-ka'sako / *ba'nika'sako 'wild dog species'
c. 'kete-ke'tebe / *ke'teke'tebe 'run a lot'
< ke'tebe 'run, follow'

Tonal information may also serve as a criterion for delimiting the PW boundaries (see CHAPTER 42: PITCH ACCENT SYSTEMS). Leben and Ahoua (1997) provide an instructive example from Baule, a Bia language of the central Tano group. The language has both a High and a Low tone. Interestingly, a sequence of High tones shows an upstepping pattern, which involves a gradual rise in pitch from a level phonetically close to Low to a Super-High level. The domain of the upsweep, as this rule is commonly referred to, is the PW. The examples in (24) demonstrate that the rule operates in monomorphemic words (24a) and noun–noun compounds (24b), but it is blocked in possessor–possessed (24c) and subject–predicate phrases (24d) (Leben and Ahoua 1997: 117–118). The difference between these two sets of formations is attributed to a difference in their respective prosodic constituencies. The former constitute a single PW, e.g. [bólí nɔ́nnɔ́n]$_{PW}$ (24b), whereas the latter are organized into two separate PWs, e.g. [bóli]$_{PW}$ [mángún]$_{PW}$ (24c). The proposed prosodic structures are enhanced with additional evidence from segmental processes (Leben and Ahoua 1997: 122ff.).

(24) a. Á k í s í 'Akisi'
[– – ¯]
b. b ó l í n ɔ́ n n ɔ́ n 'goat milk'
[_ – ¯ ¯]
c. b ó l í m á n g ú n 'goat's friend'
[_ ¯][_ ¯]
d. Á y á b ó l i 'Aya is a goat'
[– ¯][_ ¯]

4.3 Phonotactics and syllabification

The PW has been considered an important domain for phonotactic constraints, together with the syllable and the foot. In Dutch, for instance, the schwa as a featureless vowel cannot be preceded and followed by the same consonant; /a:lǝl/, for example, is an impossible morpheme. Across PW boundaries, however, as in compounds, such sequences are tolerated, e.g. *formu*/lǝ-l/*ijst* 'formula list' suggesting that the relevant Obligatory Contour Principle restriction is confined by the PW (Booij 1999: 56–57).

The boundaries of PWs can also be subject to certain restrictions. In German, a lax vowel constraint (i.e. *[−tense, −low, −long]) limits the distributional freedom of vowels that occur at the right edge of the PW (Hall 1999). In European Portuguese (Vigário 2003), a PW cannot begin with a flap [ɾ]. Furthermore, Sanskrit shrinks its consonant clusters at the right edge of the PW (Selkirk 1980), sharply contrasting in this respect with Dutch, which tolerates the occurrence of extra consonants and heavier rhymes in the same position (Booij 1995, 1999).

In several, but not all, languages syllabification is taken to be a reliable diagnostic for establishing the PW domain (e.g. Kang 1992 for Korean; Booij 1995, 1999 for Dutch; Hall 1999 for German; Peperkamp 1997 for Italian; Raffelsiefen 1999 for English). Kang (1992) argues that in Korean, aspirated consonants, which normally neutralize to stops in coda position, syllabify together with derivational suffixes (25a). This option, however, is not available when the relevant segment is in between the two stems of a compound (25b), suggesting that resyllabification is blocked by an intervening PW-boundary.

(25)	a.	$[[\text{kip}^h]_V\,\text{i}]_N$	$[\text{ki.p}^h\text{i}]$	*[kip.i]	$[\text{kip}^h\text{i}]_{PW}$
		deep-NML			'depth'
	b.	$[[[\text{ap}^h]_N\,[\text{ap}^h]_N\,\text{i}]_{Adv}$	$[\text{ap.a.p}^h\text{i}]$	$^*[\text{a.p}^h\text{a.p}^h\text{i}]$	$[\text{ap}]_{PW}\,[\text{ap}^h\text{i}]_{PW}$
		front front ADVS			'to each person'

4.4 Other diagnostics

Given that PWs dominate feet and feet are usually branching constituents, PWs should be minimally either bimoraic or disyllabic depending on whether they are composed of moraic or syllabic feet. Thus, foot binarity together with HEADEDNESS (17b) derive the notion *minimal word* (McCarthy and Prince 1986, 1993a, 1993b, 1998). Prosodic minimality is also enforced in morphological processes such as nickname truncations and various types of clippings (see McCarthy and Prince 1998 and references cited therein). In Greek, for instance, name truncations are minimal words (Topintzi 2003):

(26) a. *Ni'kolaos* 'nikos
b. *Poli'kseni* 'poli, 'kseni
c. *Aspa'sia* 'aspa

An often cited example in the literature for the identification of the PW is reduction under coordination (e.g. Booij 1985b for Dutch; Kleinhenz 1998 for German). For instance, Booij (1985b) demonstrates, on the basis of the examples in (27) that parts of complex words – i.e. compounds (27a) and complex formations with the so-called *non-cohering* affixes (27b) – can be cut off in coordinate constructions:

(27) a. natuur~~kunde~~ en scheikunde
'nature knowledge and analysis knowledge'
b. zicht~~baar~~ en tastbaar
'visible and tangible'

The condition is that this rule applies to trim off only parts of grammatical words which constitute separate PWs. That is, it can never eliminate an affix that is included in the PW of the stem it combines with: **rod~~ig~~ of groenig* 'reddish or greenish'. All that the rule "sees" is phonological structure, i.e. PW boundaries, and not the internal morphonological structure of words.

Language games constitute a resourceful supplier of evidence for the PW. Henderson (2002) reports on a play language called "Rabbit Talk" in Arrernte, a central Australian language. In Rabbit Talk, the first syllable of the word is removed from its original position, and transposed to the end of the PW (28a). Monosyllabic words skip the transposition rule, and the syllable /ej/ is prefixed instead (28b). For the purposes of this discussion, I follow Henderson in assuming that the syllable structure is VC(C). Rabbit Talk suggests that disyllabic case clitics constitute a separate PW. In (28c), the two elements of the construction count as separate domains: the first element is treated as a monosyllabic word and hence receives the /ej/ prefix. The clitic /-akerte/, on the other hand, constitutes a PW by itself and, as such, it is subjected to the transposition rule.

(28)		*ordinary speech*	*Rabbit Talk*
a.	ampangk+eme moan+PRES	/amp.angk.em/	/angk.em.amp/
b.	artwe man	/artw/	/ej.artw/
c.	irlpe-akerte ear-COMIT		/ej.irlp.ert.ak/

4.5 *Conflicting evidence and methodological issues in defining the Phonological Word domain*

Within the Prosodic Phonology framework, the methodology applied in establishing prosodic constituents in general and the PW in particular involves four basic steps: first, discovering the domain within which a given phonological process applies or is blocked; second, establishing that more rules have the same domain as their locus of application; third, matching the string of elements with a particular unit of the Prosodic Hierarchy; and fourth, giving the particulars of the mapping mechanism that determines which chunk of morphosyntactic structure is organized into that particular unit. A good theory should have a mapping mechanism with a certain degree of descriptive power. In practice, however, the weight of investigation falls primarily on substantiating a specific prosodic constituent on the basis of the phonological rules alone, with much less attention paid to the specifics of the morphosyntax. This proves to be quite problematic when the phonological component sends conflicting signals.

Cetnarowska (2000) reports that the distribution of secondary (rhythmic) stress in proclitic + host strings in Polish presumes the existence of a foot that straddles a PW boundary which is established on the basis of segmental evidence (Rubach and Booij 1990). In this case, the stress facts of the language conflict head on with the evidence provided by segmental rules. Likewise, Raffelsiefen (1999) argues that assimilation rules in English do not constitute a reliable criterion for establishing PW structure in English. This lack of agreement among different types of diagnostic criteria has led some researchers to propose units smaller or even larger than the PW in order to accommodate the "problematic" or "misbehaving" data. For instance, Rice (1993) acknowledges the existence of a *small word* in Slave, which is subject to certain rhymal constraints, as opposed to the PW, which is the domain of foot-based processes and various segmental rules.

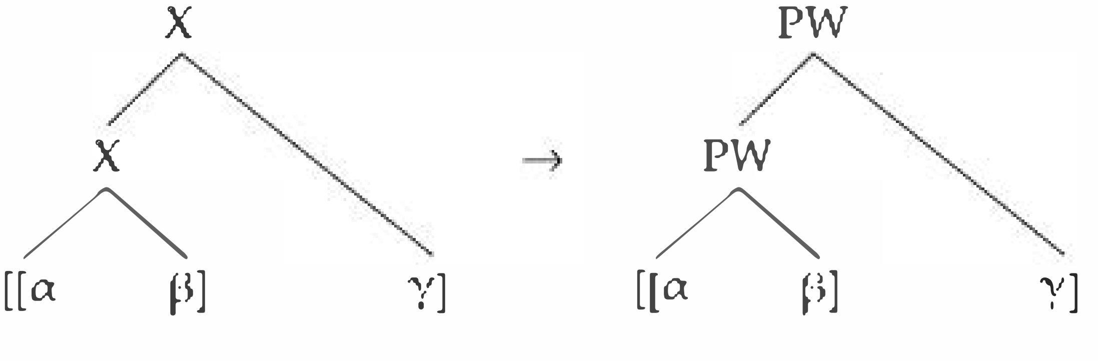
What is clear from the discussion so far is that the proponents of Prosodic Phonology have been confronted with challenging issues, and have opted for different solutions to deal with them. The first solution is to acknowledge that different types of criteria can serve as diagnostics of phonological wordhood in different languages and that each language decides on how to prioritize these criteria or even discard some of them on the basis of some sort of "scale" of relative importance. The question that naturally arises in this case is the way and the context in which this decision is made. The second solution resorts to the proliferation of prosodic domains by inserting pieces of structure with analogous prosodic behavior to distinct slots below the PW. The problem in this case is whether there is an upper bound to the proliferation of domains and, more importantly, whether these domains are universal.

There is a third approach to (partially) tackling the problem described above. Recall from §3.2 that the Weak Layering Hypothesis offers the option of constructing a recursive PW. Such an extended version of the PW has been employed to capture non-isomorphic aspects of the interface from where some (but not all) of the challenging data stem, as we will see in the following section.

5 Extending the Phonological Word

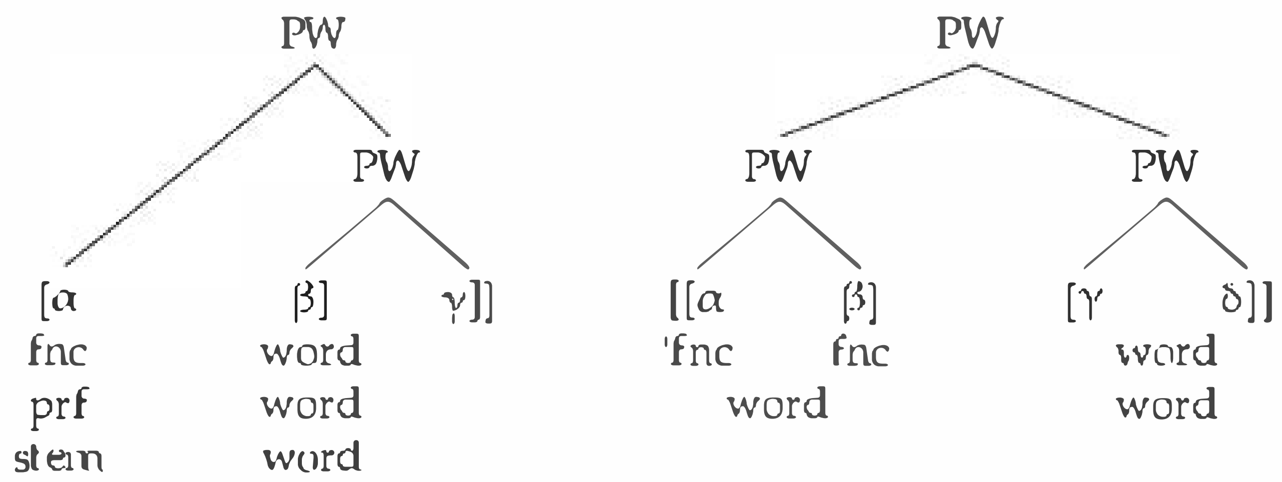
The relaxation of the SLH allowed the emergence of recursive structures. It is doubtful whether constituents lower than the PW can be recursive (Itô and Mester 2009; Kabak and Revithiadou 2009). However, recursion is commonly assumed for higher levels of the Prosodic Hierarchy, such as the PW and the PPh. Selkirk (1995), in a study on the typology of clitics, motivates PW recursion as resulting from morphosyntactic recursion. In (29), phonology "mimics" the nested structure of the morphosyntactic representation by assigning PW boundaries to the edges of constituent X (see also Kabak and Revithiadou 2009 for more examples and argumentation).

(29)



A recursive PW (PW-Rec) comes in two shapes: as a result of *adjunction* (30a) and as a *prosodic compound* (30b). The latter is typically associated with word–word compounds and sequences of function words, the first of which may be inherently stressed or stressed due to binarity.

(30) a. *adjunction* b. *prosodic compound*



In the Prosodic Phonology literature, PW recursion does not always have a morphosyntactic motivation. It is often assumed to arise either as a parsing choice of a particular language or as the result of the subcategorization requirements of individual elements (e.g. Booij and Lieber 1993 on certain Dutch clitics and prefixes). Evidence in support of recursion is primarily drawn from phonology and, specifically, the fashion in which phonological rules apply within the lower and upper PW. The main motivation for the existence of a PW-Rec is the blocking[8] or the optional application of a PW-level phonological process (e.g. Booij 1995, 1996; Peperkamp 1997; Vigário 1999, 2003). In Greek, for example, proclitics and certain prefixes are subject to the same segmental rules that typically apply within the PW (31a), such as s-voicing before a nasal or a voiced fricative:

(31)	a.	/ɣeras-'menos/	ɣera'zmenos
		old-PART	'aged'
	b.	/mas 'ðinis/	maz.'ðinis
		CLT.1PL.GEN give-2SG	'you give us'
	c.	/ðis-'mirii/	ðiz.'mirii
		twice-ten thousand-PL	'twenty thousand ones'

The fact that such sub-minimal elements are part of the extended PW and not of the PW is evidenced by the blocking of resyllabification, e.g. *ɣera.'zme.nos vs. maz.'ði.nis*, which indicates the existence of a boundary at the left edge of the word. The PW boundary prevents the proclitic/prefix from fully incorporating into the PW of its host/base, suggesting that the sub-minimal element adjoins recursively to the PW of the word: [cl/prf $[X^{\mu}]_{PW}]_{PW}$.

The recursive PW has proved to be extremely useful in accounting for attested asymmetries in the degree of cohesion that clitics, affixes, and other dependent elements show in relation to their host. For instance, in many languages enclitics incorporate to their host, whereas proclitics adjoin recursively to it (e.g. Booij 1996 for Dutch; Peperkamp 1997 for Italian; Vigário 2003 for Portuguese; Revithiadou

[8] The blocking of rule application is considered as an immediate result of adjunction. Elements of the outer layer of the PW inherit the properties of the mother constituent, but, because they are not dominated by all of its segments (Chomsky 1986), they can escape (some of) the rules (cf. Booij 1996).

and Spyropoulos 2008 for Greek). Similarly, languages may choose to incorporate suffixes but not prefixes into the PW of their base. Despite its broad use, however, the recursive PW has been called into question as a legitimate prosodic constituent mainly because it is inherently incompatible with the non-recursive nature of phonology (see e.g. Neeleman and Koot 2006; Scheer 2008; Vogel 2009).

6 Conclusions

It is clear from the discussion so far that the PW, as a constituent that lies at the heart of the interface of phonology with morphology and syntax, cannot fully escape the problems naturally associated with the complex nature of the mapping. Four key aspects of PW have been the focus of attention in this chapter: (a) the mechanism that maps a string of elements into a PW and the principles that govern it, (b) the distinction between lexical and post-lexical PW, which essentially reflects the split nature of the mapping itself, (c) the type of criteria used to motivate the PW domain, and (d) the solutions proposed to account for conflicting evidence. The discussion has revealed that all of these issues are surrounded by a number of sometimes thorny problems, and has left numerous questions open for further research. For instance, it is still undecided whether the distinction between lexical and post-lexical PW can be dispensed with or not, or whether certain types of processes are universally associated with the PW.

On the other hand, PW, as a theoretical construct, has been shown to play an important role in language acquisition (e.g. Fikkert 1994; Gerken 1994) and in language change (e.g. Lahiri 2000). Furthermore, psycholinguistic research has established a strong relation between the PW and units of production and perception (Wheeldon and Lahiri 1997, 2002), thus lending further support to the PW as a functionally useful constituent of the Prosodic Hierarchy. Future research will hopefully shed light on less clear aspects of the properties of the PW and the mode in which it is constructed and hence advance our understanding of this pivotal constituent of the Prosodic Hierarchy and, by extension, of the prosodic organization of grammatical elements.

ACKNOWLEDGMENTS

I wish to thank two anonymous reviewers, as well as the editors of the *Companion to Phonology*, especially Marc van Oostendorp and Keren Rice, for their insightful comments. All errors are of course my own.

REFERENCES

Andersen, Henning (ed.) 1986. *Sandhi phenomena in the languages of Europe*. Berlin: Mouton de Gruyter.

Booij, Geert. 1983. Principles and parameters in prosodic phonology. *Linguistics* 21. 249–280.

Booij, Geert. 1985a. The interaction of phonology and morphology in prosodic phonology. In Edmund Gussmann (ed.) *Phono-morphology: Studies in the interaction of phonology and morphology*, 23–34. Lublin: Katolicki Uniwersytet Lubelski.

Booij, Geert. 1985b. Coordination reduction in complex words: A case for prosodic phonology. In van der Hulst & Smith (1985), 143–160.

Booij, Geert. 1988. Review of Nespor & Vogel (1986). *Journal of Linguistics* 24. 515–525.

Booij, Geert. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.

Booij, Geert. 1996. Cliticization as prosodic integration: The case of Dutch. *The Linguistic Review* 13. 219–242.

Booij, Geert. 1999. The role of the prosodic word in phonotactic generalizations. In Hall & Kleinhenz (1999), 47–72.

Booij, Geert & Rochelle Lieber. 1993. On the simultaneity of morphological and prosodic structure. In Sharon Hargus & Ellen M. Kaisse (eds.) *Studies in Lexical Phonology*, 23–44. San Diego: Academic Press.

Booij, Geert & Jerzy Rubach. 1984. Morphological and prosodic domains in Lexical Phonology. *Phonology Yearbook* 1. 1–27.

Booij, Geert & Jerzy Rubach. 1990. Edge of constituent effects in Polish. *Natural Language and Linguistic Theory* 8. 427–463.

Celnarowska, Bozena. 2000. On the (non)recursivity of the prosodic word in Polish. *ZAS Papers in Linguistics* 19. 1–21.

Chen, Matthew Y. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4. 109–149.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Dixon, R. M. W. 2002. The eclectic morphology of Jarawara, and the status of word. In Dixon & Aikhenvald (2002). 125–152.

Dixon, R. M. W. & Alexandra Y. Aikhenvald (eds.) 2002. *Word: A cross-linguistic typology*. Cambridge: Cambridge University Press.

Fikkert, Paula. 1994. On the acquisition of prosodic structure. Ph.D. dissertation, University of Leiden.

Gerken, LouAnn. 1994. Young children's representation of prosodic phonology: Evidence from English speakers; Weak syllable omissions. *Journal of Memory and Language* 33. 19–38.

Grijzenhout, Janet & Barış Kabak (eds.) 2009. *Phonological domains: Universals and deviations*. Berlin: Mouton de Gruyter.

Hall, T. A. 1999. Phonotactics and the prosodic structure in German function words. In Hall & Kleinhenz (1999), 99–131.

Hall, T. A. & Ursula Kleinhenz (eds.) 1999. *Studies on the phonological word*. Amsterdam & Philadelphia: John Benjamins.

Hannahs, S. J. 1991. Prosodic structure and French morphophonology. Ph.D. dissertation, University of Delaware.

Hannahs, S. J. 1995a. *Prosodic structure and French morphophonology*. Tübingen: Niemeyer.

Hannahs, S. J. 1995b. The phonological word in French. *Linguistics* 33. 1125–1144.

Hayes, Bruce. 1989. The prosodic hierarchy in meter. In Paul Kiparsky & Gilbert Youmans (eds.) *Rhythm and meter*, 201–260. San Diego: Academic Press.

Henderson, John. 2002. The word in Eastern/Central Arrernte. In Dixon & Aikhenvald (2002), 100–124.

Hulst, Harry van der. 1984. *Syllable structure and stress in Dutch*. Dordrecht: Foris.

Hulst, Harry van der & Norval Smith (eds.) 1982. *The structure of phonological representations*. 2 parts. Dordrecht: Foris.

Hulst, Harry van der & Norval Smith (eds.) 1985. *Advances in nonlinear phonology*. Dordrecht: Foris.

Inkelas, Sharon. 1989. Prosodic constituency in the lexicon. Ph.D. dissertation, Stanford University.

Inkelas, Sharon & Draga Zec (eds.) 1990. *The phonology–syntax connection*. Chicago: University of Chicago Press.

Itô, Junko & Armin Mester. 2003. Weak layering and word binarity. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata & Shin-Ichi Tanaka (eds.) *A new century of phonology and phonological theory: A festschrift for Professor Shosuke Haraguchi on the occasion of his sixtieth birthday*, 26–65. Tokyo: Kaitakusha.

Itô, Junko & Armin Mester. 2009. The extended prosodic word. In Grijzenhout & Kabak (2009), 135–194.

Kabak, Barış & Anthi Revithiadou. 2009. An interface account to prosodic word recursion. In Grijzenhout & Kabak (2009), 105–134.

Kaisse, Ellen M. 1983. The syntax of auxiliary reduction in English. *Language* 59. 93–122.

Kaisse, Ellen M. 1985. *Connected speech: The interaction of syntax and phonology*. New York: Academic Press.

Kang, Ongmi. 1992. Korean prosodic phonology. Ph.D. dissertation, University of Washington.

Kiparsky, Paul. 1982a. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.

Kiparsky, Paul. 1982b. From Cyclic Phonology to Lexical Phonology. In van der Hulst & Smith (1982: Part I), 131–175.

Kleinhenz, Ursula. 1998. On words and phrases in phonology: A comparative study with focus on German. Ph.D. dissertation, University of Tübingen.

Lahiri, Aditi. 2000. Hierarchical restructuring in the creation of verbal morphology in Bengali and Germanic: Evidence from phonology. In Aditi Lahiri (ed.) *Analogy, leveling, markedness: Principles of change in phonology and morphology*, 70–123. Berlin & New York: Mouton de Gruyter.

Leben, William R. & Firmin Ahoua. 1997. Prosodic domains in Baule. *Phonology* 14. 113–132.

Liberman, Mark. 1975. The intonational system of English. Ph.D. dissertation, MIT.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.

McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.

McCarthy, John J. & Alan Prince. 1993a. Generalized alignment. *Yearbook of Morphology 1993*. 79–153.

McCarthy, John J. & Alan Prince. 1993b. Prosodic morphology I: Constraint interaction a nd satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University.

McCarthy, John J. & Alan Prince. 1998. Prosodic morphology. In Andrew Spencer & Arnold M. Zwicky (eds.) *The handbook of morphology*, 283–305. Oxford & Malden, MA: Blackwell.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague & Paris: Mouton.

Mohanan, K. P. 1982. Lexical Phonology. Ph.D. dissertation, MIT. Distributed by Indiana University Linguistics Club.

Neeleman, Ad & J. van de Koot. 2006. On syntactic and phonological representations. *Lingua* 116. 1524–1552.

Nespor, Marina. 1985. The phonological word in Italian. In van der Hulst & Smith (1985), 193–204.

Nespor, Marina. 1986. The phonological word in Greek and Italian. In Andersen (1986), 65–74.

Nespor, Marina. 1990. Vowel deletion in Italian: The organization of the phonological component. *The Linguistic Review* 7. 375–398.

Nespor, Marina & Irene Vogel. 1982. Prosodic domains of external sandhi rules. In van der Hulst & Smith (1982: Part I), 225–255.

Nespor, Marina & Irene Vogel. 1983. Prosodic structure above the word. In Anne Cutler & D. Robert Ladd (eds.) *Prosody: Models and measurements*, 123–140. Berlin: Springer.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
Odden, David. 1987. Kimatuumbi phrasal phonology. *Phonology Yearbook* 4. 13–36.
Odden, David. 1990. Syntax, lexical rules and postlexical rules in Kimatuumbi. In Inkelas & Zec (1990), 259–277.
Peperkamp, Sharon. 1997. Prosodic words. Ph.D. dissertation, University of Amsterdam.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.
Pyle, Charles. 1972. On eliminating BM's. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 8. 516–532.
Raffelsiefen, Renate. 1999. Diagnostics for prosodic words revisited. In Hall & Kleinhenz (1999), 133–201.
Revithiadou, Anthi. 2008. A cross-dialectal study of cliticisation in Greek. *Lingua* 118. 1393–1415.
Revithiadou, Anthi & Vassilios Spyropoulos. 2008. Greek object clitic pronouns: A typological survey of their grammatical properties. *STUF: Language Typology and Universals* 61. 39–53.
Rice, Keren. 1993. The structure of the Slave (Northern Athabaskan) verb. In Sharon Hargus & Ellen M. Kaisse (eds.) *Studies in Lexical Phonology*, 145–171. San Diego: Academic Press.
Rotenberg, Joel. 1978. The syntax of phonology. Ph.D. dissertation, MIT.
Scheer, Tobias. 2008. Why the prosodic hierarchy is a diacritic and why the interface must be direct. In Jutta M. Hartman, Veronika Hegedüs & Henk van Riemsdijk (eds.) *Sounds of silence: Empty elements in syntax and phonology*, 145–192. Amsterdam: Elsevier.
Selkirk, Elisabeth. 1972. The phrase phonology of English and French. Ph.D. dissertation, MIT.
Selkirk, Elisabeth. 1978a. The French foot: On the status of "mute" e. *Studies in French Linguistics* 1. 141–150.
Selkirk, Elisabeth. 1978b. On prosodic structure and its relation to syntactic structure. Paper presented at the Conference on Mental Representation in Phonology.
Selkirk, Elisabeth. 1980. Prosodic domains in phonology: Sanskrit revisited. In Mark Aronoff & Mary-Louise Kean (eds.) *Juncture*, 107–129. Saratoga: Anma Libri.
Selkirk, Elisabeth. 1981a. On prosodic structure and its relation to syntactic structure. In Thorstein Fretheim (ed.) *Nordic prosody II*, 111–140. Trondheim: Tapir.
Selkirk, Elisabeth. 1981b. On the nature of phonological representation. In Terry Myers, John Laver & John Anderson (eds.) *The cognitive representation of speech*, 379–388. Amsterdam: North Holland.
Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371–405.
Selkirk, Elisabeth. 1995. The prosodic structure of function words. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 439–469. Amherst: GLSA.
Selkirk, Elisabeth & Tong Shen. 1990. Prosodic domains in Shanghai Chinese. In Inkelas & Zec (1990), 313–338.
Topintzi, Nina. 2003. Prosodic patterns and the minimal word in the domain of Greek truncated nicknames. *Proceedings of the 6th International Conference of Greek Linguistics*. Available (April 2010) at: www.philology.uoc.gr/conferences/6thICGL.
Vigário, Marina. 1999. On the prosodic status of stressless function words in European Portuguese. In Hall & Kleinhenz (1999), 253–293.
Vigário, Marina. 2003. *The prosodic word in European Portuguese*. Berlin & New York: Mouton de Gruyter.
Vogel, Irene. 1985. On constraining prosodic rules. In van der Hulst & Smith (1985), 217–233.

Vogel, Irene. 1986. External sandhi rules operating between sentences. In Andersen (1986), 55–64.

Vogel, Irene. 1991. Level ordering in Italian Lexical Phonology? In Pier Marco Bertinetto, Michael Kenstowicz & Michele Loporcaro (eds.) *Certamen Phonologicum II: Papers from the 1990 Cortona Phonology Meeting*, 81–101. Turin: Rosenberg & Sellier.

Vogel, Irene. 2009. The status of the Clitic Group. In Grijzenhout & Kabak (2009), 15–46.

Wheeldon, Linda R. & Aditi Lahiri. 1997. Prosodic units in speech production. *Journal of Memory and Language* 37. 356–381.

Wheeldon, Linda R. & Aditi Lahiri. 2002. The minimal unit of phonological encoding: Prosodic or lexical word. *Cognition* 85. B31–B41.

52 Ternary Rhythm

Curt Rice

1 The facts

Most languages with iterative stress patterns show a simple rhythmic alternation between stressed and unstressed syllables (CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE; CHAPTER 41: THE REPRESENTATION OF WORD STRESS). But in a few cases, stress appears not on every second syllable, but rather on every third one. Patterns of this nature reveal the phenomenon of ternary rhythm.

Ternary rhythm is most easily seen in a language with a stress system that ignores the internal structure of syllables, i.e. a quantity-insensitive system (see CHAPTER 57: QUANTITY-SENSITIVITY). Cayuvava, now extinct, but formerly spoken in parts of Bolivia, is a language well documented in the work of Key (e.g. Key 1961, 1967). It is classified as an isolate, with no established genetic relationship to other languages. Key's fieldwork documents ternary rhythm in Cayuvava and no relevant syllable quantity. Stress in this language appears on every third syllable counting from the right edge of the word. To see the pattern schematically, consider the representations in (1). Each number represents a syllable: "0" represents a syllable with no stress, "1" represents a syllable with primary stress, and "2" represents secondary stress. The pattern is claimed to emanate from the right edge of the word and the representations here are therefore right-justified. The pattern clearly emerges from this schematic representation.

(1) *Ternary alternation patterns of Cayuvava*

a.	10
b.	100
c.	0100
d.	00100
e.	200100
f.	0200100
g.	00200100
h.	200200100
i.	0200200100

The transcribed data from the literature on Cayuvava flesh out the schematic patterns. We can see forms from Key's work in (2), which correspond to the patterns already sketched in (1).

(2) *Cayuvava*

a.	ˈda.pa	'canoe'
b.	ˈto.mo.ho	'small water container'
c.	a.ˈri.po.ro	'he already turned around'
d.	a.ri.ˈpi.ri.to	'already planted'
e.	ˌa.ri.hi.ˈhi.be.e	'I have already put the top on'
f.	ma.ˌra.ha.ha.ˈe.i.ki	'their blankets'
g.	i.ki. ˌta.pa.ra.ˈre.pe.ha	'the water is clean'

To see a more complex instance of ternary rhythm, we turn to Tripura Bangla. Das (2001) describes Tripura Bangla as a dialect of Bangla, resulting from a complicated sociolinguistic situation in the small Indian state of Tripura, where it is a commonly used lingua franca.

One complication in the pattern of Tripura Bangla when compared with Cayuvava is the relevance of syllable structure for stress assignment. Before illustrating this, we can discern the default pattern through a consideration of words consisting only of light syllables. In such strings, we find main stress on the initial syllable and secondary stress emanating rightward in a ternary rhythm. However, a final light syllable cannot bear stress. When the pattern would place stress on a final syllable – e.g. in strings of four or seven syllables – that stress is not realized. This means, for example, that a word consisting of exactly four light syllables will have only one stress, namely the main stress on the word-initial syllable.

(3) *Tripura Bangla default pattern*

a.	ˈra.za	'king'
b.	ˈgo.ra.li	'ankle'
c.	ˈnɛ.ta	'leader'
d.	ˈboi.ra.gi	'mendicant'
e.	ˈbɛ.na.ro.ʃi	'Benaras silk'
f.	ˈbi.βɛ.sɔ.na	'consideration'
g.	ˈʃɔ.ma.lɔ.ˌsɔ.na	'criticism'
h.	ˈo.nu.kɔ.ˌro.ni.jɔ	'imitable'
i.	ˈɔ.no.nu.ˌda.βo.ni.jɔ	'unintelligible'
j.	ˈɔ.no.nu. ˌkɔ.ro.ni.ˌjɔ.ta	'inimitability'

These patterns can be perturbed by closed syllables. Closed syllables can under certain circumstances tolerate stress in word-final position and they can also draw stress off from a word-initial open syllable. A simple generalization is that a stressed light syllable cannot be immediately followed by a closed syllable. When this would happen, the closed syllable bears stress instead. Further complications assign stress to the third syllable when it is heavy and to a word-final closed syllable, unless immediately preceded by a stressed syllable. Providing analyses at this level of detail is not the aspiration here, but both Das (2001) and Houghton (2006) discuss these patterns in detail. The data in (4) have closed syllables in various positions. In positions where we expect stress anyway, the patterns are as in (3). In other cases, the heavy syllable interrupts the default pattern.

(4) *Tripura Bangla quantity effects*

a.	'mal.ʃa	'big metal bowl'
b.	'ɸa.til	'earthen pot'
c.	'ʃɔr.kar	'government'
d.	ɔ.'hɔŋ.kar	'pride'
e.	'ʃɔŋ.rɔk.ˌkɔn	'reservation'
f.	o.'big.ga.ˌɸɔn	'intimation'
g.	'o.nu.ˌbik.kɔn	'microscope'
h.	'baɸ.zaiʈ.ʈa.mi	'adamancy'
i.	'ɔ.no.ˌnu.kɔ.ˌrɔn	'non-imitation'
j.	'ɸɔdʒ.dʒa.lɔ.ˌsɔ.na	'deliberation'
k.	'ʃɔŋ.rɔk.ko.ˌni.jɔ.ta	'preservability'
l.	'ɸa.rɔ.ˌdoʃ.ʃi.kɔ.ta	'expertness'

Ternary rhythm is also visible in Chugach Alutiiq, a Yup'ik language spoken by a small number of individuals in Alaska. This language is most extensively documented in a series of important works by Leer (e.g. 1985a, 1985b). These data and Leer's discussion figure prominently in the literature on ternary rhythm, including many of the theoretical works cited in the present chapter.

Quantity is also relevant for the placement of stress in Chugach Alutiiq, but unlike in Tripura Bangla, a syllable must have a long vowel to perturb the pattern. Closed syllables with short vowels – except when in word-initial position – do not attract stress. Assuming that the default stress pattern is revealed in strings with no relevant quantity distinctions, stress in Chugach appears on the second syllable and then every third syllable thereafter. A word in Chugach Alutiiq with five or six light syllables will have stress on the second and fifth. A word with four, however, will have stress on the second and fourth. Syllables with long vowels always attract stress.

(5) *Chugach Alutiiq*

a.	mu.'lu.kan	'if she takes a long time'
b.	a.'ku.ta.'mek	'*akutaq* (a food) (ABL SG)'
c.	ta.'qa.ma.lu.'ni	'apparently getting done'
d.	a.'ku.tar.tu.'nir.tuq	'he stopped eating *akutaq*'
e.	ma.'ŋar.su.qu.'ta.qu.'ni	'if he (REFL) is going to hunt a porpoise'
f.	'taa.'taa	'her father'
g.	'taa.ta.'qa	'my father'
h.	'naa.ma.ci.'quq	'it will suffice'
i.	'naa.qu.ma.'lu.ku	'apparently reading it'
j.	'naa.ma.'ci.'qua	'I will suffice'
k.	mu.'u.'kuut	'if you take a long time'
l.	u.'lu.te.ku.'ta.'raa	'he's going to watch her'

The three cases presented above are the clearest examples of ternary rhythm that have been uncovered to this point. The languages include some very long words, and even in the quantity-sensitive languages, there are words consisting of long strings of light syllables. Leer's transcriptions of those strings indicate stressed syllables that are separated by two unstressed syllables. This is the empirical basis for the claim that ternary rhythm is a real phenomenon and that a metrical theory of stress assignment must have formal tools that can generate such patterns. While the clearest cases are presented above, there are other languages which have been analyzed as having ternary rhythm, at least in some subset of the data. Most familiar among these are Ho-Chunk, Sentani, and Munster Irish.

Having established the basis for the claim that ternary rhythm is an empirical fact, we turn now to metrical theory and the major strategies that that literature offers for the analysis of these data.

2 Theory and analysis

The preceding section has established that a plausible theory of metrical structure must offer a strategy for modeling ternary rhythm. We turn now to a brief review of the emergence of this issue in the literature and the general tendencies that can be identified.

Hints about the treatment of ternary rhythm can be found very early in the development of a generative theory of stress assignment. As CHAPTER 40: THE FOOT discusses in detail, early work in generative phonology treated stress as the realization of a phonological feature [stress]. In this way, stress was analyzed with tools parallel to those used in the analysis of place of articulation – e.g. [coronal] or [dorsal] – or manner of articulation – e.g. [voice] or [continuant]; cf. Chomsky and Halle (1968) (see CHAPTER 17: DISTINCTIVE FEATURES).

A breakthrough in the study of stress systems came with Liberman's (1975) proposal that stress should be characterized not as a feature with absolute values, but rather as a relation in which two elements differ in their relative prominence. Along with this proposal came hierarchical representations and the introduction of the metrical foot into the generative literature, further developed in Liberman and Prince (1977). The foot naturally invited a more extensive theory of prosodic structure, incorporating segments into syllables, syllables into feet, and so on up the prosodic tree to the phrase or utterance; cf. Nespor and Vogel (2008) (see also CHAPTER 40: THE FOOT; CHAPTER 51: THE PHONOLOGICAL WORD; CHAPTER 84: CLITICS; CHAPTER 50: TONAL ALIGNMENT). This is the context in which any proposed modifications of metrical theory find themselves today. The first extensive typological work on stress systems is found in Hayes (1980). Hayes studies the stress systems of many languages, and identifies a number of parameters that can be used to characterize the variation shown in these languages. Parameters specify points of variation, such as the direction of foot construction, sensitivity to syllable-internal quantitative structure, trochaic or iambic headedness of the feet, whether feet are binary or unbounded, the edge of the word that hosts main stress, and whether or not peripheral material can be excluded from the initial parse through extrametricality. And, indeed, it is precisely the discussion leading up to the proposal of extrametricality that includes the earliest considerations of ternary rhythm. Before turning to the treatment of iterative ternary rhythm in metrical theory and Optimality Theory (OT), the relevance of extrametricality and its competitor are discussed. For a more thorough overview of metrical theory, see van der Hulst (1999) or Hammond (1995).

3 Extrametricality *vs.* ternary feet

Extrametricality as a theoretical tool arose in response to apparent ternary rhythm at the edges of words. The stress pattern of English nouns offers a relevant illustration. In sufficiently long words, we can see that English displays alternating, binary stress assignment, in words such as *Apalachicola, Minnesota, candelabra.* But when we examine the right edges of words more closely, we quickly find that stress is sometimes found not on one of the final two syllables, but rather on the antepenultimate syllable, as in *America, cinema, analysis.* In this way, we identify a fundamentally binary system that has a ternary component, namely a three-syllable window at the right edge of the word. A model that only constructs binary feet over an entire word would not be able to generate this pattern. Specifically, the construction of feet from right to left in English nouns would always result in penultimate stress. How can a binary foot "reach in" far enough to position primary stress on the antepenult? To model antepenultimate stress, two possible enhancements of the theory were entertained early on. One of these is extrametricality.

Extrametricality is a theoretical tool that does not explicitly entail enhancement of the inventory of feet. Instead, it provides a particular strategy for foot construction, or parsing a string of syllables. In particular, extrametricality excludes a peripheral syllable from the string to be parsed into feet. In the case of English nouns, exclusion of the final syllable, followed by construction of a binary left-headed foot, will place stress on the antepenultimate syllable. Extrametricality is also illustrated in CHAPTER 43: EXTRAMETRICALITY AND NON-FINALITY. As we will see below, some later work on iterative ternary rhythm relativizes the peripherality requirement, such that syllables can be excluded from the string not only when they are word-peripheral but also, for example, when they are foot-peripheral.

A conceptually different approach from extrametricality would be to enhance the model such that it also includes ternary feet. Data of the type described for English nouns would then be modeled by building a ternary foot at the edge, followed by the construction of binary feet iterating leftward. Since the stress that is found in the three-syllable window is the primary stress, this amounts to a proposal that primary stress can be modeled through the use of one kind of foot while secondary stress requires another. Such proposals can be found for other points of parametric variation for foot construction, as well. For example, primary stress may require the use of a quantity-sensitive foot, while iterative secondary stress seems to be quantity-insensitive (van der Hulst 1984, 1999).

One strategy for modeling an edgemost ternary domain – extrametricality – enhances the parsing strategies available in the theory, while the other strategy – a ternary foot – enhances the inventory of feet available in the theory.

4 Modeling iterative ternary rhythm

The parameters of metrical theory specify the nature of feet and control their construction across words in languages. The feet that are constructed are constituents that create a domain for the assignment of relative prominence. Prince (1983) offers an alternative approach without internal constituency, representing relative prominence instead only with a grid, and reconstruing some of Hayes's parameters such that their effects can be replicated without binary constituents (see also CHAPTER 41: THE REPRESENTATION OF WORD STRESS). The debate about constituency includes argumentation based on sensitivity of non-stress phenomena to feet, as reviewed in Kenstowicz (1993). This debate is present in much of the subsequent literature, finding one of its most extensive and significant considerations in Halle and Vergnaud (1987).

Halle and Vergnaud's constituentized grid representation integrates grids and feet. Grids are built but the gridmarks are grouped and these groupings represent constituents. The construction process implements parameter settings here, too, also in pursuit of a typology of stress systems. And it is here, in Halle and Vergnaud's opus, that we find the first discussion of iterative ternary stress presented in a major work on stress system typology.

Halle and Vergnaud of course draw on papers and presentations regarding ternarity that were floating about in the immediately preceding years, with some issues already nascent in McCarthy (1979). The discussion of iterative ternary rhythm and its implications for the typologies under consideration in the relevant literature was initiated by Levin (1985), which was ultimately published in a significantly modified form as Levin (1988). Levin's work drew on the data from Cayuvava in (2).

Halle and Vergnaud (1987) discuss neither Tripura Bangla nor Chugach Alutiiq. Regarding the latter, Leer's (1985a, 1985b) careful and important results would soon influence the details of the constituentized grid theory. Leer's work was picked up on in Rice (1988), where an analysis in the spirit of Halle and Vergnaud (1987) is advanced. This, in turn, influenced subsequent revisions of the theory, as presented in Halle (1990).

The theory developed by Halle and Vergnaud models ternary rhythm through the construction of ternary feet, extending to the problem of iterative ternary rhythm the spirit of the approach discussed above in the context of word-final three-syllable stress windows.

A competing approach also reflects that earlier debate. This competitor maintains a size limit such that feet are maximally binary. Ternary rhythm is achieved with a parsing strategy that leaves occasional syllables unincorporated into feet, extending the basic notion of extrametricality; cf. Hammond (1990) and Hayes (1995).

These two general approaches, to be illustrated presently, form the heart of the theoretical debate occasioned by ternary stress patterns. As we will see in the discussion of ternary rhythm and OT below, the debate persists there, too. We turn now to the chronologically first approach, namely an analysis of ternary rhythm using ternary feet.

5 Ternary feet

5.1 Amphibrachs

At first glance, the Cayuvava stress patterns in (1) and (2) suggest an analysis with dactylic feet (strong–weak–weak), built from right to left. If we maintain a parametric strategy for constructing feet, then the independently established presence in the theory of a parameter placing heads at the left or right edge of the foot means that the admission of dactyls to the inventory of derivable feet would imply the introduction of anapests (weak–weak–strong) as well. Allowing a ternary foot with its head at the left edge implies via the relevant parameter the possibility of constructing a ternary foot with its head at the right edge. With Cayuvava as the only known case of iterative ternary rhythm at the time of this theoretical work, generating dactyls would lead to the phenomenon of overgeneration, i.e. being able with the tools of the theory to generate patterns not known to exist.

In pursuit of a restrictive theory, Levin (1988) therefore takes a different tack, relaxing metrical theory just enough to allow for exactly one type of ternary foot, instead of two; dactyls and anapests are disallowed, but the theory now permits amphibrachs, i.e. ternary feet with prominence on the middle syllable, employing a strategy described below. When combined with final extrametricality – which can be overridden when necessary to build at least one foot on the (minimal) disyllabic words – the construction of amphibrachs will yield a footing of the schematic patterns in (1) that correctly locates stress, as seen in (6). Parentheses indicate feet and angled brackets mark extrametricality. In longer words, initial lone syllables are left unfooted, by stipulation.

(6) *Ternary alternations parsed into amphibrachs*

a.	(10)
b.	(10)⟨0⟩
c.	(010)⟨0⟩
d.	0(010)⟨0⟩
e.	(20)(010)⟨0⟩
f.	(020)(010)⟨0⟩
g.	0(020)(010)⟨0⟩
h.	(20)(020)(010)⟨0⟩
i.	(020)(020)(010)⟨0⟩

Halle and Vergnaud (1987) adopt Levin's strategy, and also limit Universal Grammar (UG) to this one type of ternary foot. They parameterize the requirement that the head of a constituent be at its edge. When this parameter is set such that the head is not required to be at the edge of a constituent, the only ternary foot that can emerge is an amphibrach; cf. Rice (1988, 1990) for related discussion. The approach developed by Levin and widely discussed in publications by Halle and Vergnaud effectively views iterative ternary rhythm as evidence for expanding the inventory of feet. Constituents may have one head, but as many as two non-heads. For them, there is no hierarchical structure within the foot, so that this approach generates flat ternary feet.

Another main thrust of the literature also sees a proposal with ternary feet, but now with internal hierarchical structure. Early proponents of this include Dresher and Lahiri (1991) and Rice (1992), building on Rice (1990). Leer's (1985b) article offers the leading idea, namely identifying the quantitative equivalence of two light syllables with a single heavy syllable, allowing either of those configurations to be the head of a foot. Taking into account a non-head consisting of a light syllable, a foot might consist of three light syllables, two of which are themselves a subconstituent. Hence, ternary feet become an option.

In the foot typology of Hayes (1980), some languages were identified in which the heads of feet must be heavy, a foot type dubbed the obligatory branching foot.

Rice (1992) in particular draws a parallel between Hayes's obligatory branching feet and the analysis of ternarity under consideration, since the head consisting of one heavy syllable or two light ones could be construed as obligatory branching. That analysis is also relevant to the contrast between (6e) and (6d): in the former case, two word-initial syllables concluding the right-to-left parse are sufficient for a foot, while the single syllable in the latter case is not. The analysis in Rice (1992) suggests that degenerate feet must have a head, and in the case of the ternary feet constructed for Cayuvava, two lights are required to constitute a head, hence the minimal foot (and word) is binary. In the approach with flat ternary feet, it is unclear why a minimum of two syllables is necessary for a degenerate foot. Additional discussion related to this approach can be found in Everett (1988), Hewitt (1992), Rice (1993), Blevins and Harrison (1999), van der Hulst (1999), Rifkin (2003), and other references mentioned below.

5.2 *Weak local parsing*

The appearance of iterative ternary stress patterns in the literature on metrical phonology triggered, as noted above, a second strategy. Instead of increasing the set of possible feet, this second strategy increased the set of possible parsing strategies. This approach is developed in Hayes (1995), drawing on earlier work by Hammond (1990). In Hayes's approach, universal grammar allows only three kinds of feet, as in (7) (see also CHAPTER 44: THE IAMBIC–TROCHAIC LAW).

(7) *The Hayesian foot typology*

a.	Syllabic trochee	(x	.)		
		σ	σ		
b.	Moraic trochee	(x	.)		(x)
		L	L	*or*	H
c.	Iamb	(.	x)		(x)
		L	σ	*or*	H

No exhaustive parsing of a string with any of these feet will give an iterative ternary pattern. But non-exhaustive parsing can do that. Hayes proposes that UG include a weak local parsing parameter that creates the possibility of leaving an unparsed syllable between each foot. The unparsed syllable can by stipulation only be a light one. Having an unparsed light syllable between each foot yields a ternary pattern using only binary feet, as in (8).

(8) *Ternary alternations parsed into non-exhaustive binary feet*

a. (10)
b. (10)(0)
c. 0(10)(0)
d. 00(10)(0)
e. (20)0(10)(0)
f. 0(20)0(10)(0)
g. 00(20)0(10)(0)
h. (20)0(20)0(10)(0)
i. 0(20)0(20)0(10)(0)

For the Cayuvava pattern, trochees are constructed from right to left; whether they are moraic or syllabic trochees is irrelevant, since there is no quantity distinction in this language. The parsing also uses final extrametricality and, of course, weak local parsing. The final syllable in (8c) is extrametrical and the initial syllable is unfooted, because it is too little to be a foot, since no degenerate feet are allowed. In (8d) the influence of weak local parsing is seen; in this form, there is in fact sufficient material at the left edge of the word to form a foot. Since doing so would result in adjacent feet – which is not allowed with weak local parsing – no foot can be formed. Not until we have six syllables, as in (8e), is there sufficient space to build two non-adjacent – i.e. weakly local parsed – feet. A detail beyond the scope of this chapter is that adjacency can be tolerated in the case of adjacent heavy syllables – as in Chugach – suggesting that the requirement to incorporate heavy syllables into feet has priority over the prohibition on adjacent feet under weak local parsing.

5.3 *Summary*

At the level of analysis, we have seen that there are two primary strategies for constructing constituents across strings when the goal is to achieve iterative ternary alternations. One strategy is to expand the inventory of constituents, and here there are also two approaches. In the approach developed by Levin (1988) and Halle and Vergnaud (1987), flat ternary feet are allowed. Any non-head in a foot must be adjacent to its head. This allows exactly one kind of ternary foot, namely an amphibrach, where the head is not found at the edge of the foot, but rather is flanked by two non-heads. The second inventory-expanding approach, as seen primarily in Dresher and Lahiri (1991) and Rice (1992), relaxes the requirement that the head of a foot can span only one syllable. Feet that require heavy heads can draw their material either from one heavy syllable or from two light ones.

The alternative to expanding the foot inventory is expanding the strategies available for constructing binary feet, and the primary representative of this approach is Hayes (1995). In this approach, binary feet are constructed in a new way. There are two necessary properties to the weak local parsing of a string: feet must be non-adjacent and they must be minimally non-adjacent. These requirements lead to iterative construction of feet that are separated by one light syllable.

There was, as noted, an early debate in the metrical literature regarding the need for constituency; perhaps stress systems can be modeled simply with a theory of prominence as represented with a grid, and feet are superfluous. This debate has not shown itself in the context of ternary rhythm, insofar as the literature lacks a grid-only analysis of iterative ternary rhythm.

6 Ternary rhythm in Optimality Theory

The typological enterprise in generative grammar has enjoyed enhanced prominence in the era of OT (McCarthy and Prince 1993; Prince and Smolensky 1993). One of the core foci in the OT literature is typology (Archangeli and Langendoen 1997; Roca 1997; Kager 1999). Classic OT achieves typological insights by having a universally fixed set of constraints. Variation is modeled with constraint reranking. The factorial typology of constraint rankings defines the range of possible grammars.

Stress patterns and metrical theory have played an important role in the construction and exploration of optimality-theoretic approaches to modeling grammar. Indeed, one of the important early discussions of the power of violable constraints was built around the pursuit of a parallelist strategy for achieving the effects of directionality. The insight in this discussion is that minimal violation of a requirement that all feet be at one edge of the word (ALLFT-L or ALLFT-R), when combined with the force of a requirement that all syllables be parsed into feet (PARSE-σ), will yield as optimal a parse identical with serial foot construction from one edge of a string to the other.

Less present in the OT literature, however, has been discussion of ternary rhythm. There has been almost no debate about the contrast between analyses using ternary feet and those using non-exhaustive parsing with binary feet. Indeed, Rice (2007) is to the best of my knowledge the only publication in which the issue is even mentioned, although Hyde (2002) also offers relevant perspectives on the nature of ternary parsing.

The most prominent discussions of ternary rhythm in OT mimic the weak local parsing approach, as in Ishii (1996) and Elenbaas and Kager (1999). Elenbaas and Kager take an important principled position on methodology. In particular, they articulate and adopt the goal of deriving iterative ternary rhythm with tools that are already necessary to account for other phenomena. This laudable position of theirs contrasts with the too frequent practice in OT analyses of positing new constraints to give new analyses. That practice has substantial implications, in light of the methodology of the factorial typology noted above; introducing a new constraint introduces many new grammars, and the restricted typological enterprise as construed in classic OT is substantially challenged with every new constraint that is introduced.

The analysis of ternary rhythm in OT based on underparsing has two crucial components. First, PARSE-σ – which requires that syllables be incorporated into feet – must be relatively low-ranked. This will be important in allowing optimization of an incomplete parse along the lines seen in (8). But simply ranking PARSE below a requirement that all feet align with the right edge of the word will yield a parse with only one foot. Note that ALLFT-R awards a violation for every syllable intervening between the right edge of a foot and the right edge of the word, for each foot.

(9) *Underparsing with low-ranked* PARSE

σσσσσσ	ALLFT-R	PARSE
a. ('σσ)('σσ)('σσ)	*!*****	
b. σ('σσ)σ('σσ)	*!**	**
☞ c. σσσσ('σσ)		****

This brings us to the second crucial component. To counter the pressure of ALLFT-R, parsing of at least some of the other syllables must be rewarded. The solution offered builds on well-established insights that lapses in long parses should be avoided; cf. Selkirk (1984). There are various *LAPSE constraints in the literature (e.g. Kager 1994; Green 1995; Gordon 2002), where the leading idea is that a string of more than two unstressed syllables is disfavored. With *LAPSE ranked above ALLFT-R, the optimal parse will show only as much parsing as is necessary to minimize violations of *LAPSE, and will favor options in which the feet are relatively toward the right.

(10) *Ternary rhythm with* *LAPSE

σσσσσσ	*LAPSE	ALLFT-R	PARSE
a. ('σσ)(σ'σ)(σ'σ)		****!**	
b. (σ'σ)σ(σ'σ)σ		****!*	**
☞ c. σ(σ'σ)σ(σ'σ)		***	**
d. σσ(σ'σ)(σ'σ)	*!	**	**

The OT analysis of ternary stress requires consideration of many more details and much more discussion, which is to be found in the cited works. For present purposes, it is sufficient to note that an analysis akin to Hayes's weak local parsing strategy is achieved through the interaction of *LAPSE and ALLFT-R. Although the methodology of pursuing an analysis built simply on the reranking of independently motivated constraints is commendable, that goal has not yet been achieved. For example, careful study reveals that multiple versions of *LAPSE will be necessary, one of which is specifically designed for the ternary cases; this is made laudably explicit in Houghton (2006). While the use of ternary-specific tools is not an a priori flaw of these analyses, it nonetheless keeps them from clearing the high bar set in pursuit of an analysis by pure reranking of constraints that are not ternary-specific.

In addition to facilitating an illustration of the strategy that has been most thoroughly pursued in providing an analysis of ternary rhythm within OT, the patterns under consideration here raise another important methodological point. Future work in OT that considers the relative merits of the two main types of analyses illustrated above – ternary feet or underparsing – must consider related issues about the division of labor among the modules of the theory. Consider, for example, the possibility that a particular analysis intends to optimize a parse that does not use flat ternary feet, or amphibrachs. How will such feet and their optimization be avoided?

If it is possible that prosodic structure is present in inputs – a possibility required by the richness of the base methodology – then amphibrachs are possibly present in inputs. One way in which these can be prevented from being selected as optimal is with a constraint that rules them out. Introducing an anti-amphibrach constraint, however, implicitly raises the possibility that it could be ranked relatively low, which in turn could open the door to the optimization of such structures. If one's position is that amphibrachs are never optimal, then it is unfortunate to achieve this universal exclusion with a constraint. The alternative is to provide structure to Gen, such that the output of Gen cannot include amphibrachs. The tension between these possibilities is the focus of Rice (2007), and it is one of the general theoretical issues raised in OT by the study of ternary rhythm.

7 Implications and directions for future research

As noted earlier, one strategy for modeling an edgemost ternary domain – extrametricality – enhances the parsing strategies available in the theory, while the other strategy – a ternary foot – enhances the inventory of feet available in the theory. On what basis can a theoretician select the model to be pursued? Are these two options really different from one another in some meaningful sense? Does one approach allow for the description of some kind of situation that the other one does not?

If the approaches do not make different predictions, are there other strategies available for selecting among them? One possibility would be to appeal to general principles of theory construction or to findings in other realms of cognitive science. Such principles or findings may have implications for selecting among competing theories of phonology.

To see one example of argumentation for selecting among competing theories, we can turn to Hayes's (1980) argumentation for extrametricality over ternary feet. This argument is based primarily on identifying differences in the types of systems the competing theories predict. In the cited work, Hayes compares his foot inventory at that point with a proposal made by Morris Halle, attributed to "class notes" (Hayes 1980: 114ff.). Halle had introduced ternary feet into his version of the theory, while Hayes had developed the approach using extrametricality.

Hayes begins his argument against Halle's inventory by stating the following: "I know of no languages whose stress patterns could simply be described using feet of the [ternary] form" (Hayes 1980: 115). This quotation suggests that a gap in the typology – namely the absence of languages with iterative ternary rhythm – can be used as an argument against a theory that could in fact model that.

At the conclusion of the section, we are again encouraged to adopt an inventory with binary feet and peripheral extrametricality, in part because it provides "an explanation for why feet which have [ternary] surface forms . . . are never assigned iteratively" (Hayes 1980: 122).

This example from the early literature on ternary rhythm illustrates an argument based on overgeneration. The theories are evaluated, and the one that generates a pattern not known to exist is dispreferred on those grounds. Gaps in the typology become a criterion for theory selection. However, as we now know from the sections above, this gap was soon revealed to be accidental.

This is not the only example in the generative literature of argumentation based on gaps in the empirical record. We might use this occasion to ask whether such experiences are relevant as we hone our methodology for identifying the properties of UG. The following paragraphs present some of the broader implications that may be explored on the basis of our discussion of ternary rhythm.

The typological enterprise as widely practiced in generative phonology aspires to model grammatical variation through simple formal manipulations of various components of the theory: reranking of constraints, different rules or rule orders, or the setting of parameters to different values.

Generative linguists often see themselves as doing work grounded in a reliable typology of the structures found in natural languages. Our goal – cf. Odden (forthcoming) – is to identify the limits of the human linguistic capacity and to model that knowledge. What is a possible grammar, and what is not? What cognitive structures must be posited to restrict the outcome of the language acquisition process to possible grammars, thereby rendering unattainable the impossible ones? And regarding the case at hand, what does the fact of ternary rhythm force us to posit in our theory?

These questions are sometimes studied from "above." Researchers could take as their starting point a theory of cognitive capacity and build a model of linguistic knowledge within the context of that theory. Demonstrated incompatibility of a conceivable linguistic structure with a known fact about our cognitive system would be an argument for genuine universal ungrammaticality: cf. Odden (2008) and Reiss (2008).

Alternatively, one could develop a theory from "below." In this case, one would approach the matter through a deep study of one language or one family of languages, or through a carefully selected set of unrelated languages. Regardless of the starting point one adopts, any predictive model of linguistic knowledge will be held accountable to facts about natural languages.

We work to enhance the empirical foundation for developing theories of grammars by studying individual languages. The very act of documenting, describing, and analyzing the multifarious properties and subsystems that are found in natural language entails engagement in a typological enterprise; cf. Newmeyer (2005) and Hyman (2009). A model of linguistic knowledge that allows the construction of a particular grammar gains credibility *vis à vis* its competitors when a specific language is identified that requires precisely that grammar.

We might call this the *positive typological enterprise*: certain structures are attested in the set of well-studied natural languages, and any theory of grammar must sanction the generation of such structures.

But there is also a *negative typological enterprise*. Work on this side of the program aspires to model the absence of unattested structures. If we imagine competing models of linguistic knowledge, all of which satisfy the positive side of the enterprise insofar as all of them generate those structures that are known to exist, then we need some criterion for choosing among them. Proposed models of linguistic knowledge are therefore routinely criticized from the negative side, i.e. for allowing the construction of a grammar that is not known to be instantiated by any natural language; this is the state of overgeneration, of which we saw an example above.

A model that overgenerates is in principle inferior to one that does not. A model that completely fails to overgenerate matches the systems that it cannot generate with those that are unattested.

While this seems at first glance to be an important goal, the danger we must guard against when attempting to eliminate aspects of a theory that overgenerates is the equation of *unattested* with *impossible*. How can we know whether a structure that is absent from the empirical database is merely unattested – as was the case with iterative ternary rhythm – or genuinely beyond the grasp of UG?

Finding an answer to this question seems insurmountably forbidding when we realize that distinguishing the merely unattested from the cognitively impossible would be no easier if all languages were thoroughly documented, studied, and analyzed. Those activities, of course, may fill gaps in our knowledge, and they will certainly generate a richer base for our research enterprise, thereby contributing to a deeper understanding of our cognitive capacity. But even if all languages were deeply understood, any linguist would still be able to posit conceivable albeit unattested structures, and all theories that overgenerate would make this challenge easy. Such conceivable but unattested structures can be assumed to be universally ungrammatical only if we assume that all possible structures in fact do appear somewhere in the set of human languages. This assumption, alas, is no more plausible than an assumption that all possible structures for eyes, for example, are attested somewhere in the animal kingdom.

This realization presents a significant confrontation to the bottom-up approach to linguistic theory construction. Building theories on the basis of what is attested and unattested tends to confuse *does not exist* with *cannot exist*. The work of linguists is not to explain linguistic structures that do exist, but rather to explain linguistic structures that can exist. While the starting point may be the empirical record, that cannot be the ending point. The empirical record does not and cannot show us the limits of human linguistic capacity. The empirical record cannot reveal what is necessarily beyond the realm of our grammatical competence; cf. Isac and Reiss (2008).

8 Conclusions

Although it is not the purpose of this chapter to go deeply into methodological issues, it is essential nonetheless to highlight the importance of doing so. The study of ternary rhythm and the history of analyses of this phenomenon in the generative literature raise issues of the kind presented here. The remaining challenge is to take these discussions further, not only in this context, but whenever we enter into discussions arguing for the selection of one theoretical model over another.

There is more work to be done in metrical theory and on ternary rhythm. This must address several issues, both in the context of a specific analysis and with respect to the theory with which that analysis is built. Naturally, any analysis has to satisfy the positive typological enterprise by allowing the generation of those patterns that are attested. An analysis must take a position on the kind of constituents that are constructed. Are they binary or ternary?

If there are ternary feet, i.e. feet with three terminals, is there any internal structure to those constituents, or are they flat ternary structures? If there are not ternary feet, how are binary feet constructed with iterative non-exhaustivity, i.e. periodic non-parsing as with the earlier theory of weak local parsing? What is the fate of those syllables that are left unparsed? How are they incorporated into the prosodic structure of the word, and what implications does this have for a theory of layering in prosodic structure (Selkirk 1984)?

Beyond the basic matter of developing an analysis, future work on this topic should also address the role of typological evidence in the theory being offered, specifically the extent to which patterns absent from the empirical record are considered to be not merely unattested but unattestable. In this way, the typological enterprise will remain an important topic of discussion, and the languages showing ternary rhythm will play their significant role in future refinements of metrical theory and the methodologies of research in generative linguistics.[1]

[1] Works relevant to the study of ternary rhythm that have not been mentioned in this article, but which students of this topic should consult, include the following: Crowhurst (1992); Idsardi (1992); Kager (1993); Green and Kenstowicz (1995); Rowicka (1996); van de Vijver (1998); Elenbaas (1999); Hyde (2001); Gordon (2002); McCartney (2003); Karttunen (2006); Buckley (2009).

ACKNOWLEDGMENTS

For feedback on an early draft of this chapter, I thank Sylvia Blaho, Peter Jurgec, Bruce Morén-Duolljá, Dave Odden, Charles Reiss, Bridget Samuels, and two anonymous reviewers. I also thank Sylvia Blaho for extensive support during the editing process. I am indebted to phonologists too many to name, with whom I have had fruitful discussions about metrical theory and ternarity over the years. Some of the perspectives advanced here owe an intellectual debt to Hale and Reiss (2008), which is hereby gratefully acknowledged.

REFERENCES

Archangeli, Diana & D. Terence Langendoen. 1997. *Optimality Theory: An overview.* Cambridge, MA & Oxford: Blackwell.

Blevins, Juliette & Sheldon P. Harrison. 1999. Trimoraic feet in Gilbertese. *Oceanic Linguistics* 38. 203–230.

Buckley, Eugene. 2009. Locality in metrical typology. *Phonology* 26. 389–435.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Crowhurst, Megan J. 1992. Minimality and foot structure in metrical phonology and prosodic morphology. Ph.D. dissertation, University of Arizona.

Das, Shyamal. 2001. Some aspects of the phonology of Tripura Bangla and Tripura Bangla English. Ph.D. dissertation, Central Institute of English and Foreign Languages, Hyderabad (ROA-493).

Dresher, B. Elan & Aditi Lahiri. 1991. The Germanic foot: Metrical coherence in Old English. *Linguistic Inquiry* 22. 251–286.

Elenbaas, Nine. 1999. A unified account of binary and ternary stress: Considerations from Sentani and Finnish. Ph.D. dissertation, Utrecht University.

Elenbaas, Nine & René Kager. 1999. Ternary rhythm and the lapse constraint. *Phonology* 16. 273–329.

Everett, Daniel L. 1988. On metrical constituent structure in Pirahã phonology. *Natural Language and Linguistic Theory* 6. 207–246.

Gordon, Matthew. 2002. A factorial typology of quantity-insensitive stress. *Natural Language and Linguistic Theory* 20. 491–552.

Green, Thomas. 1995. The stress window in Pirahã: A reanalysis of rhythm in Optimality Theory. Unpublished ms., MIT (ROA-45).

Green, Thomas & Michael Kenstowicz. 1995. The lapse constraint. *Proceedings of the 6th Annual Meeting of the Formal Linguistics Society of the Midwest.* 1–14 (ROA-101).

Hale, Mark & Charles Reiss. 2008. *The phonological enterprise.* Oxford: Oxford University Press.

Halle, Morris. 1990. Respecting metrical structure. *Natural Language and Linguistic Theory* 8. 149–176.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress.* Cambridge, MA: MIT Press.

Hammond, Michael. 1990. Deriving ternarity. Unpublished ms., University of Arizona, Tucson.

Hammond, Michael. 1995. Metrical phonology. *Annual Review of Anthropology* 24. 313–342.

Hayes, Bruce. 1980. A metrical theory of stress rules. Ph.D. dissertation, MIT. Published 1985, New York: Garland.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Hewitt, Mark. 1992. Vertical maximization and metrical theory. Ph.D. dissertation, Brandeis University.

Houghton, Paula. 2006. Ternary stress. Unpublished ms., University of Massachusetts, Amherst (ROA-836).

Hulst, Harry van der. 1984. *Syllable structure and stress in Dutch*. Dordrecht: Foris.

Hulst, Harry van der. 1999. Word accent. In Harry van der Hulst (ed.) *Word prosodic systems in the languages of Europe*, 3–115. Berlin & New York: Mouton de Gruyter.

Hyde, Brett. 2001. Metrical and prosodic structure in Optimality Theory. Ph.D. dissertation, Rutgers University (ROA 476).

Hyde, Brett. 2002. A restrictive theory of metrical stress. *Phonology* 19. 313–359.

Hyman, Larry M. 2009. How (not) to do phonological typology: The case of pitch-accent. *Language Sciences* 31. 213–238.

Idsardi, William J. 1992. The computation of prosody. Ph.D. dissertation, MIT.

Isac, Daniela & Charles Reiss. 2008. *I-language: An introduction to linguistics as cognitive science*. Oxford: Oxford University Press.

Ishii, Toru. 1996. An optimality theoretic approach to ternary stress systems. *UCI Working Papers in Linguistics* 2. 95–111.

Kager, René. 1993. Alternatives to the iambic-trochaic law. *Natural Language and Linguistic Theory* 11. 381–432.

Kager, René. 1994. Ternary rhythm in alignment theory. Unpublished ms., Utrecht University (ROA-35).

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Karttunen, Lauri. 2006. The insufficiency of paper-and-pencil linguistics: The case of Finnish prosody. Unpublished ms., Stanford University (ROA-818).

Kenstowicz, Michael. 1993. Evidence for metrical constituency. In Kenneth Hale & Samuel J. Keyser (eds.) *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, 257–273. Cambridge, MA: MIT Press.

Key, Harold H. 1961. Phonotactics of Cayuvava. *International Journal of American Linguistics* 27. 143–150.

Key, Harold H. 1967. *Morphology of Cayuvava*. The Hague: Mouton.

Krauss, Michael (ed.) 1985. *Yupik Eskimo prosodic systems: Descriptive and comparative studies*. Fairbanks: Alaska Native Language Center.

Leer, Jeff. 1985a. Prosody in Alutiiq (the Koniag and Chugach dialects of Alaskan Yupik). In Krauss (1985), 77–133.

Leer, Jeff. 1985b. Toward a metrical interpretation of Yupik prosody. In Krauss (1985), 159–172.

Levin, Juliette. 1985. Evidence for ternary feet and implications for a metrical theory of stress rules. Unpublished ms., University of Texas, Austin.

Levin, Juliette. 1988. Generating ternary feet. *Texas Linguistic Forum* 29. 97–113.

Liberman, Mark. 1975. The intonational system of English. Ph.D. dissertation, MIT.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.

McCarthy, John J. & Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University.

McCartney, Steven J. 2003. Ternarity through binarity. Ph.D. dissertation, University of Texas, Austin.

Nespor, Marina & Irene Vogel. 2008. *Prosodic phonology*. Berlin & New York: Mouton de Gruyter. 1st edn. 1986, Dordrecht: Foris.

Newmeyer, Frederick J. 2005. *Possible and probable languages: A generative perspective on linguistic typology*. Oxford: Oxford University Press.

Odden, David. 2008. Ordering. In Vaux & Nevins (2008), 61–120.

Odden, David. Forthcoming. Rules v. constraints. In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.) *The handbook of phonological theory*. 2nd edn. Malden, MA & Oxford: Wiley-Blackwell.

Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14. 19–100.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Reiss, Charles. 2008. Constraining the learning path without constraints, or the OCP and NoBanana. In Vaux & Nevins (2008), 252–301.

Rice, Curt. 1988. Stress assignment in the Chugach dialect of Alutiiq. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 24. 304–315.

Rice, Curt. 1990. Pacific Yupik: Implications for metrical theory. *Coyote Papers* 3. Tucson: Department of Linguistics, University of Arizona.

Rice, Curt. 1992. Binarity and ternarity in metrical theory: Parametric extensions. Ph.D. dissertation, University of Texas, Austin.

Rice, Curt. 1993. A note on ternary stress in Sentani. *University of Trondheim Working Papers in Linguistics* 17. 67–71.

Rice, Curt. 2007. The roles of GEN and CON in modeling ternary rhythm. In Sylvia Blaho, Patrik Bye & Martin Krämer (eds.) *Freedom of analysis?*, 233–255. Berlin & New York: Mouton de Gruyter.

Rifkin, Jay I. 2003. Ternarity is prosodic word binarity. In Jeroen van de Weijer, Vincent J. van Heuven & Harry van der Hulst (eds.) *The phonological spectrum*, vol. 2: *Suprasegmental structure*, 127–150. Amsterdam & Philadelphia: John Benjamins.

Roca, Iggy (ed.) 1997. *Derivations and constraints in phonology*. Oxford: Clarendon Press.

Rowicka, Grażyna. 1996. 2+2=3: Stress in Munster Irish. In Henryk Kardela & Bogdan Szymanek (eds.) *A Festschrift for Edmund Gussmann*, 217–238. Lublin: University Press of the Catholic University of Lublin.

Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Vaux, Bert & Andrew Nevins (eds.) 2008. *Rules, constraints, and phonological phenomena*. Oxford: Oxford University Press.

Vijver, Ruben van de. 1998. The iambic issue: Iambs as a result of constraint interaction. Ph.D. dissertation, University of Leiden.

53 Syllable Contact

Misun Seo

1 Introduction

Syllable contact is a notion introduced to describe the sonority relation between adjacent segments across a syllable boundary, that is, between heterosyllabic coda and onset segments. According to Hooper (1976), Murray and Vennemann (1983), and Vennemann (1988), among others, there is a cross-linguistic preference to avoid rising sonority across a syllable boundary; this tendency is formulated as the Syllable Contact Law (SCL). This law states, for example, that *al.ta*, with falling sonority, is preferred to *at.la*, with rising sonority. This law has been adduced to account for both diachronic and synchronic sound alternations in coda–onset clusters.

In this chapter, I will review the notion of syllable contact by referring to sound alternations involving coda–onset clusters observed in different languages. The chapter is organized as follows. In §2, I offer an overview of sonority and sonority scales on which the SCL is crucially based. In addition, I review previous proposals on syllable contact and different types of diachronic and synchronic phonological changes analyzed as repair strategies for bad syllable contact. In §3, I present the issues and debates on syllable contact such as the necessity of syllable contact, the categorical *vs.* gradient nature of syllable contact, language-specific variation, and problems of syllable contact (see also CHAPTER 49: SONORITY).

2 Syllable contact

The notion of syllable contact is essentially based on the relative sonority of segments. Thus, in §2.1, I first review sonority and then the SCL. In §2.2, I provide examples of different types of diachronic and synchronic phonological changes which have been analyzed using the notion of syllable contact.

2.1 *Sonority and the Syllable Contact Law*

The SCL is based on the concept that speech sounds can be classified into different categories according to their relative sonority. There have been many proposals for the definition of sonority. For example, Ladefoged (1982) defines sonority as the loudness of a sound relative to that of other sounds with the same length, stress, and pitch, and Clements (1990) defines it in terms of a set of major class features. In Venneman (1988), sonority is described as an inverse restatement of strength, which is based on "degree of deviation from unimpeded (voiced) air flow." In addition, Parker (2002, 2008) claims that sonority has a phonetic basis in intensity measured as "sound level differences in decibels between a target segment and a constant reference segment in the environment."

There exist many competing sonority scales, obtained on the basis of different definitions of the sonority. Different sonority scales result from the controversial relative sonority of laterals and rhotics, voiced and voiceless obstruents, stops, fricatives, and affricates, and high, mid, and low vowels. However, it is generally agreed among most researchers that the uncontroversial sonority hierarchy is as follows (Bell and Hooper 1978; Harris 1983; van der Hulst 1984; Clements 1987, 1990; Kenstowicz 1994; Smolensky 1995; Holt 1997; van Oostendorp 1999):

(1) *Sonority hierarchy*

more sonorous
↑ vowels
glides
liquids
nasals
↓ obstruents
less sonorous

For extensive discussion of sonority and sonority scales, see CHAPTER 49: SONORITY.

With reference to the relative sonority of segments, Hooper (1976) proposes a constraint on heterosyllabic coda-onset clusters. According to the principle, the sonority of a coda consonant must exceed that of a following onset consonant. Murray and Vennemann (1983) and Vennemann (1988) extend the principle as follows:

(2) *Syllable Contact Law* (Vennemann 1988: 40)

A syllable contact A.B is the more preferred, the less the consonantal strength of the offset A and the greater the consonantal strength of the onset B.

The SCL can be rephrased as in (3), using the concept of sonority, which is the reverse of strength and more commonly used in current phonology.

(3) *Syllable Contact Law (sonority version)* (Davis and Shin 1999: 286)

A syllable contact A.B is the more preferred, the greater the sonority of the offset A and the less the sonority of the onset B.

The SCL has also been invoked as a family of related OT constraints (Bat-El 1996; Davis 1998; Ham 1998; Davis and Shin 1999; Rose 2000; Baertsch 2002; Gouskova 2004; Holt 2004; Zec 2007).

2.2 *Phonological change as optimization of syllable contact*

2.2.1 *Diachronic change*

The notion of syllable contact has been employed in motivating various diachronic changes attested in coda–onset sequences (see also CHAPTER 93: SOUND CHANGE). For example, Hooper (1976) claims that various phonological processes have applied to the /nr/ sequence diachronically in different Spanish dialects to remedy the sequence violating the SCL, as illustrated below:

(4) *venirá* > *venrá* > *vendrá*	(by stop insertion)	'(it) will come'
verná	(by metathesis)	
verrá	(by assimilation)	

In the example above, after the occurrence of syncope, the resulting form *venrá* is unacceptable since the onset /r/ in the /nr/ sequence is more sonorous than the coda /n/, thus violating the SCL. Therefore, phonological processes such as stop insertion, metathesis, and assimilation apply in different Spanish dialects to improve the less preferred heterosyllabic /nr/ sequence.

Stop insertion also applied to various heterosyllabic consonant–liquid clusters in Old Spanish and Old French, as shown below.

(5) a. *Stop insertion in Old Spanish* (Martínez-Gil 2003)

Latin	Old Spanish	
fem(i)na	*(f)embra*	'female'
hum(e)ru	*hombro*	'shoulder'
trem(u)lare	*temblar*	'to shake, shiver'
ingen(e)rare	*engendrar*	'to engender, beget'
mel(io)rare	*me(l)drar*	'to grow'

b. *Stop insertion in Old French* (Walker 1978; Morin 1980; Wetzels 1985; Picard 1987; Martínez-Gil 2003)

Latin			Old French	
cam(e)ra	(> *chamre*)	>	*chambre*	'room'
sim(u)lare	(> *semler*)	>	*sembler*	'to resemble'
ten(e)ru	(> *tenre*)	>	*tendre*	'tender'
mol(e)re	(> *molre*)	>	*molder*	'to grind'
laz(a)ru	(> *lazre*)	>	*la(z)dre*	'beggar'
ess(e)re	(> *esre*)	>	*estre*	'to be'
spin(u)la	(> *espinle*)	>	*espingle*	'pin'

According to Martínez-Gil (2003), all the heterosyllabic consonant–liquid clusters in the examples above violate the SCL by showing more sonorous onsets than codas, and thus the stop is inserted within the clusters as a strategy to improve bad syllable contact.

Holt (2004) analyzes the metathesis observed in the heterosyllabic /dn/ and /dl/ sequences of Old Spanish as a repair strategy to optimize syllable contact. In Old Spanish, the heterosyllabic /dn/ and /dl/ sequences brought about by syncope in Late Spoken Latin or by morpheme concatenation underwent many repair strategies, such as dissimilation, palatalization, stop insertion, deletion, and strengthening, e.g. *antenatu* (Latin) > *adnado* ~ *andado* ~ *andrado* ~ *alnado* ~ *anado* ~ *annado* (Old Spanish) 'stepchild'. The examples in (6) illustrate that metathesis is optional in Old Spanish forms, while only metathesized forms occur in Modern Spanish.

(6) a. /dn/

Latin	Old Spanish	Modern Spanish	
CAT(E)NATU	*cadnado* ~ *candado*	*candado*	'padlock'
ANTENATU	*adnado* ~ *andado*	*andado*	'stepchild'
LEGITIMU	*lidmo* ~ *lindo*	*lindo*	'pretty'
RETINA	*riedna* ~ *rienda*	*rienda*	'rein'

b. /dl/

Latin	Old Spanish	Modern Spanish	
SPATULA	*espadla* ~ *espalda*	*espalda*	'back'
CAPITULU	*cabidlo* ~ *cabildo*	*cabildo*	'town council'
FOLIATILE	*hojalde* ~ *hojaldre*	*hojaldre*	'puff pastry'
TITULO	*tidle* ~ *tilde*	*tilde*	'written accent'

According to Holt, metathesis applied to /dn/ and /dl/ in Old Spanish, since the sequences have rising sonority over a syllable boundary and thus show bad syllable contact. (See also CHAPTER 59: METATHESIS for a discussion of synchronic metathesis processes in response to phonotactic requirements more generally.)

In addition, West Germanic word-medial gemination after a short vowel and before *j, *r, and *l has been analyzed by Murray and Vennemann (1983) as being driven by the need to avoid an onset which is more sonorous than the coda.

(7) *Gemination in West Germanic*
(Murray and Vennemann 1983; Braune and Eggers 1987; Ham 1998)[1]

	East Germanic	West Germanic	
a.	Go. *skapjan*	OS *skeppian*, OE *scieppan*	'to create'
	Go. *bidjan*	OS *biddian*, OE *biddan*	'to ask for'
	Go. *hafjan*	OHG *heffan*, OE *hebban*	'to lift'
	ON *framja*	OHG *fremmen*, OE *fremman*	'to carry out (a task)'
	Go. *halja*	OHG *hella*, OS *hellia*	'hell'
b.	ON *bitr*	OHG *bittar* (<**bittr*)	'bitter'
	Go. *akrs*	OHG *ackar*, OS *akkar* (<**akr*)	'field'
	VL *facla*	OHG *faccala* (<**faccla*)	'torch'
	ON *epli*	OE *æppel*, OS *appel* (<**appl*)	'apple'

According to Murray and Vennemann, Proto-Germanic had stem-initial accent, suggesting that the Proto-Germanic forms of the examples in (7) were syllabified as *VC.jV, *VC.rV, and *VC.lV in order to maintain a heavy stressed syllable. Murray

[1] Go. = Gothic; OS = Old Saxon; OE = Old English; OHG = Old High German; ON = Old Norse; VL = Vulgar Latin.

and Vennemann argue that West Germanic gemination was triggered by the SCL; that is, by gemination, coda–onset sequences with rising sonority across a syllable boundary (i.e. VC.jV, VC.rV, and VC.lV) were improved to ones with a sonority plateau (i.e. VC.CjV, VC.CrV, and VC.ClV). Ham (1998) also motivates gemination before *r and *l in West Germanic by appeal to the concept of syllable contact, arguing that gemination before *j resulted from the well-attested Germanic avoidance of *j in onset-initial position.

2.2.2 *Synchronic change*

Syllable contact has also been considered as a factor motivating different types of synchronic sound alternations, such as assimilation, metathesis, glide formation, and epenthesis in heterosyllabic coda–onset clusters. For example, Rice and Avery (1991), Rice (1992), Iverson and Sohn (1994), and Davis and Shin (1999) view the assimilation observed in Korean nasal + liquid sequences as being due to the SCL. In Korean, two types of assimilation processes are attested when a liquid is preceded by a nasal. When /n/ is followed by a liquid, the nasal assimilates to the following liquid, resulting in [ll], as the examples in (8a) illustrate. On the other hand, when a liquid is preceded by a non-coronal nasal /m/ or /ŋ/, the liquid assimilates to the previous nasal segment in nasality and thus the /ml/ and /ŋl/ sequences are realized as [mn] and [ŋn], respectively, as shown in (8b).

(8) a. *Assimilation in /nl/*

/non+li/	[nolli]	'logic'
/han+ljaŋ/	[halljaŋ]	'limit'
/cʰɔn+li/	[cʰəlli]	'natural law'

b. *Assimilation in /ml/ and /ŋl/*

/sam+lju/	[samnju]	'third rate'
/tam+ljək/	[tamnjək]	'courage'
/jəŋ+li/	[jəŋni]	'profit'
/saŋ+lju/	[saŋnju]	'the upper stream'

According to Iverson and Sohn (1994) and Davis and Shin (1999), the two types of sonorant assimilations occur in a nasal + liquid sequence in Korean because the sequence is unacceptable due to violating the SCL.

Bradley (2007) analyzes obstruent–nasal metathesis in Sidamo as occurring in order to avoid bad syllable contact.

(9) *Obstruent–nasal metathesis in Sidamo* (Hudson 1975, 1995; Hume 1999)

/hab+nemmo/	[hambemmo]	'we forget'
/gud+nonni/	[gundonni]	'they finished'
/it+noommo/	[intoommo]	'we have eaten'
/has+nemmo/	[hansemmo]	'we look for'
/duk+nanni/	[duŋkanni]	'they carry'
/ag+no/	[aŋgo]	'let's drink'

As can be seen from the examples in (9), a stem-final obstruent metathesizes with a following suffix-initial /n/ in Sidamo and then the nasal assimilates regressively in place. According to Bradley, metathesis applies to obstruent–nasal clusters with rising sonority to produce nasal–obstruent clusters, which do not violate the SCL.[2]

In addition, Bradley (2007) accounts for the metathesis of /nr/ clusters in irregular future and conditional forms of Judeo-Spanish verbs by employing the concept of syllable contact.

(10) *Metathesis of /nr/ in Judeo-Spanish*
(Lamouche 1907; Baruch 1930; Bradley 2007)

terné	'I will have'	cf. tener	'to have'
verné	'I will come'	cf. venire	'to come'

Glide formation in German is also motivated by the SCL. According to Hall (2007), [i] and [j] are in complementary distribution in German and /i/ is realized as [j] when occurring to the left of a vowel as in *labial* [la'bjaːl] 'labial'. In /VCCiV/ sequences, glide formation applies if the two consonants show a sonority fall, as in (11a), while it does not apply if the two consonants show a sonority rise, as in (11b).

(11) a. *Glide formation*

liquid + obstruent	*Skorpion*	[skɔʀpjoːn]	'scorpion'
	Celsius	[tsɛlzjus]	'celsius'
nasal + obstruent	*Indien*	[ɪndjən]	'India'
liquid + nasal	*Kalifornien*	[kalifɔʀnjən]	'California'
	Fermium	[fɛʀmjom]	'fermium'

b. *No glide formation*

obstruent + liquid	*Atrium*	[aːtʀium]	'atrium'
	Bibliothek	[biblioteːk]	'library'
obstruent + nasal	*Hafnium*	[hafniom]	'hafnium'

The notion of syllable contact has also been employed in explaining vowel epenthesis patterns in loanwords. When CVC languages borrow loanwords with complex onsets, a vowel is inserted and, according to Gouskova (2001), the position of the inserted vowel is determined in consideration of syllable contact.

(12) *Epenthesis patterns in loanwords* (Gouskova 2001)

a. *Falling or flat sonority: Edge epenthesis*

		gloss/source
Hindi	ɪskul	'school'
	isfɪər	'sphere'
Bengali	iʃkul	'school'
Central Pahari	ispiitʃ	'speech'
Sinhalese	istri	Sanskrit stri 'woman'
Wolof	estati	'statue'
Uyghur	istatistika	Russian statistika 'statistics'

[2] Hume (1999) accounts for obstruent–nasal metathesis in Sidamo in terms of the perceptibility of the two consonants involved. According to her, metathesis in Sidamo occurs because having distinctive place in both nasals and unreleased stops in pre-consonantal position is perceptually disfavored.

b. *Rising sonority: Internal epenthesis*

		gloss/source
Hindi	fɪrut	'fruit'
	pəfaizər	'Pfizer'
Bengali	gelaʃ	'glass'
Central Pahari	silet	'slate'
Sinhalese	tijage	Sanskrit tjage 'gift'
Wolof	kalas	'class'
Uyghur	kulub	Russian klub 'club'

As shown in (12a), a vowel is inserted before the onset consonants with falling or flat sonority, since the resulting coda–onset clusters do not violate the SCL. However, as illustrated in (12b), a vowel is inserted between the two consonants of the onset with rising sonority to avoid the coda-onset clusters violating the SCL.

According to Rose (2000), syllable contact plays a role in determining the position of an epenthetic vowel in Chaha (see CHAPTER 67: VOWEL EPENTHESIS). For example, when [i] is epenthesized in /VCCCV/ for structural reasons, either [VCCɨCV] or [VCɨCCV] can surface with the aim of achieving good syllable contact. Thus, /t-n-k'rət'm-nə/ 'while we are cutting' surfaces as [tink'irət'imnə], since the alternative form *[tɨnɨk'rət'ɨmnə] has the coda–onset cluster [k'.r], with bad syllable contact. On the other hand, /t-n-msəkr-nə/ 'while we are testifying' is realized as [tinimsəkinnə] rather than *[tinmɨsəkinnə], with a coda-onset cluster with flat sonority, which is avoided in the language.

In addition, epenthesis in Picard clitics (Auger 2003) is said to be influenced by the preference for achieving good syllable contact, and Pons (2005) argues that regressive manner assimilation, rhotacism, gliding, onset strengthening, epenthesis, and deletion attested in Romance languages are strategies triggered to avoid bad syllable contact.

3 Issues and debates on syllable contact

3.1 *Sonority Dispersion Principle and the Syllable Contact Law*

As seen in §2.2, the notion of syllable contact has been employed in motivating different types of diachronic and synchronic sound changes. However, there have been debates regarding whether or not it is necessary to posit a separate law just for coda-onset sequences. Clements (1990) claims that the Sonority Dispersion Principle makes the SCL dispensable. According to the Sonority Dispersion Principle, sonority rise is required to be maximal from the onset to the nucleus and sonority drop is required to be minimal from the nucleus to the coda. Thus, for example, the principle says that [ta] is more preferred than [la] as the onset and [al] is more preferred than [at] as the coda. From the Sonority Dispersion Principle, it can be predicted that maximal sonority drop is preferred across syllable boundaries, as stated in the SCL. Thus, Clements proposes that the SCL follows from the more general Sonority Dispersion Principle.

However, based on the data from Kazakh in (13), Davis (1998) and Gouskova (2004) argue that the SCL cannot be reduced to the Sonority Dispersion Principle.

(13) *Syllable contact in Kazakh* (Davis 1998)

a. *No onset desonorization*

/alma-lar/	[al.ma.lar]	'apples'
/mandaj-lar/	[man.daj.lar]	'foreheads'
/kijar-lar/	[kijar.lar]	'cucumbers'
/kol-ma/	[kol.ma]	'hand-INTERROG'
/kijar-ma/	[ki.jar.ma]	'cucumber-INTERROG'

b. *Onset desonorization*

/kol-lar/	[kol.dar]	'hands'
/murin-lar/	[mu.rin.dar]	'noses'
/koŋɯz-lar/	[ko.ŋɯz.dar]	'bugs'
/murin-ma/	[mu.rin.ba]	'nose-INTERROG'
/koŋɯz-ma/	[ko.ŋɯz.ba]	'bug-INTERROG'

In Kazakh, consonants of any sonority can be onsets if they are preceded by vowels or by consonants of higher sonority. Thus, no onset desonorization is attested in the examples of (13a). However, when the onset is preceded by a consonant with lower or the same sonority, it desonorizes, as shown in (13b). Therefore, to explain the sound alternations of onsets in the Kazakh examples, it is essential to refer to the sonority relation of the coda and the following onset. For this reason, both Davis and Gouskova claim that the Sonority Dispersion Principle cannot completely replace the SCL.

3.2 Nature of syllable contact: Categorical vs. gradient

Since Hooper (1976), the SCL has been extended to account for more cross-linguistic data, and the nature of syllable contact is now viewed as gradient rather than categorical (see CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY). Hooper (1976) originally proposed the SCL for Spanish, where the sonority of a syllable-final consonant must exceed that of a following syllable-initial consonant. According to this, the syllable contact is categorical in nature. Different types of coda–onset clusters are equally fine as long as the sonority of the coda exceeds that of the onset. Thus, although the sonority distance between the coda and the onset is greater in *al.ta* than in *al.na*, both are equally fine clusters with respect to syllable contact since the coda is more sonorous than the onset. Likewise, both *at.la* and *an.la* are equally non-optimal since the coda is less sonorous than the onset.

Murray and Vennemann (1983) and Vennemann (1988) propose an extended version of the SCL whose nature is gradient: two adjacent heterosyllabic segments are more preferred, the greater the sonority of the first segment and the less the sonority of the second segment. Clements (1990) paraphrases the extended SCL as follows:

(14) *The extended Syllable Contact Law* (Clements 1990: 319)

The preference for a syllabic structure A.B, where A and B are segments and *a* and *b* are the sonority values of A and B respectively, increases with the value of *a* minus *b*.

Davis and Shin's (1999) version of the SCL in (3) is also a gradient one. In the gradient view of the syllable contact phenomenon, coda-onset clusters with falling sonority are not considered to exhibit the same degree of optimality. In addition, not all violations of the SCL by clusters with rising sonority are considered equally serious. Depending on the relative sonority distance between the coda and the onset, one cluster is more optimal or less non-optimal than the others. For example, *al.ta* is more optimal than *al.na*, since the sonority distance between the coda and onset is greater in *al.ta* than in *al.na*. In addition, a sequence such as *at.la* constitutes a more severe violation of the law than a sequence such as *an.la*.

In line with this, Clements (1990) provides the following aggregate complexity score of each of the coda–onset sequences by summing the complexity values of each of the demisyllables (nucleus-coda and onset–nucleus sequences) that constitute it.

(15)

C1 \ C2	obstruent	nasal	liquid	glide
obstruent	5	6	7	8
nasal	4	5	6	7
liquid	3	4	5	6
glide	2	3	4	5

According to Clements (1990), obstruent–glide sequences such as *at.wa* violate the SCL most seriously and glide–obstruent sequences such as *aw.ta* satisfy the law most faithfully.

A more fine-grained syllable contact scale is proposed by Gouskova (2004), based on Jespersen's (1904) sonority scale.

(16) a. *Sonority scale* (Jespersen 1904)
glides > rhotics > laterals > nasals > voiced fricatives > voiced stops > voiceless fricatives > voiceless stops
(abbreviated as: w > r > l > n > z > d > s > t)

b. *Syllable contact scale* (Gouskova 2004)
more harmonic ←——————→ less harmonic

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
w.t	w.s	w.d	w.z	w.n	w.l	w.r	w.w	r.w	l.w	n.w	z.w	d.w	s.w	t.w
	r.t	r.s	r.d	r.z	r.n	r.l	r.r	l.r	n.r	z.r	d.r	s.r	t.r	
		l.t	l.s	l.d	l.z	l.n	l.l	n.l	z.l	d.l	s.l	t.l		
			n.t	n.s	n.d	n.z	n.n	z.n	d.n	s.n	t.n			
				z.t	z.s	z.d	z.z	d.z	s.z	t.z				
					d.t	d.s	d.d	s.d	t.d					
						s.t	s.s	t.s						
							t.t							
–7	–6	–5	–4	–3	–2	–1	0	+1	+2	+3	+4	+5	+6	+7

←—— drop ——————— flat ——————— rise ——→

In Gouskova's syllable contact scale in (16b), glide + voiceless stop clusters (represented as *w.t*), for example, are characterized as sequences with a sonority drop of –7, since the sonority of the voiceless stop is lower than that of the glide by seven steps according to Jespersen's sonority scale in (16a). As can be seen in (16b), Gouskova proposes that the syllable contact scale has 15 strata, with glide + voiceless stop sequences (represented as *w.t*) on stratum 1 the most harmonic, and voiceless stop + glide sequences (represented as *t.w*) on stratum 15 the least harmonic.

Clements and Gouskova employ different sonority hierarchies, and thus propose different gradient versions of the sonority contact scale. Gouskova's more fine-grained syllable contact scale can account for the different patterning of /ll/ and /rl/ in Kazakh, while this is not possible within Clements' syllable contact scale. According to Davis (1998), coda-onset sequences with flat or rising sonority undergo desonorization in Kazakh in order to improve syllable contact, as in /murin-lar/ → [murindar] 'noses'. On the other hand, no desonorization applies to coda-onset sequences with falling sonority, as can be seen from /mandaj-ga/ → [mandajga] 'forehead + DIRECT'. In the case of /ll/ with flat sonority, desonorization applies, as in /kol-lar/ → [kol.dar] 'hands', while /rl/ surfaces as [rl], as in /kijar-lar/ → [kijarlar] 'cucumbers'. The different phonological patterning of /ll/ and /rl/ in Kazakh suggests that /r/ is more sonorous than /l/, and that /rl/ is the sequence with falling sonority, as in Gouskova's syllable contact scale. Note that Clements' syllable contact scale cannot account for the different patterning of /ll/ and /rl/ in Kazakh, since /r/ is assumed to be as sonorous as /l/.

The gradient nature of syllable contact makes it possible to explain language-specific patterns of the syllable contact phenomenon. According to Gouskova (2004), languages differ in the level of complexity they tolerate, and thus language-specific patterns of syllable contact are attested. For example, Kirghiz and Kazakh require sonority to drop from the coda to the onset but the two languages are different in the thresholds of acceptable sonority drop. In Kazakh, sonority is merely required to drop from the coda to the onset and need not be maximal. Thus, as can be seen in (13b), onset desonorization applies to coda–onset sequences with flat or rising sonority, but not to coda–onset sequences with falling sonority, as in (13a). On the other hand, as can be seen from the examples in (17), sonority drop alone is not sufficient in Kirghiz, which requires maximal sonority drop.

(17) *Syllable contact in Kirghiz*
(Hebert and Poppe 1964; Kasymova *et al.* 1991; Gouskova 2004)

/konok-lar/	+5	konok.tar	0	'guest (PL)'
/taʃ-lar/	+4	taʃ.tar	−1	'stone (PL)'
/konok-nu/	+4	konok.tu	0	'guest (OBJ)'
/taʃ-nu/	+3	taʃ.tɯ	−1	'stone (OBJ)'
/atan-lar/	+1	atan.dar	−2	'gelded camel (PL)'
/rol-lar/	0	rol.dar	−3	'role (PL)'
/atan-nu/	0	atan.dɯ	−2	'gelded camel (OBJ)'
/kar-lar/	−1	kar.dar	−4	'snow (PL)'
/rol-nu/	−1	rol.du	−3	'role (OBJ)'
/aj-lar/	−2	aj.dar	−5	'moon (PL)'
/aj-nu/	−3	aj.dɯ	−5	'moon (OBJ)'
/kar-du/	−4	kar.dɯ	−4	'snow (OBJ)'
/too-lar/	—	too.lar	—	'mountain (PL)'
/too-nu/	—	too.nu	—	'mountain (OBJ)'

As can be seen from the examples above, desonorization in Kirghiz applies to suffix-initial sonorants after any consonant, in order for sonority to drop maximally from the coda to the onset. Note that the Kazakh and Kirghiz data above cannot be explained properly within the categorical version of the SCL, which predicts that all coda-onset clusters with falling sonority are equally acceptable.

3.3 *Syllable contact as a language-specific constraint on minimal sonority distance*

With respect to syllable contact phenomena, language-specific variations can be observed. First, some languages respecting the SCL allow coda-onset clusters with flat sonority while others disallow them. For example, as shown in (8), Korean prohibits coda-onset clusters with rising sonority. However, a coda can be as sonorous as a following onset, as can be seen from /sallim/ → [sallim] 'housekeeping', /simmun/ → [simmun] 'interrogation', etc. On the other hand, as shown in (13), coda–onset clusters with flat sonority as well as rising sonority are not sanctioned in Kazakh. Thus, /kol-lar/ is realized as [kol.dar] 'hands', to improve syllable contact.

Such syllable contact variation of coda–onset sequences with equal sonority has been accounted for by positing two versions of the SCL.

(18) a. *Syllable Contact Law (strict)* (Rose 2000: 401)
The first segment of the onset of a syllable must be lower in sonority than the last segment in the immediately preceding syllable.

b. *Syllable Contact Law (loose)* (Bat-El 1996; Davis and Shin 1999)
The first segment of the onset of a syllable must not be of greater sonority than the last segment of the immediately preceding syllable.

Coda–onset sequences of equal sonority are not allowed within the SCL in (18a), while they are permitted within the loose version of the SCL in (18b).

Syllable contact variations that cannot be explained by the two versions of the SCL are also found. The SCL has generally been employed to constrain coda–onset clusters with rising sonority. However, languages which allow only falling sonority from a coda to a following onset can vary with respect to minimal sonority distance. According to Gouskova (2004), both Kirghiz and Sidamo allow coda–onset sequences with falling sonority, forbidding sequences with flat or rising sonority. However, the thresholds of acceptable sonority drop are different in the two languages. As can be seen from the examples in (17), minimal sonority drop required in Kirghiz is −4, and coda–onset sequences with sonority drop of less than −4 undergo phonological change of desonorization to remedy bad syllable contact. Thus, desonorization applies to /aj-nu/ 'moon (OBJ)', which is realized as [aj.dɯ], with a sonority drop of −3. On the other hand, /kar-du/ 'snow (OBJ)', with a sonority drop of −4, is not targeted by any phonological process in order to remedy bad syllable contact, and surfaces as [kar.dɯ]. In the case of Sidamo, unlike Kirghiz, the acceptable minimal sonority drop is −2.

(19) *Syllable contact in Sidamo* (Moreno 1940; Gouskova 2004)

a. *Sonority rises: Metathesis*

/duk-nanni/	+4	duŋ.kanni	−4	'they carry'
/huʧ-nanni/	+4	hun.ʧanni	−4	'they pray/beg/request'
/has-nemmo/	+3	han.seemo	−3	'we look for'
/hab-nemmo/	+2	ham.bemmo	−2	'we forget'

b. *Sonority drops less than −2 or is flat: Gemination*

/af-tinonni/	−1	affinonni	—	'you (PL) have seen'
/lelliʃ-toti/	−1	lelliʃʃoti	—	'don't show!'
/ful-nemmo/	−1	fullemmo	—	'we go out'
/um-nommo/	[illegible]	ummommo	—	'we have dug'

c. *Sonority drops more than −2: Place assimilation only*

/mar-toti/	−5	mar.toti	−5	'don't go'
/ful-te/	−5	ful.te	−5	'your having gone out'
/qaram-tino/	−4	qaran.tino	−4	'she worried'

In Sidamo, coda-onset clusters with rising sonority are repaired by metathesis, as in (19a), and the clusters with flat sonority or sonority drop of less than −2 are modified by gemination, as in (19b). On the other hand, when coda–onset sequences have a sonority drop of −2 or above, no modification applies to improve syllable contact, although the sequences undergo place assimilation, as in (19c).

In addition, different languages can permit different degrees of minimal sonority rise, which cannot be explained within the SCL in (18). For example, Faroese and Icelandic set different acceptable sonority distances between a coda and a following onset. According to Gouskova (2004), Faroese permits a sonority rise of +4 or below, as illustrated in (20).

(20) *Syllable contact in Faroese* (Gouskova 2004)

a. *Sonority rise of 5 points or more: Complex onsets*

aː.ʰkvamarɪn	+7	'beryl'
vɛaː.ʰkrɪr	+6	'beautiful (MASC PL)'
aiː.ʰtrantɪ	+6	'poisonous'
d̥ɛaː.ʰprɪr	+6	'sad'
miː.ʰkhr	+5	'great (MASC PL)'
eː.ʰplɪ	+5	'potato'

b. *Sonority rise of fewer than 5 points: Heterosyllabic coda–onset clusters*

sɪg̊.rɪ	+4	'further south'
b̥aʰt.na	+4	'to improve'
ɪd̥.la	+3	'or'
vɛs.na	+3	'to worsen'
ʝar.na	+2	'gladly'
rɔʰk.tɪ	0	'smoked (SG)'
vɛs.tur	−1	'west'
hɛn.d̥ur	−2	'hands'
ʝœr.d̥ɪ	−4	'did (SG)'
nod̥.d̥ɪ	—	'approached (SG)'

In Faroese, where initial syllables are always stressed and heavy, syllable contact plays a crucial role in syllabification. As can be seen in (20a), when two intervocalic consonants show a sonority rise of +5 or more, the two consonants are syllabified into a complex onset and the preceding vowel is long. On the other hand, when two intervocalic consonants exhibit a sonority rise of +4 or less, the two consonants are syllabified as heterosyllabic and the preceding vowel is short, as in (20b). Thus, the examples in (20) show that the maximal sonority rise from a coda to a following onset permitted in Faroese is +4 or less.

In Icelandic, unlike Faroese, the acceptable sonority distance between a coda and a following onset is +5 or less.

(21) *Syllable contact in Icelandic* (Gouskova 2004)

a. *Sonority rise of 6 points or more: Complex onsets*

vɪː.t⁽ʰ⁾ja	+7	'to visit'	skɔː.p⁽ʰ⁾ra	+6	'to roll'
vœː.k⁽ʰ⁾va	+7	'to water'	tvɪː.svar	+6	'twice'
aː.k⁽ʰ⁾rar	+6	'fields'	ɛː.sja	+6	'the mountain Esja'
tʰɪː.t⁽ʰ⁾ra	+6	'to vibrate'			

b. *Sonority rise of fewer than 6 points: Heterosyllabic coda–onset clusters*

ɛʰp.lɪ	+5	'apple'	tʰɛm.ja	+3	'to domesticate'
ɛʰk.la	+5	'lack'	vɛl.ja	+2	'to choose'
ăɪʰt.la	+5	'to intend'	vɛr.ja	+1	'to defend'
b̥ɪð.ra	+4	'to ask'	tʰɛv.ja	0	'to delay'
stœð.va	+4	'to stop'	hɛs.tʏr	−1	'horse'
hăiɣ.rɪ	+3	'right'	ɛv.rɪ	−1	'upper'
b̥lað.ra	+3	'balloon'	av.laɣa	−2	'to bend out of shape'
sɪg̊.la	+3	'to sail'	d̥vɛr.g̊ʏr	−4	'dwarf'
vɪs.na	+3	'to wither'			

As can be seen from the examples in (21a), two consonants with a sonority rise of +6 or more are syllabified as a complex onset and the preceding vowel is lengthened to make the first stressed syllable heavy. On the other hand, when two consonants show a sonority rise of +5 or less, as in (21b), they are syllabified as heterosyllabic, and the preceding vowel is realized as a short vowel since the first stressed syllable is heavy, due to the coda.

According to Gouskova (2004), the language-specific patterns of syllable contact discussed above suggest that languages differ in the thresholds of acceptable sonority distance between a coda and a following onset and that the SCL is not a single constraint forbidding rising sonority. To account for such language variations within Optimality Theory, Gouskova reformulates the SCL as a hierarchy of constraints derived from sonority scales, which target negative, flat, or positive sonority distance across a syllable boundary.

3.4 *Problems with the Syllable Contact Law*

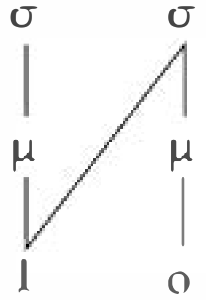
Seo (2003) points out that the SCL has two problems. First of all, the SCL is problematic, since it cannot explain why a consonant + liquid cluster is targeted by a phonological process when a syllable boundary cannot be assumed between two segments in the cluster, for example, in word-initial position. Recall that the SCL crucially refers to a syllable boundary in explaining phonological change of a consonant + liquid cluster by assuming that the change is motivated to avoid cases where the onset B has higher sonority than the coda A in a heterosyllabic A.B sequence.

As illustrated below, in Leti an /nl/ sequence is realized as [ll], whether it occurs word-initially or intervocalically. Note that vowel deletion occurs in the examples given.

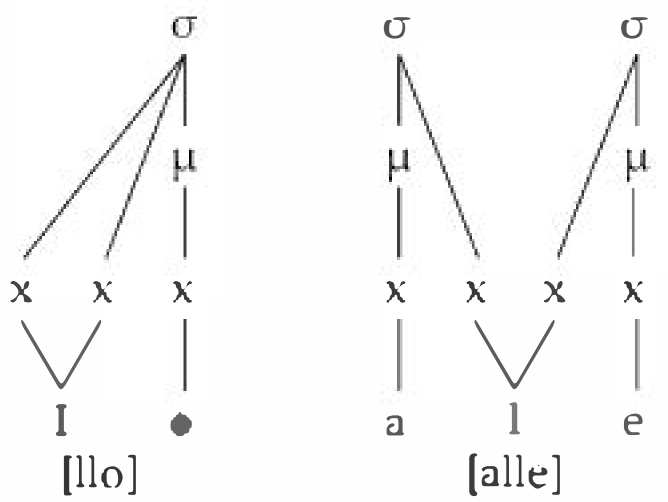
(22) *Leti* (van Engelenhoven 1995; Hume *et al.* 1997)

/na+losir/	→	[llosir]	'3SG-to follow'
/a:na+leti/	→	[a:lleti]	'Alety (clan)' (child + Leti)

In line with the view that the change of a consonant + liquid cluster is motivated due to the SCL, the word-initial geminate [ll] in [llosir] might be considered to be bisyllabic, as represented below (see also CHAPTER 37: GEMINATES):

(23)


However, as discussed in Hume *et al.* (1997), there is a problem with the representation in (23). Leti has a minimal word requirement that lexical words must be minimally bimoraic. With the representation of the geminate in (23), the minimal word requirement is satisfied in words consisting of an initial geminate and a vowel, for example [p.pe]. Thus, it is expected that there will exist words with such a structure in Leti. However, none are attested. On the other hand, if word-initial geminates are assumed to be part of the onset of a single syllable, words containing initial geminates such as [ppu.na] 'nest' are bisyllabic, conforming to the dominant tendency that lexical words are made up of two syllables in Leti. Thus, Hume *et al.* (1997) propose that geminates in Leti have phonological structure with a root node multiply linked to two timing slots, as in (24).

(24) *Geminates and long vowels in Leti* (Hume *et al.* 1997)


In view of the minimal word requirement, it cannot be assumed that an initial geminate is bisyllabic in Leti. Therefore, the modification of the word-initial tautosyllabic /nl/ sequence in Leti cannot be motivated by the SCL since a syllable boundary cannot be referred to. On the other hand, phonological change of the intervocalic /nl/ sequence in Leti could be accounted for by relying on the avoidance of rising sonority over a syllable boundary. Thus, even though both heterosyllabic and tautosyllabic /nl/ sequences show the same pattern in Leti, the syllable contact account cannot provide a unified account of the modification.

The SCL is also argued to be problematic in Seo (2003), since it cannot provide a unified account of the same types of phonological changes found in nasal + liquid and liquid + nasal sequences in Korean.

(25) *Modifications of nasal/liquid sequences in Korean* (Davis and Shin 1999)

/nl/	→	[ll]	/non-li/	→	[nolli]	'logic'
/ln/	→	[ll]	/səl+nal/	→	[səllal]	'New Year's Day'
cf. /lm/	→	[lm]	/pul+mjən/	→	[pulmjən]	'insomnia'

As illustrated in (25), in Korean /nl/ surfaces as [ll]. According to the SCL, this change is expected since the sequence violates the law. However, the same change applies to /ln/, while /lm/ (with the same sonority distance as /ln/) does not undergo any change. Thus it must be assumed within the syllable contact account that the modification of a nasal + liquid sequence and that of a liquid + nasal sequence result from different factors, although the same types of phonological changes occur in both types of sequence. Korean is not the only language that has the same types of alternations in nasal/liquid sequences. As shown in (26), the same type of phonological change applies to both /nl/ and /ln/ sequences in Leti, Toba Batak, and Boraana Oromo.

(26) *Modifications of nasal/liquid sequences*

a. *Leti* (van Engelenhoven 1995; Hume *et al.* 1997)

/nl/	→	[ll]	/aːna+leti/	→	[aːlleti]	'Alety (clan)'
/ln/	→	[ll]	/vulan/	→	[vulla]	'moon'

b. *Toba Batak* (Nababan 1981)

/nl/	→	[ll]	/laɔn+laɔn/	→	[laɔllaɔn]	'eventually'
/ln/	→	[ll]	/bal+na/	→	[balla]	'his ball'

c. *Boraana Oromo* (Stroomer 1995)

/nl/	→	[ll]	/hin+lool+a/	→	[hilloola]	'I/he will fight'
/ln/	→	[ll]	/kofl+na/	→	[kofalla]	'we smile'

Within the syllable contact account, the modification of /ln/ would have to be motivated by a factor other than syllable contact, although both /ln/ and /nl/ undergo the same type of phonological change.

Seo (2003) proposes that the phonological processes found in phonological modifications of heterosyllabic coda-onset sequences could be viewed as resulting from a segment contact phenomenon which is closely related to speech perception, rather than from the SCL (see CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). According to Seo, contrasts of weak perceptibility triggered by phonetic similarity between two members of a cluster are a key factor in motivating the alternations in the cluster. Thus it is expected that phonological modifications will apply when two consonants are perceptually similar to each other, and when they occur in a sequence, regardless of the presence or absence of a syllable boundary and the order of the two consonants.

4 Conclusion

In this chapter, I have reviewed four major issues concerning the SCL. First, different opinions regarding the necessity of the SCL were discussed – while Clements (1990) claims that the SCL is not required with the Sonority Dispersion Principle, Davis (1998) and Gouskova (2004) argue that the SCL cannot be reduced to the Sonority Dispersion Principle based on the Kazakh data. Second, the change in viewpoint on the nature of syllable contact was reviewed, and it was shown that a gradient of syllable contact makes it possible to explain language-specific syllable contact patterns while a categorical view does not. Third, after providing examples illustrating cross-linguistic variations of syllable contact, I introduced Gouskova's (2004) proposal of the SCL as a language-specific constraint on minimal sonority distance. Finally, based on Seo (2003), after pointing out problems associated with the SCL, I argued that phonological modifications of heterosyllabic coda-onset sequences could be viewed as a segment contact phenomenon, which is closely related to speech perception.

REFERENCES

Auger, Julie. 2003. Pronominal clitics in Picard revisited. In Núñez-Cedeño *et al.* (2003), 3–20.

Baertsch, Karen. 2002. An optimality theoretic approach to syllable structure: The split margin hierarchy. Ph.D. dissertation, Indiana University.

Baruch, Kalmi. 1930. El judeo-español de Bosnia. *Revista de Filología Española* 17. 113–154.

Bat-El, Outi. 1996. Selecting the best of the worst: The grammar of Hebrew blends. *Phonology* 13. 283–328.

Bell, Alan & Joan B. Hooper. 1978. Issues and evidence in syllabic phonology. In Alan Bell & Joan B. Hooper (eds.) *Syllable and segments*, 3–22. Amsterdam: North-Holland.

Bradley, Travis G. 2007. Constraints on the metathesis of sonorant consonants in Judeo-Spanish. *Probus* 19. 171–207.

Braune, Wilhelm & Hans Eggers. 1987. *Althochdeutsche Grammatik*. 14th edn. Tübingen: Niemeyer.

Clements, G. N. 1987. Phonological feature representation and the description of intrusive stops. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 29–50.

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 283–333. Cambridge: Cambridge University Press.

Davis, Stuart. 1998. Syllable contact in Optimality Theory. *Korean Journal of Linguistics* 23. 181–211.

Davis, Stuart & Seung-Hoon Shin. 1999. The syllable contact constraint in Korean: An optimality-theoretic analysis. *Journal of East Asian Linguistics* 8. 285–312.

Engelenhoven, Aone van. 1995. A description of the Leti language as spoken in Tutukei. Ph.D. dissertation, University of Leiden.

Gouskova, Maria. 2001. Falling sonority onsets, loanwords and syllable contact. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 37. 175–186.

Gouskova, Maria. 2004. Relational hierarchies in Optimality Theory: The case of syllable contact. *Phonology* 21. 201–250.

Hall, T. A. 2007. German glide formation and its theoretical consequences. *The Linguistic Review* 24. 1–31.

Ham, William. 1998. A new approach to an old problem: Gemination and constraint reranking in West Germanic. *Journal of Comparative Germanic Linguistics* 1. 225–262.

Harris, James W. 1983. *Syllable structure and stress in Spanish: A nonlinear analysis.* Cambridge, MA: MIT Press.

Hebert, Raymond & Nicholas Poppe. 1964. *Kirghiz manual.* The Hague: Mouton.

Holt, D. Eric. 1997. The role of the listener in the historical phonology of Spanish and Portuguese: An optimality-theoretic account. Ph.D. dissertation, Georgetown University (ROA-278).

Holt, D. Eric. 2004. Optimization of syllable contact in Old Spanish via the sporadic sound change metathesis. *Probus* 16. 43–61.

Hooper, Joan B. 1976. *An introduction to natural generative phonology.* New York: Academic Press.

Hudson, Grover. 1975. Suppletion in the representation of alternations. Ph.D. dissertation, University of California, Los Angeles.

Hudson, Grover. 1995. Phonology of Ethiopian languages. In John A. Goldsmith (ed.) *The handbook of phonological theory,* 782–797. Cambridge, MA & Oxford: Blackwell.

Hulst, Harry van der. 1984. *Syllable structure and stress in Dutch.* Dordrecht: Foris.

Hume, Elizabeth. 1999. The role of perceptibility in consonant/consonant metathesis. *Proceedings of the West Coast Conference on Formal Linguistics* 17. 293–307.

Hume, Elizabeth, Jennifer Muller & Aone van Engelenhoven. 1997. Non-moraic geminates in Leti. *Phonology* 14. 371–402.

Iverson, Gregory K. & Hyang-Sook Sohn. 1994. Liquid representation in Korean. In Young-Key Kim-Renaud (ed.) *Theoretical issues in Korean linguistics,* 1–19. Stanford: CSLI.

Jespersen, Otto. 1904. *Lehrbuch der Phonetik.* Leipzig & Berlin: Teubner.

Kasymova, Bella, Kurmanbek Toktonalijev & Asan Karybajev. 1991. *Izucajem Kyrgyzskii jazyk.* Frunze: Mektep.

Kenstowicz, Michael. 1994. *Phonology in generative grammar.* Cambridge, MA & Oxford: Blackwell.

Ladefoged, Peter. 1982. *A course in phonetics.* 2nd edn. New York: Harcourt Brace Jovanovich.

Lamouche, L. 1907. Quelques mots sur le dialecte espagnol parlé par les Israélites de Salonique. *Romanische Forschungen* 23. 969–991.

Martínez-Gil, Fernando. 2003. Consonant intrusion in heterosyllabic consonant–liquid clusters in Old Spanish and Old French: An optimality theoretical account. In Núñez-Cedeño *et al.* (2003), 39–58.

Moreno, Martino Mario. 1940. *Manuale di Sidamo.* Milan: Mondadori.

Morin, Yves-Charles. 1980. Morphologisation de la épenthèse en ancient français. *Canadian Journal of Linguistics* 32. 365–375.

Murray, Robert W. & Theo Vennemann. 1983. Sound change and syllable structure in Germanic phonology. *Language* 59. 514–528.

Nababan, P. W. J. 1981. *A grammar of Toba Batak.* (*Pacific Linguistics* D37.) Canberra: Australian National University.

Núñez-Cedeño, Rafael, Luis López & Richard Cameron (eds.) 2003. *A Romance perspective on language knowledge and use.* Amsterdam & Philadelphia: John Benjamins.

Oostendorp, Marc van. 1999. Syllable structure in Esperanto as an instantiation of universal phonology. *Esperantologio/Esperanto Studies* 1. 52–80.

Parker, Steve. 2002. Quantifying the sonority hierarchy. Ph.D. dissertation, University of Massachusetts, Amherst.

Parker, Steve. 2008. Sound level protrusions as physical correlates of sonority. *Journal of Phonetics* 36. 55–90.

Picard, Marc. 1987. On the general properties of consonant epenthesis. *Canadian Journal of Linguistics* 32. 133–142.

Pons, Claudia. 2005. It is all downhill from here: The role of syllable contact in romance languages. Paper presented at the 13th Manchester Phonology Meeting (ROA-802).

Rice, Keren. 1992. On deriving sonority: A structural account of sonority relationships. *Phonology* 9. 61–99.

Rice, Keren & Peter Avery. 1991. On the relationship between laterality and coronality. In Carole Paradis and Jean-François Prunet (eds.) *The special status of coronals: Internal and external evidence*, 101–124. San Diego: Academic Press.

Rose, Sharon. 2000. Epenthesis positioning and syllable contact in Chaha. *Phonology* 17. 397–425.

Seo, Misun. 2003. A segment contact account of the patterning of sonorants in consonant clusters. Ph.D. dissertation, Ohio State University.

Smolensky, Paul. 1995. On the structure of the constraint component *Con* of UG. (ROA-86.)

Stroomer, Harry. 1995. *A grammar of Boraana Oromo (Kenya): Phonology, morphology, vocabularies*. Cologne: Rüdiger Köppe Verlag.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin: Mouton de Gruyter.

Walker, Douglas C. 1978. Epenthesis in Old French. *Canadian Journal of Linguistics* 23. 66–83.

Wetzels, W. Leo. 1985. The historical phonology of intrusive stops: A non-linear description. *Canadian Journal of Linguistics* 30. 285–333.

Zec, Draga. 2007. The syllable. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 161–194. Cambridge: Cambridge University Press.

54 The Skeleton

Péter Szigetvári

Only theories of phonology that attach significance to representations of phonological objects and, in addition, subscribe to an autosegmental version of these representations face the question of what the phonological skeleton looks like. Therefore, this chapter presupposes an autosegmental view of phonological representations.

The motivation for the autosegmental model is the fact that the segmentation of the speech signal can never result in absolutely discrete segments. Here segmentation is taken to mean practically the conversion of the continuous speech signal into the alphabetical symbols of the IPA. Some of these symbols pertain to more than one segment: for example, the stress mark to the syllable after it, tones potentially to even longer stretches. Take the question *you live by the sea?* Its last word, carrying the most prominent stress in the sentence, the tonic, might be transcribed as [↑siː]. In this transcription, the tone mark has a scope lasting all through the word (basically its only vowel): the pitch rises on steadily until the end of the utterance. The same holds if the string after the tone mark is longer, for example, as in *you live by the* | ↑*seaside, Martin?* It would take a very complicated mechanism to maintain that pitch was a property of individual segments and in some cases this rising pitch was realized on a single vowel, while in others it was split into low, higher, even higher, and highest pitch and added to several other vowels following. Tone is clearly not an immanent property of a vowel; it is an ephemeral phenomenon (from the point of view of a vowel) controlled by syntactic and pragmatic factors. If so, it is useful to represent it separately from the rest of the properties of the sound string. Such autonomous sound properties came to be known as autosegments.

If the phonological shape of an utterance is represented as a string of discrete feature bundles, the only option of representing the rising pitch in ↑*sea* includes a feature [rising tone] (here R) in the set of features corresponding to the vowel, as in (1a). In ↑*seaside, Martin*, on the other hand, a set of features [low tone], [higher tone], [even higher tone], etc. (here 0H, 1H, etc.) has to be assigned to the vowels following the tonic, as in (1b).

(1) a. siː^{R}
    b. siː^{0H} $\text{saɪ}^{\text{1H}}\text{d}$ $\text{mɑː}^{\text{2H}}\text{tɪ}^{\text{3H}}\text{n}$

One flaw of such representations is obvious: there appears to be nothing in common between the two rising tones, i.e. nothing to indicate their relationship. It is clear that the same tone is spread over the available vowels, but this is not shown in (1b).

Not only tone but many other sound properties turn out to be similarly promiscuous, with the potential of simultaneously belonging to several segments, and being manipulable independently of the segment(s) they belong to. (For further discussion, see CHAPTER 45: THE REPRESENTATION OF TONE and CHAPTER 14: AUTOSEGMENTS.)

The more sound properties extracted from their feature bundles, the fewer there remain. There are two widespread views on how many: according to one – historically the earlier – one feature, [syllabic], remains in the "bundle" (e.g. McCarthy 1979; Halle and Vergnaud 1980; Clements and Keyser 1983); according to the other, no feature remains (e.g. Levin 1985; Lowenstamm and Kaye 1986). The string of segmental positions thus vacated is called the phonological skeleton (a name suggested by Halle and Vergnaud 1980: 83) – or, alternatively, the timing tier or skeletal tier. The former type, in which the skeletal positions hold the feature [syllabic], is the CV skeleton (discussed in §2); the latter one, which is completely empty, is the X skeleton (discussed in §3). A non-segment-based framework involving only syllables and moras is introduced in §4. I will then argue that there is a way of incorporating moras in the old CV skeleton with clear advantages over the moraic framework (§5).

To begin with, let us examine the types of relation that may exist between skeletal positions and phonetic features associable to them.

1 Melody–skeleton relations

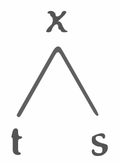
Skeletal positions represent the presence of a segment, and serve as an anchoring site for the phonetic properties associated with that portion of the speech signal. If the relationship of feature bundles – referred to as melody, again following Halle and Vergnaud (1980), among many others – and skeletal positions were always one to one, the latter would be superfluous. But, as we have already seen in the case of tone, this is not so. Let us take the non-trivial options one by one.

1.1 *One-to-many relations*

The standard textbook examples for this type of skeleton–melody relationship are affricates and prenasalized plosives (see CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION). With respect to the former, there has been much debate among phonologists about the feature [±delayed release] (and the marginal oppositions it creates), which Chomsky and Halle (1968) introduced to distinguish affricates from plosives. An alternative approach, that affricates are bisegmental (discussed by Gimson 1989: 172f. and Roca 1994: 3, among others), as suggested by the IPA symbols used to represent them, is undermined by many facts. In most cases, the distribution of affricates shows that they are not clusters, but single segments. It is even possible that an affricate in a system does not contrast with a homorganic fricative (e.g. Castilian Spanish has [ʧ], but not [ʃ]), rendering a cluster analysis improbable.

The separation of quantity (skeleton) and quality (melody) offers an opportunity for handling the quantitatively simplex, but qualitatively complex, affricates in an intuitive way, as so-called contour segments. (2) depicts the view of the affricate [ts] along these lines. (The skeletal slot is represented as "x", but this is not meant to indicate a standpoint in the CV *vs.* X skeleton debate.)

(2) *An affricate as a contour segment*

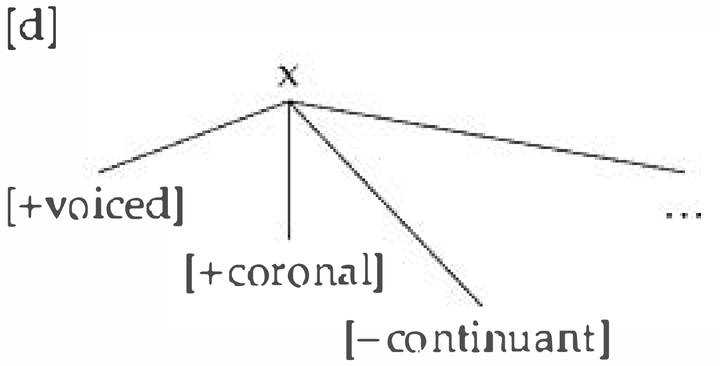


The representation in (2), however, incorporates a misconception, namely, that the melody of segments, without the slot they attach to, forms some kind of unit, two of which are here associated with a single skeletal slot. In reality, the symbols "t" and "s" above have no theoretical status. What exist in an autosegmental framework (or for that matter any other phonological theory since the middle of the twentieth century) are features, many of which occur in both parts of the affricate (e.g. place of articulation, laryngeal properties). Another difficulty with the contour model of affricates lies in the interpretation of autosegmental representations. Any melody linked to a slot of the skeleton – also known as the timing tier – is interpreted simultaneously. Temporal sequencing is managed by the skeleton, i.e. what is linked to an earlier slot is interpreted earlier than what is linked to a later slot. Associating the stop part of the affricate to the left leg of the contour segment and the fricative part to the right is then no more than a graphical trick, which cannot have any realizational consequences. The standard solution to this problem, involving root nodes, is discussed in §4. As Clements (1999) and Lin (CHAPTER 16: AFFRICATES) argue, affricates are best thought of as non-contour segments (strident stops), as Jakobson *et al.* (1952) have proposed.

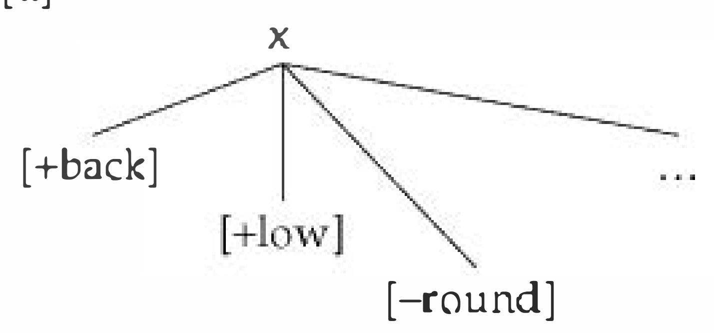
It seems then that we are left without one-to-many relations between the skeleton and melodic material. In fact, such relations are the most common occurrences in representations, since it is not segments but features that are associated with the slots of the skeleton. Thus most segments embody the one-to-many relation, as the partial representations of two very common segments, [d] and [ɑ], show in (3).

(3) *Partial autosegmental representations*

a. [d]



b. [ɑ]



1.2 *Many-to-one relations*

The long–short contrast of a vowel could be encoded in a feature [long], so that the long vowel is [+long] and the short one [–long]. It is evident, however, that this is not an adequate way of modeling length contrasts. Vowel length (or consonant length for that matter, on which see CHAPTER 37: GEMINATES) is not a property like vowel height (or the voicing of obstruents): it does not harmonize or trigger or undergo assimilation of any type (see CHAPTER 20: THE REPRESENTATION OF VOWEL LENGTH). Furthermore, changes in segmental length are usually unlike common assimilatory changes. Take, for example, the Rhythmic Law of Slovak, which shortens a suffixal long vowel after a long vowel in the stem. The agentive *-ník* (the acute accent marks length) inherently contains a long vowel (e.g. *rol-ník* [roljniːk] 'farmer'), which shortens when added to a long-vowelled stem (e.g. *stráž-nik* [straːʒɲik] 'guard'; Kenstowicz and Rubach 1987). The rule could be categorized as a dissimilatory process. What is conspicuously missing in languages is any assimilation of this type: i.e. changes where a short vowel would lengthen in the vicinity of a long vowel, and, crucially, because of that long vowel, or a long vowel would shorten purely because of the shortness of a neighboring vowel.

An even more telling phenomenon is compensatory lengthening (see also CHAPTER 64: COMPENSATORY LENGTHENING).[1] A synchronic comparison of the forms of the 1st singular copula in two varieties of Ancient Greek, Attic [eːmi] and Aeolic [emːi] 'I am', suggests a simple shift in the host of the alleged feature [+long]. In light of the reconstructed Proto-Greek etymon *[esmi], however, a different analysis is called for. The loss of the [s] triggers the lengthening of one of the neighboring segments, the preceding vowel in Attic and the following consonant in Aeolic. If length were encoded by a feature, the change could only be described by a pair of rules applying simultaneously, one deleting the coda consonant, the other lengthening the segment next to the deletion site. It is clear that the two rules are interrelated: spontaneously, open syllable lengthening is not attested in Attic, nor is intervocalic gemination in Aeolic – these changes only occur in tandem with the loss of the coda consonant. It is therefore difficult to understand why these two rules so commonly co-occur. If the quantity of segments is stored separately from their quality, this process, and any similar one, has a very neat explanation: it is only the quality (melody) of the coda [s] that is lost (more precisely, only its association with the skeleton); its place, that is, the time it occupied in the string of sounds, is retained (cf. e.g. Ingría 1980; Steriade 1982; Hock 1986; Hayes 1989). It is this empty place that one of the neighboring segments fills in, as shown in (4).

(4) *Compensatory lengthening: The stability of the skeleton*

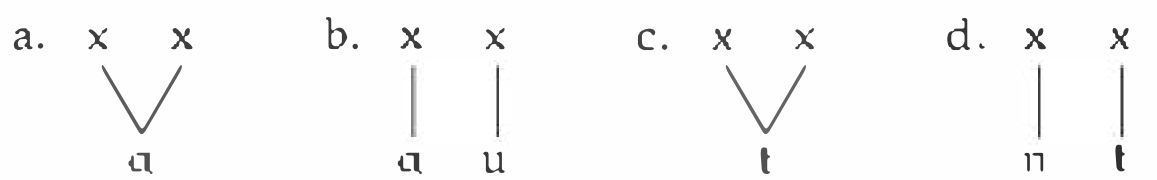
a. *Attic* b. *Aeolic*

[1] Much of the literature limits the term *compensatory lengthening* to cases involving the lengthening of a vowel. The lengthening of a consonant is called inverse compensatory lengthening by Hayes (1989: 280–281). Here I will refer to both processes by the same name.

While cases like the above could also be analyzed as the total assimilation of the [s] to the preceding vowel or the following consonant, there are more complicated types of compensatory lengthening, for which such an analysis is not at all viable. Cases in point include Middle English *tale* [talə] > [taːl] (Minkova 1982), Old Church Slavonic *bógŭ* > pre-Serbo-Croatian *bôg* [bóog] 'god', *bobŭ* > *bôb* [boób] 'bean' (Hock 1986: 435), and Old Hungarian [hida] > *híd* [hiːd] 'bridge', [lɛvɛlɛ] > [lɛvɛːl] (Modern Hungarian *levél* [lɛveːl] 'leaf', with subsequent closing of the second vowel; E. Abaffy 2003: 331). The English case is controversial (see Lahiri and Dresher 1999 for a different analysis), but for the others there is evidence that they are not cases of open syllable lengthening ensued by apocope. In Slavonic the original bisyllabic stress pattern is preserved on the long vowel of the monosyllabic forms. In Hungarian no lengthening takes place before suffixes that retain the stem-final vowel: Modern Hungarian *hídam* [hidɑm] 'my bridge', *levelek* [lɛvɛlɛk] 'leaves'. Lengthening due to a minimal word constraint is also excluded by the last example: the process takes place in monosyllabic and polysyllabic words alike.

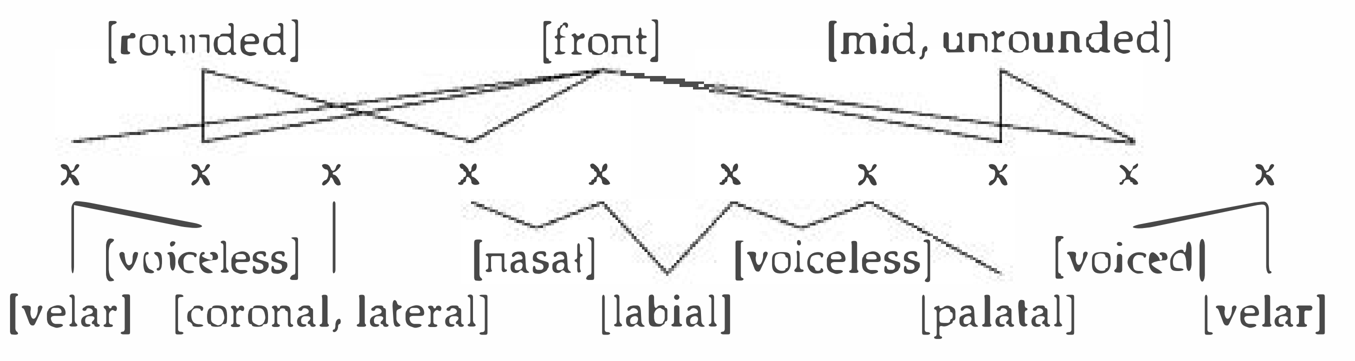
The bipositional analyis of long vowels is also made likely by the fact that they behave similarly to "vowel clusters," i.e. diphthongs. In English, for example, neither category occurs before non-coronal consonant clusters, and both occur word-finally, unlike short monophthongs (Fudge 1969: 272f.; Harris 1994: 37; Gussmann 2002: 20–23; see Prince 1984 for the same conclusion in Finnish for both vowels and consonants). Accordingly, there is a general consensus that long vowels ought to be represented as in (5a), and long (i.e. geminate) consonants as in (5c). The representation of diphthongs and other consonant clusters is given in (5b) and (5d), for comparison.

(5) *The autosegmental representation of vowel and consonant clusters*



It is not only complete segments that may be linked to more than one skeletal position. The standard situation in fact is that features (autosegments) are multiply linked. Take, for example, the Hungarian word *különbség* [k̟ylø̃mpʃeːg] 'difference', depicted in (6). (The features only serve illustrative purposes; their exact identity and location is irrelevant here.)

(6) *Multiply linked features in the representation of* [k̟ylø̃mpʃeːg]



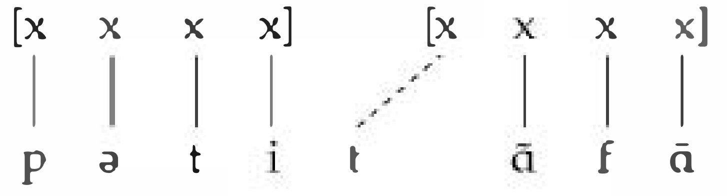
Chaotic as it seems, the diagram in (6) does not contain all the relevant features specifying the segmental content of the string [k̟ylø̃mpʃeːg] – manner of articulation features, for example, are lacking. Nevertheless, it can clearly be seen that it is more common for a feature to be associated with several skeletal slots than to be associated exclusively with one. In (6) this is because of voicing, place of articulation assimilations, vowel nasalization, and consonant fronting, as well as vowel harmony. Even when feature sharing is not a result of such phonological processes, i.e. in monomorphemic items, the multiple association of a single feature is dictated by the Obligatory Contour Principle (Leben 1973; McCarthy 1986; Odden 1986).

1.3 One-to-zero and zero-to-one relations

As we have seen, many-to-one and one-to-many relations between the skeleton and melody are very common. Two further options are discussed in this section. It is possible that a skeletal position is not associated with any melodic material. The opposite case may also occur: features unlinked to any point on the skeleton.

French liaison exemplifies both of these possibilities (see CHAPTER 112: FRENCH LIAISON). In this phenomenon, a word-final consonant is pronounced when the next word begins with a vowel, but not when it begins with a consonant. (The syntactic conditions on liaison need not concern us here.) Thus in the phrase *petit garçon* 'little boy' the first element ends in a vowel ([pəti garsõ]), in *petit enfant* 'little child' a [t] is pronounced ([pətit ɑ̃fɑ̃]). According to one analysis (Prunet 1987: 226) *petit* comes with only four skeletal slots, but five segments; *enfant*, on the other hand, has an extra skeletal slot – it begins with an initial consonantal slot which is empty. The situation is shown in (7).

(7) *Liaison*



The [t] at the end of *petit* is not associated to the skeleton; it is said to be floating. Floating melody fails to be pronounced unless it is able to associate to the skeleton. Vowel-initial words supply an empty skeletal position that the floating melody can associate to. The floating [t] at the end of *petit* must be lexically determined: there are other liaison consonants besides [t], whose identity is unpredictable (e.g. *gros enfant* [groz ɑ̃fɑ̃] 'fat child', *mon enfant* [mõn ɑ̃fɑ̃] 'my child', *gentil enfant* [ʒɑ̃tij ɑ̃fɑ̃] 'nice child', *long article* [lõg artikl] 'long article', etc., where the consonant before the space appears only if the next word begins with a vowel). Therefore this consonant must be included in the lexical representation. It is also not unjustified to suppose that vowel-initial words carry an empty skeletal slot at their left side. It is true for all languages that at least some words (and syllables) begin with a consonant (CHAPTER 55: ONSETS). For some languages this is not optional, but obligatory; crucially, though, there are no languages where an onset is not possible. One may argue that a syllable-initial consonantal position is in fact obligatory in all languages, the optionality lying in whether this position may or may not be left empty (see e.g. Kaye 1989: 134). Thus consonant-initial words do not carry an empty skeletal slot to their left, while vowel-initial words do, and as a result, the latter can host the floating consonantal melody at the end of the preceding word. Apparently, even languages that allow syllable-initial consonantal positions to be empty prefer them to be filled.

Hypothesizing that there is an empty skeletal position between two vowels in hiatus and that languages make an effort to fill it also explains the common phenomenon of hiatus filling (CHAPTER 61: HIATUS RESOLUTION). Unless a language manages to get rid of this consonantal position (often together with one of the neighboring vowels), an intervocalic consonantal position is filled by some melody associating to it from one of the vowels. (8) illustrates this with English *skier* and Hungarian *síel* [ʃiːɛl] 's/he skies'.

(8) *Hiatus filling*

The hiatus between [iː] and [ə] or [ɛ] is filled by the melody of the first vowel, resulting in the forms [skiːjə] (Gimson 1989: 215, 2001: 213) and [ʃiːjɛl] (Siptár and Törkenczy 2000: 283).

The possibility of vocalic positions being empty is considerably more controversial; this issue will be taken up in §5.

2 The CV skeleton

The notion of the CV tier was originally developed for the analysis of the non-concatenative morphology of Classical Arabic by McCarthy (1979, 1981). As in other Semitic languages, a large number of morphological categories are not expressed by linking morphemes after one another, but by fusing individually unpronounceable components into one. A similar, but much less elaborate case is the ablaut found in Germanic languages, e.g. English *sing*, *sang*, *sung*, and *song*, where the consonants carry the lexical entry and the vowel the grammatical category. (See also CHAPTER 105: TIER SEGREGATION and CHAPTER 108: SEMITIC TEMPLATES.)

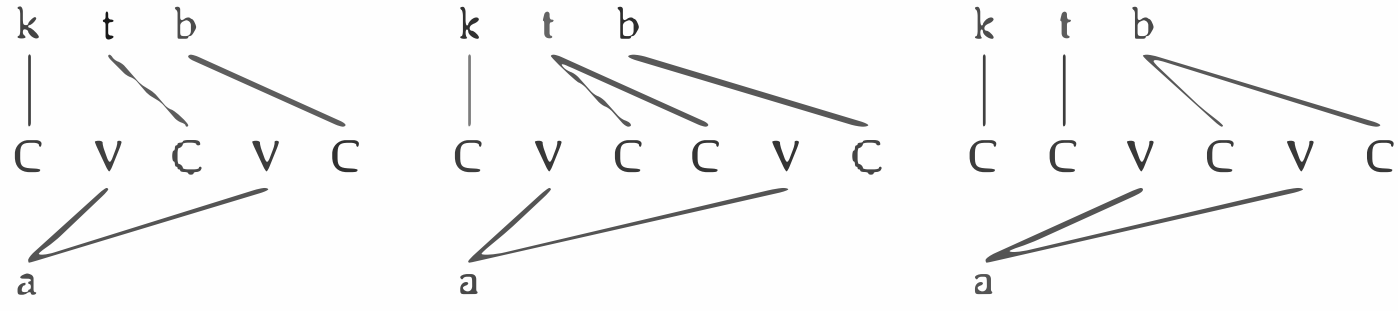
Paradigms in Arabic are classified into groups traditionally called conjugations – or, as McCarthy refers to them, binyans. The prime phonological property of a binyan is the order in which consonants and vowels are arranged. Roots of three (sometimes two or four) consonants contribute a lexical field to the meaning; the vowels are often responsible for grammatical categories like tense and voice. A portion of McCarthy's (1979: 244) table depicting the forms for the root /ktb/ 'to write' is given in (9).

(9) *Some forms of* /ktb/

binyan	*perf act*	*perf pass*
I	katab	kutib
II	kattab	kuttib
III	kaatab	kuutib
...		
IX	ktabab	—

The CV skeletons of the first three binyans are CVCVC, CVCCVC, and CVVCVC, respectively; that of binyan IX is CCVCVC. The root consonants and the vowels supplied by the grammatical category are mapped onto this skeleton more or less according to the association conventions elaborated by Goldsmith (1976). Three cases are shown in (10).

(10) a. katab b. kattab c. ktabab

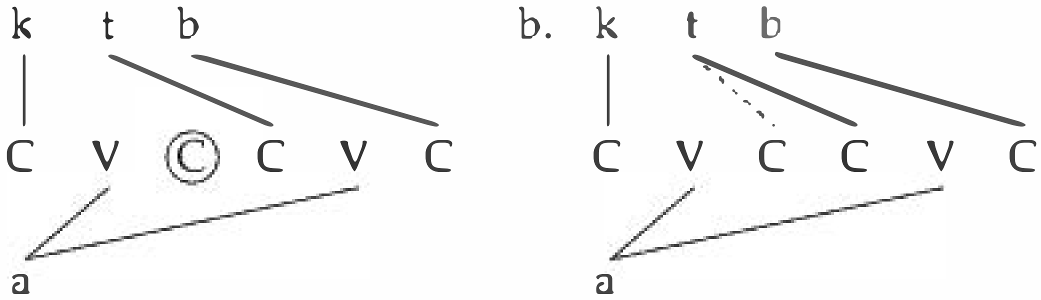


In (10a), the consonants are linked to the C slots of the skeleton, one by one. It is vital that the consonantal and vocalic skeletal slots be distinguished, since the linking of the root consonants and the vowel(s) can be done as required only thus. The case of (10c) shows that association takes place from left to right: with three consonants to four positions, the last consonant is linked to the surplus position ([ktabab]). (10b) poses a problem in this respect: either association is idiosyncratically edge-in in this case, or some extra mechanism is needed. McCarthy (1979: 256) uses brute force here: he assumes the expected *[katbab] in the first round, with a later rule delinking the first linkage of [b] ([katCab], where C represents the slot from which the melody of the [b] is delinked). This is automatically followed by the spreading of the [t] ([kattab]), much like an instance of compensatory lengthening (see §1.2).

A slightly less powerful solution is proposed by Lowenstamm and Kaye (1986: 117–118), who claim that association to the first position is inhibited from the start, thus each consonant of the root occupies its final position in the first round, as shown in (11a). The resulting configuration (empty C followed by filled C) is interpreted as a geminate, as in (11b). I have adapted the original to the previous diagrams of this chapter to aid comparison. We will see below (§3) that Lowenstamm and Kaye use a significantly different scheme.

(11) *The mapping of a geminate* ([kattab])

a. b.



Note that McCarthy's second-round-spreading solution cannot apply after Lowenstamm and Kaye's first-round blocking, since this would yield the unattested form *[kaktab].[2]

[2] Neither analysis gives a reason for delinking or inhibiting the association of the consonant encircled in (11a), so that the unattested form *[katbab] is avoided. Following Hoberman (1988), we may assume that long-distance geminates (those separated by a vowel) are more marked (their inhibition is ranked higher) than local ones and that word-initial geminates are even more marked. This explains why [kattab] is preferred to *[katbab], but [ktabab] to *[kkatab].

In McCarthy's analysis, the CV skeleton of Arabic words is a morpheme (a prosodic template, in his words), identifying the binyan of the word form, contributing to the semantic elements of the specific binyan (as if the Attic–Aeolic difference between [eːmi] and [emːi] represented a difference in morphological categories).

Clements and Keyser (1983: 11) apply the CV skeleton as a universal phonological device, the mediator between the syllable and autosegments, its two types of members, C and V representing "the useful but ill-defined notion of 'phonological segment'." The C for them is an anchor for anything [−syllabic] and the V for [+syllabic] segments. Prince (1984) shows that such impoverished representations adequately capture the templates of, for example, verbal person endings in Finnish, which are -C in the singular and -CC[e] in the plural, with the melody [m] in the 1st, and [t] in the 2nd person. The surface forms are thus 1st singular [-n] (by an independently motivated rule turning [m] to [n] word-finally), 2nd singular [-t], 1st plural [-mme] and 2nd plural [-tte].

3 The X skeleton

Simultaneously with the development of theories of the CV skeleton there evolved an alternative view that considered the distinction of C and V slots redundant, and argued that skeletal slots are uniform, usually marked with dots or x's (e.g. Lowenstamm and Kaye 1986; Levin 1985). Proponents of the X skeleton have put forward a number of arguments against skeletal positions predestined for syllabicity.

3.1 Reduplication in Mokilese

Levin (1985: 35–41) shows some unusual cases of reduplication from Mokilese, which, she believes, are analyzable only with an X skeleton. The point is that the reduplicant is a copy of the first three segments of the first syllable of the stem, irrespective of their being consonants or vowels. Levin argues that the template of the reduplicant must therefore also lack this information. The relevant data are given in (12).

(12) *Mokilese reduplication*

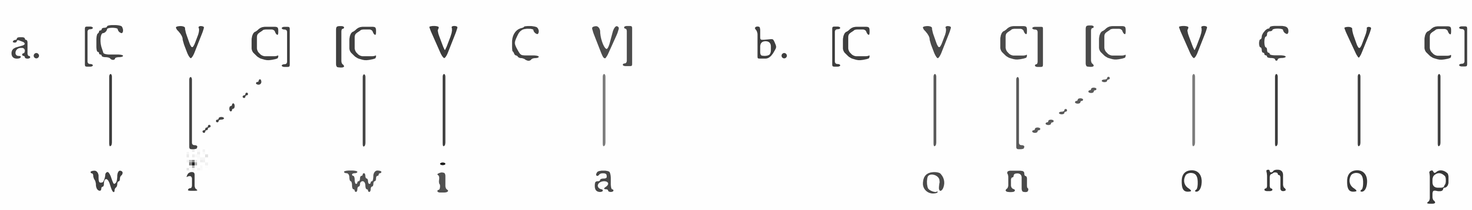
	stem	*progressive*	
a.	pɔdok	pɔdpɔdok	'plant'
b.	kasɔ	kaskasɔ	'throw'
c.	pa	paːpa	'weave'
d.	wia	wiːwia	'do'
e.	caːk	caːcaːk	'bend'
f.	onop	onnonop	'prepare'
g.	andip	andandip	'spit'

Levin contends that the reduplicant must be a totally specificationless skeleton, $_{\sigma}$[xxx], to which the copy of the melody of the stem is associated following universal conventions. The case of (12a), (12b), and (12g) is now straightforward. When the stem is too short, as in [pa], (12c), the last melody is multiply linked. The fact that the reduplicant is a single syllable inhibits the second vowels of [wia] and [onop] from associating to the skeleton (12d) and (12f). As a result, the preceding vowel or consonant is lengthened again. There is a problem with the stem [caːk] though, (12e). The melody of the stem comprises three segments, "|c|", "|a|", and "|k|", and the expected reduplicated form is therefore *[cakcaːk] rather than the attested [caːcaːk]. Levin has to stipulate that multiple melodic associations like that of the long [aː] are transferred in reduplication. A further problem of this analysis lies in the interpretation of the reduplicant: it is specified as a syllable, but it is not one in [on.n-onop] or [an.d-andip] (where the hyphen indicates the boundary between the reduplicant and the stem), since a word-internal pre-vocalic consonant forms a syllable with the following vowel, as the universal onset maximalization principle requires. Yet the constraint on the reduplicant being a syllable cannot be relaxed, because if the first three segments were copied without reference to a syllable, undesirable results like *[wi.a-wia] or *[o.no-onop] would emerge. In fact, Moravcsik says that in her survey of reduplication types she has never come across formulations like "reduplicate the first two [or, in our case, three – szp] segments (regardless of whether they are consonants or vowels)" (1978: 307–308). If in a language reduplication copies the first CVC part of the stem for consonant-initial stems, it will copy VC (not VCV) of vowel-initial stems.

Actually, a simpler account is available for the data in (12). Theoretically it is no more plausible than Levin's, but needs fewer stipulations, and thus invalidates her analysis *as an argument* for the X skeleton. Suppose, as in §1.3 above, that syllable onsets are always represented on the skeleton, either as a filled or as an empty C position. (This immediately explains Moravcsik's observation.) The reduplicant then is a copy of the first CVC part of the stem, melody and skeleton included. The cases of (12a–c) are obvious. The third slot for (12c) is automatically filled by the vowel of the reduplicant, just as for Levin. The objection (also put forward by Broselow 1995: 184) that vowels cannot spread onto a consonantal slot is mistaken: a C slot is not meant to host consonants exclusively, but non-syllabic segments. If a syllable has one syllabic segment, then a long vowel is hosted by a VC sequence on the skeleton, as Clements and Keyser (1983: 12) argue.

In (12d), the empty intervocalic C position is involved in the copying, but, as it is preconsonantal in the reduplicant, it serves as an anchor for the preceding vowel, unlike in the stem, where it is pre-vocalic. This is shown in (13a). Pre-vocalic stems blindly copy the initial empty C position, and so only the first two "real" segments form the reduplicant. (12g) seems to cause a problem now: here the reduplicant appears to be [and-], i.e. VCC, instead of the expected VC. Raimy (1999) suggests an obvious solution: if [nd] is analysed as [ⁿdd], a geminate pre-nasalized stop, then the situation is identical to that in (12f). The stem-initial empty C must be filled to satisfy onset maximalization: it is impossible to have a coda consonant followed by an empty onset. This is illustrated in (13b). (The reduplicant and the stem are enclosed in brackets for easier identification.)

(13) *Reduplication and empty onsets*



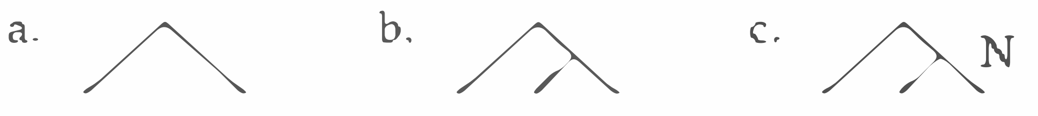
In fact, Levin herself suggests the empty-C-slot analysis as an escape hatch for the CV skeleton, but rejects the idea on the grounds that the vowel of the causative prefix [ka-] does not lengthen when prefixed to vowel-initial stems (e.g.

[ka+adanki] > [ka:danki] 'to name', [ka+uru:r] > [kauru:r] 'to be funny'). Vowels do not usually lengthen by filling a pre-vocalic empty C position (cf. Hayes 1989: 281); what is more, it is hard to expect a long vowel or a diphthong to further lengthen. The conclusive test, the prefix [ak-], which is expected to geminate its consonant if prefixed to a vowel-initial stem if there was an empty consonantal slot, "was only found prefixed to C-initial stems" (Levin 1985: 40). We can conclude that the hypothesis that vowel-initial stems carry an empty consonantal position at their left edge is not refuted by Levin's data.

3.2 *Redundancy of C and V*

A better argument against CV skeletons is that specifying syllabicity on the skeleton is redundant if the same information can be read off higher prosodic structures, like syllabic constituents, especially the nucleus. Lowenstamm and Kaye (1986) argue that simple syllable trees, like those in (14), adequately define the slots of the skeleton.

(14) *Syllable trees*

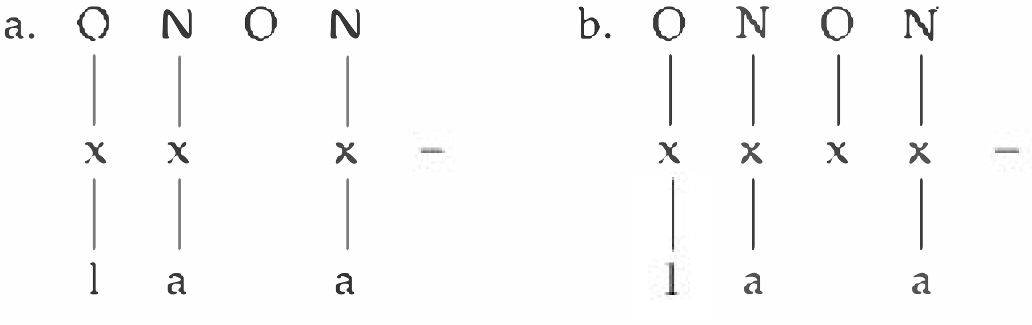


They suggest that labeling the trees is unnecessary since this information also follows from the configuration. Nevertheless, some minimal labeling (N, i.e. "nucleus") is necessary to distinguish CVC, (14b), and CVV, (14c), syllables – consider, for example, the Arabic templates of binyans II ([kattab]) and III ([kaatab]) (see (9)).

Lowenstamm and Kaye (1986) raise the issue of whether the skeleton is an independent level in phonological representations, or merely a projection of higher prosodic structure, in particular syllable structure. A consequence of this assumption is that the nodes representing syllabic constituents (like onset or nucleus) cannot be distinguished from the skeletal position(s) that they dominate. That is, it is impossible to conceive of skeletal positions not dominated by higher prosodic structure, or of a syllabic constituent that does not dominate a skeletal position.

Charette utilizes "pointless onsets" in an analysis of h-*aspiré* words in French (1991: 90f). She claims that "normal" vowel-initial words begin with an onset that does not dominate any skeletal position, while those which contain h-*aspiré* – words that phonetically begin with a vowel, but phonologically behave as consonant-initial – begin with a regular "pointful" onset, dominating a skeletal position which is not associated with any melody. The vowel of the definite article is unpronounced before vowel-initial words, but it is pronounced before consonant- and h-*aspiré*-initial words. (15) illustrates the first part of the two cases using Charette's examples: *l'amie* [lami] 'the girlfriend' and *la hache* [la aʃ] 'the axe'.

(15) *Two types of empty onset*



According to Charette's analysis, the vowel of the article is deleted before a pointless onset as a result of the Obligatory Contour Principle, since the two nuclei are "adjacent" if the onset between them lacks a skeletal slot, as in (15a). When such an onset is linked to a skeletal slot, it inhibits the deletion process, as in (15b). This analysis faces difficulties on several counts. On the one hand, the Obligatory Contour Principle controls the appearance of identical *melodic elements* linked to adjacent skeletal positions. The nodes labeled "nucleus" do not qualify as such. On the other hand, liaison calls for the opposite representation of the two types of vowel-initial words. As mentioned in §1.3, some morphemes that are vowel-final preconsonantally exhibit a consonant when followed by a vowel-initial word. The plural of the definite article is an example: *les amies* [lez ami] 'the girlfriends' *vs. les haches* [le aʃ] 'the axes' (recall that h-*aspiré*-initial words behave as if they were consonant-initial). The final [z] of the cliticized article is pronounced when there is no skeletal position for it to anchor to, and it is not pronounced when there is one, i.e. without further stipulations, Charette's analysis predicts just the opposite of the attested liaison facts. The impoverished structures of (14) are also impossible if labels like "onset" and "nucleus" are treated separately from what they label: the skeletal slots.

To summarize, there is no compelling reason to distinguish skeletal points and the syllabic constituents containing them. Allowing pointless constituents or constituentless skeletal points makes unnecessary contrasts possible. But then, if prosodic nodes like onset and nucleus are not distinct from skeletal slots, then skeletal slots do carry the basic information of syllabic status: such a skeleton does contain Cs and Vs, irrespective of whether this is penciled on paper as Cs and Vs, Os and Ns, or something else. The two levels must, nevertheless, be kept distinct if more than one skeletal slot can be associated with a single syllabic constituent, i.e. if branching onsets and nuclei are posited. §5 discusses a model where even these are claimed not to exist.

4 Moras

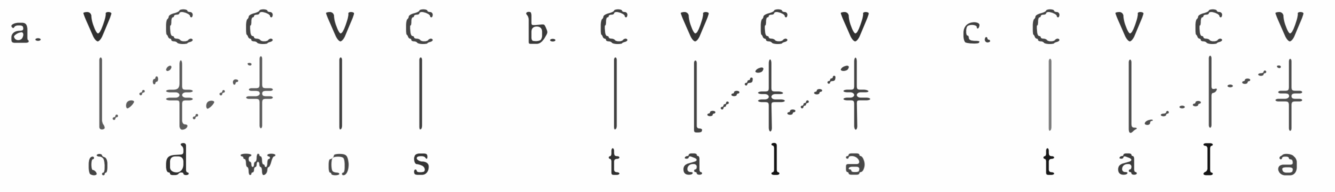
As we have seen in the case of Mokilese reduplication (§3.1), preconsonantal empty C positions are available as targets for the spreading of a preceding vowel; intervocalic ones are not. In many languages, a similar asymmetry characterizes these two consonantal positions. Stress processes, for example, may treat a preconsonantal consonant as being on a par with vowels, but not pre-vocalic consonants.[3]

Hock (1986) argues that the notion of mora must be (re)introduced into phonological theory. The mora has been around in linguistic discussions for at least two centuries (see Allen 1974: 100); it is its theoretical status that is at issue here (see also CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE). Hock's proposal is to introduce

[3] It is common at this point to offer a disclaimer with respect to Everett and Everett (1984) (who claim that Pirahã is different in this respect) or to Davis (1988) (who collects cases where the quality of the onset seems to play a role in stress assignment). However, as Hayes states: "I believe that the ability of Moraic Theory to account for wide-spread patterns of markedness should be given more weight in assessing the evidence than any particular awkwardness in the analysis of individual languages" (1989: 303). This is probably true for any theory. Furthermore, some of the very few onset-sensitive systems have been shown to be re-analyzable so that they are not onset-sensitive (Goedemans 1996; Takahashi 1999).

the mora as an autosegment, rather similar to tones: in his proposal tones are indeed linked to moras. If compensatory lengthening could only lengthen a vowel in compensation for the loss of a tautosyllabic consonant, the "standard" CV or X skeleton would be fully capable of dealing with the process. We have seen, however, that compensatory lengthening also occurs at a distance: the loss of a vowel in the following syllable may lead to lengthening across an intervening onset consonant. Relevant cases are Greek glide loss (e.g. Proto-Greek [odwos] > Ionic [oːdos] 'threshold'; Steriade 1982: 118) and Middle English schwa apocope (e.g. [talə] > [taːl] 'tale'; Minkova 1982). In both cases the melody delinks and the vowel spreading is separated by a consonant that apparently remains linked to the skeleton.

(16) *Problematic cases of compensatory lengthening*



Actually, as (16a, b) show, the consonant standing in the way of compensatory lengthening is shifted to the right by one slot in both cases. This process, proposed by Steriade (1982: 126–128), is referred to as "double flop" by Hayes (1989: 265–267). The Greek case in (16a) can be explained by universal principles: the loss of [w] leaves us with an empty onset (provided that the syllabification is [od.wos]). The resulting [od.os] violates the onset maximalization principle, thus resyllabification ensues. But the skeletal position does not resyllabify, since there is an empty onset slot, recently vacated by [w]. It is to this slot that the [d] associates, leaving its original slot empty, triggering the lengthening of the preceding vowel.

The lengthening triggered by apocope, exemplified by Middle English [talə] > [taːl] in (16b), is more problematic for a theory which lacks moras. The mechanism appears to be the same as in (16a), but now the consonant before the disappearing word-final vowel is supposed to flop to a vocalic position, to the nuclear slot of the last syllable. In addition, the position it leaves is not one that should cause lengthening of the preceding vowel. The alternative, whereby the vowel spreads out immediately to the vacated vocalic slot, as in (16c), is even worse, as it violates the axiomatic constraint inhibiting the crossing of association lines.[4]

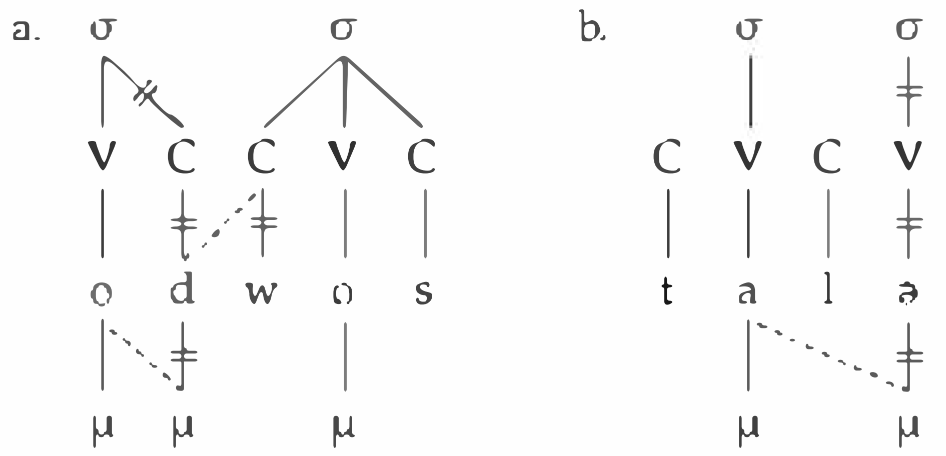
In fact, with both CV and X skeletons it is hard to explain why the spreading of a vowel to some consonantal slots should cause lengthening, while in other cases an apparently similar vowel spreading does not. For example, the empty onset in Hungarian *pia* [pijɑ] 'drink' is filled by the spreading of the melody of the preceding vowel, as in (8).[5] Yet the result is not a long vowel, which it is in *film* 'film', for which the pronunciation [fiːm] is possible (Siptár and Törkenczy 2000: 281) (cf. Hayes 1989: 281–283).

[4] This problem could be avoided by placing vowels and consonants on separate autosegmental planes (as in (10) and (11)); however, such a modification would loosen the theory beyond desirable limits: we would now find it hard to explain why so many processes deemed possible by the framework never occur.

[5] While it may be argued that *pia* is underlyingly [pijɑ], the question still holds why the same structure, the melody of [i] doubly linked to a V and a C slot, is [ij] in one case and [iː] in the other.

Hock's (1986) proposal is to attach a mora (μ) to each weight-bearing position, that is, to each vocalic position, as well as to some consonantal positions, notably codas. The two cases are shown in (17).

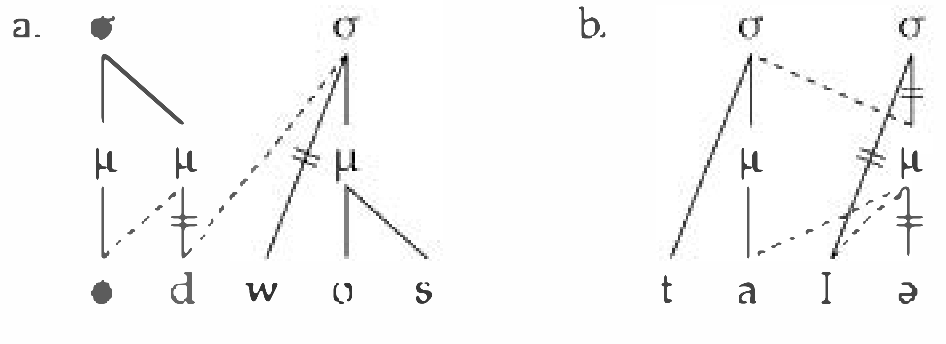
(17) *Double flop with moras*



The moraic analysis of [odwos] > [oːdos] in (17a) is not significantly different from the moraless one, shown in (16a). It nevertheless suggests a reason for the asymmetry between onset and coda consonants: the former do not possess a mora, while the latter do. The advantage of the mora analysis becomes clear in the lengthening of a vowel caused by apocope: [talə] > [taːl], (17b). The intervening onset consonant is not affected by the process at all, since it is not associated with a mora. Thus, the mora left floating after the final vowel is lost can associate to the stem internal vowel "above the head" (or rather "below the foot") of the intervening moraless consonant, much like in a vowel harmony process, where intervening consonants not possessing the relevant vocalic feature are transparent.

Hayes (1989) rearranges the relationship of the syllable and the mora by making the latter an integral part of prosodic structure, dominated by the syllable node. In a more radical innovation he also gets rid of the skeleton as previously conceived. In his view, the function of the skeleton is taken over by moras, and moraless consonants are either associated directly with the syllable node or share a mora with the moraic segment. Accordingly, the two processes displayed in (16) and (17) would be represented as in (18).

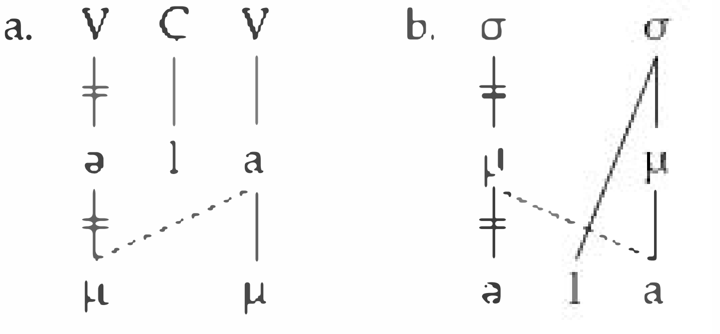
(18) *Moras as the skeleton*



The simple double-flop case of Greek glide deletion in (18a) does not require much comment, as the mechanism is the same as before. For Middle English apocope, however, Hayes needs an extra stipulation, called parasitic delinking: the loss of an overt nucleus in a syllable entails the dissolution of the whole syllable. What is now left of the last syllable, a mora and an [l], is joined to the first one, yielding the correct result. In Hock's analysis, on the other hand, the [l] remains in place, and does not have to be delinked and relinked, as can be seen in (17b).

Despite this complication, Hayes's model has definite advantages over Hock's use of moras. On the theoretical count, it is simpler in that it lacks the CV or X skeleton. On the empirical count, it predicts that compensatory lengthening of a vowel is only caused by the loss of a moraic segment that follows the vowel, never by the loss of one that precedes it. As (19a) shows, Hock's representations easily allow the latter case, which is not attested according to Hayes. His hypothetical example is [əla] > [la:].

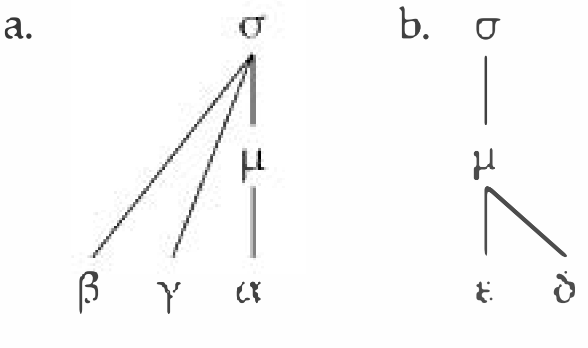
(19) *Compensatory lengthening triggered by loss of preceding vowel*



In Hayes's model, (19b), the freed mora of the first syllable cannot be captured by the second mora, because the onset consonant inhibits this. The price to pay for this solution is the stipulative parasitic delinking mentioned above: if the moraic segment of a syllable is delinked, the onset consonant is also delinked, as in (18b). Without this an onset will always block the linking of a heterosyllabic mora. Note that in Hock's model not only the loss of a vowel, but also the loss of a moraic consonant, could lead to the lengthening of a following vowel (e.g. Proto-Greek [esmi] > hypothetical *[emi:]). Such changes also seem to be unattested, as predicted by Hayes.

While theoretically attractive, dispensing with the skeleton has serious repercussions. Recall that linking IPA symbols to elements of higher prosodic structure (slots of the skeleton, moras, syllables) is misleading, since segments are not atomic. In partial trees like those in (20), where the Greek letters α–ε stand for (auto)segments, the temporal order of these autosegments is not specified. The string βγ is usually referred to as a branching onset; δ is a moraless coda, which may occur word finally even in languages with moraic codas, like English.

(20) *Autosegmental representations without a timing tier*



Accordingly, the order of two adjacent tautosyllabic or tautomoraic segments must be given by some stipulation. Kaye, for example, provides such a stipulation: "By universal convention the less sonorous of the two elements associated to the same point is produced first in the speech chain" (1985: 289). It remains to be seen if this can be maintained. For syllable-initial consonants, (20a), this is exactly what the sonority sequencing principle dictates. In the domain of single segments, affricates follow this convention, but the existence of prenasalized stops casts some doubt on its validity. Apart from light diphthongs (like French [wa] in *trois* [trwa] 'three'), monomoraic rhymal sequences, as in (20b), obviously cannot be subject to this generalization, since they are invariably ordered in the opposite way, with the more sonorous (vowel) first and the less sonorous (consonant) second. Be that as it may, without some similar (set of) principle(s) an autosegmental representation without a timing tier is uninterpretable.

To overcome this difficulty, one might wish to introduce root nodes, a notion familiar from frameworks organizing features into hierarchical structures, "feature geometries" (Clements 1985; Sagey 1986; McCarthy 1988; CHAPTER 27: THE ORGANIZATION OF FEATURES). The root node is the topmost node of such a hierarchy, containing all of the features making up the given segment, that is, the entirety of the segment. If the graphical order of root nodes specified their temporal order as well – as assumed in the contour-segment model of affricates – then "root node" would be just another name for "skeletal slot," i.e. one would simply reintroduce the skeleton into the representation. The skeleton is apparently indispensable.

5 A return to the CV skeleton

The modern career of the mora was launched by the need to distinguish onset consonants from coda consonants. Only the latter are capable of contributing to the weight of a syllable, that is, of behaving like a vowel; onsets are not (CHAPTER 55: ONSETS and CHAPTER 57: QUANTITY-SENSITIVITY). A mora is therefore assigned to consonants in the rhyme, but not to those in the onset. Note, however, that the reasoning is circular: codas are equipped with a mora because we observe that they behave differently, and then refer to these moras to explain their difference. But we could just as well imagine an alternative world in which onsets were moraic and codas were not. There is no inherent property of coda consonants that means they are fated to be moraic, as opposed to onset consonants. To make things worse, we will see that it is not exactly true that onset consonants are never moraic, or at least that their loss never entails compensatory lengthening. It turns out to be an oversimplification to tie diverse phenomena like compensatory lengthening, stress assignment algorithms, the assignment of tone-bearing units, etc. to a single property of the representation, i.e. moras (Hayes 1995: 299; Gordon 2004).

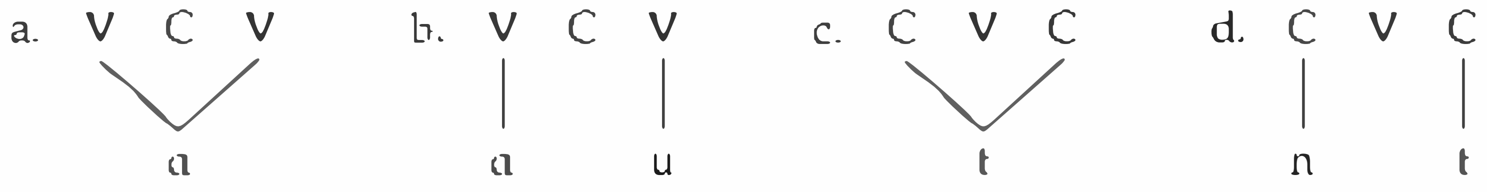
It is a version of the once rejected CV skeleton that might bring us closer to understanding this asymmetry in the behavior of consonants at the two edges of the syllable. To distinguish it from the McCarthy and Clements and Keyser type of CV skeleton, I will refer to it as the "strict CV skeleton." In §1.3 and §3.1, we saw why it is useful to suppose that some skeletal positions are empty. So far, we have only seen empty consonantal positions, but there is no particular reason why emptiness, i.e. the state of not being associated to any melodic material, should be limited to consonantal positions. The claim that the host of the vowel (the nucleus) is the head of the syllable, and therefore cannot be missing, is not a very strong one. Syntactic heads, for example the complementizer of a complementizer phrase, may remain empty (e.g. *I know* [$_{CP}$ [$_{C}$ Ø] *she'll come*]).[6] But other prosodic units like the foot may also exist without an overt head: in the previous sentence

[6] In fact, in English it is by default empty in non-questions, that is, there is an empty complementizer at the beginning of the matrix clause too.

the first headed foot begins with *know*, and the pronoun *I* before it forms a headless, degenerate foot. Feet and syllables are similar types of prosodic units, and headless syllables are therefore not inconceivable entities.

If nuclei may remain unpronounced, a very restricted syllable structure becomes available.[7] Lowenstamm (1996) proposes that underlyingly all languages have the same skeleton, the simplest one available, comprising non-branching onsets and non-branching nuclei in strict alternation. Accordingly, no two consonants and no two vowels are adjacent on the skeleton, as they are always separated by a position of the opposite type. (21) gives the four cluster types of (5) in a strict CV model.

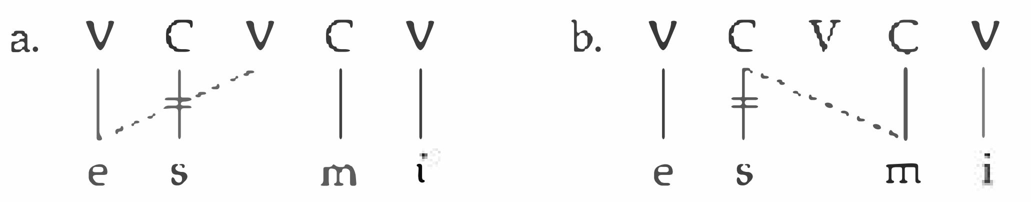
(21) *The strict CV representation of vowel and consonant clusters*



Recall the discussion in §3.2: the CV skeleton contains redundant information that can be read off higher prosodic structures. But this only holds if there is any higher prosodic structure. In fact, strict CV analyses generally do not call for the recognition of such structure, and certainly not of any further syllabic constituency.

The two cases of compensatory lengthening in Greek – Proto-Greek [esmi] > Attic [eːmi] and Aeolic [emːi] – are illustrated in (22).

(22) *Compensatory lengthening in a strict CV skeleton*

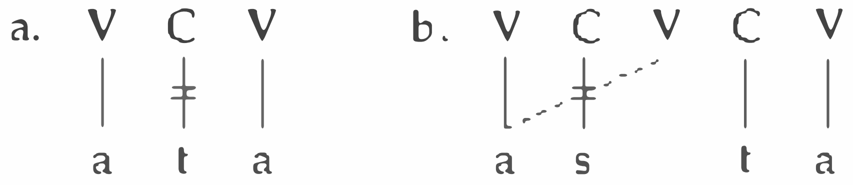


With the delinking of the [s], two skeletal slots are opened up for association: both the consonantal slot of the delinked coda and the vocalic slot enclosed within the original [sm] cluster. The choice is apparently controlled by a dialect-specific parameter, just as in any other theory of the skeleton.

The mora of moraic theories is an independent entity, which can be assigned to segments as the analyst needs it – it is only empirical considerations that stop them from assigning a mora to onsets. In the strict CV approach, moras are an inevitable consequence of the way the skeleton is built up (Scheer and Szigetvári 2005). A coda consonant is moraic because it is followed by an unpronounced vocalic slot. That is, the moraicness of the coda is only apparent: it is the following vocalic slot that carries weight. In this view, only vocalic slots are moraic. The loss of an intervocalic consonant does not free any "buried" empty vocalic slot, as (23a) shows. The loss of a preconsonantal consonant, on the other hand, makes an otherwise unreachable vocalic slot available as a target for spreading, as in (23b).

[7] Note that "empty" and "unpronounced" are not equivalent. In a privative feature framework, empty skeletal positions may be phonetically interpreted, as a sound maximally lacking any contrast, like e.g. [ə] or [ʔ]. Some empty skeletal positions may thus be pronounced; others may remain unpronounced if they satisfy certain conditions. See Kaye *et al.* (1985, 1990), Charette (1991), and Harris (1994) for details.

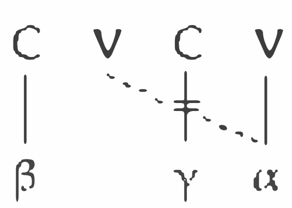
(23) *The loss of an intervocalic and a preconsonantal consonant*



The weight of closed syllables containing a short vowel is language-specific. For example, in English and Cairene Arabic such syllables count as heavy, while in Khalkha Mongolian and Yidin[y] they count as light (Zec 1995: 89). This parametric variation is trivially encoded in moraic frameworks: coda consonants are sometimes assigned a mora, and sometimes not. In the strict CV model, the same fact is encoded by parameterizing whether or not an unpronounced vocalic slot is counted by the relevant process. Crucially, however, since the shape of the skeleton is constant – it is always a strict alternation of vocalic and consonantal positions – the uncounted vocalic slot is there even when it is not counted by a certain process (say, stress assignment). One prediction running counter to those of Moraic Theory follows from this fact: compensatory lengthening of a vowel should be possible even if coda consonants are not moraic in a language. Kavitskaya (2002) claims that at least two languages, Piro and Ngajan, are exactly like this. One could claim that the mora associated with the coda in such languages is one which does not contribute to weight but does allow compensatory lengthening (as an anonymous reviewer points out). This then means that there are two types of mora, a "weight mora" and a "compensatory lengthening mora." The strict CV model predicts exactly this: there are two types of Vs. Pronounced Vs obligatorily contribute to weight, unpronounced ones are parameterizable.

In the strict CV framework, when an empty vocalic position enclosed between two consonants is "unearthed," compensatory lengthening may ensue, irrespective of whether this target of spreading is to the left or to the right of the vowel to lengthen. That is, the loss of an onset consonant may result in the lengthening of the vowel that follows it, as (24) shows.

(24) *Onset loss yielding compensatory lengthening*



The theory dictates that this option is available only for postconsonantal onsets, not for intervocalic ones (see (23a)). Confirmation of this prediction comes from southwestern dialects of Finnish where gradated [k] is lost, with compensatory lengthening. The data in (25) come from Kiparsky (2008); doubled vowels are long, as in standard Finnish orthography.

(25) *Compensatory lengthening in southwestern Finnish dialects*

input	*SW dialect*	*standard*	
/jalka-t/	jalaat	jalat	'legs'
/nælæ-n/	nælææn	næljæn	'hunger-GEN'
/halko-t/	haloot	halvot	'logs'

In the Finnish data, the lost consonant is always preceded by another consonant, and is never intervocalic. This is important, because the empty vocalic slot is available *between* two consonants, but not after a vowel, as (23a) shows.

Samothraki Greek exhibits a similar type of compensatory lengthening. In this dialect, pre-vocalic [r] is lost, and is only retained in preconsonantal position – a mirror image of the distribution in non-rhotic dialects of English. The loss of postconsonantal [r] is illustrated in (26a). Intervocalic [r] is lost without trace, as in (26b), as expected. (The data are from Topintzi 2006, who attributes them to Katsanis 1996.)

(26) *Loss of /r/ in Samothraki Greek*

	input	*output*	
a.	/'protos/	['po:tus]	'first'
	/'frena/	['fe:na]	'brakes'
	/'xroma/	['xo:ma]	'color'
	/'ɣrafo/	['ɣa:fu]	'I write'
b.	/'leftirus/	['leftius]	'free'
	/va'reʎ/	[va'eʎ]	'barrel'
	/'mera/	['mia]	'day'
	/'skara/	['skaa]	'grill'

To provide the missing mora, Hayes (1989: 283) has to hypothesize an epenthesis stage before the loss of the [r]: ['frena] > [fe'rena] > [fe'ena] > ['fe:na]. The strict CV analysis is rather similar, though the only difference is a very important one: the slot of the "epenthetic" vowel is lexically available, since any two consonants are always separated by such an empty slot. The relevance of this difference between the two analyses is that there is no empirical evidence for epenthesis in this case. Furthermore, this assumption creates a paradox in the ordering of the historical events (Kavitskaya 2002: 98), and Hayes's hypothesis is therefore not plausible. The strict CV skeleton, however, has a vocalic position to which the vowel can spread without any extra process.

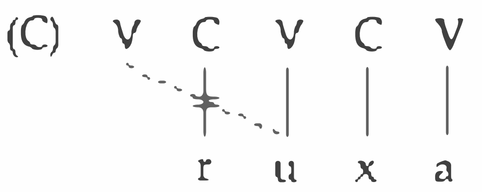
But even the strict CV model seems to be taken by surprise when it comes to the loss of word-initial [r]: this loss also triggers compensatory lengthening, as the words in (27) show.

(27) *Loss of /r/ in Samothraki Greek*

input	*output*	
/'ruxa/	['u:xa]	'clothes'
/'rema/	['e:ma]	'stream'

Scheer and Ségéral (2001) introduce the notion of "coda mirror." The coda is a typical lenition environment, being the position in the word that is not followed by a vowel, i.e. a preconsonantal or word-final position. The coda mirror is the opposite case: it is the position not preceded by a vowel, i.e. a postconsonantal or word-initial position, which is claimed to be the strong position, where lenition is unlikely. Scheer and Ségéral's theory is built on the strict CV skeleton: for them, "not followed by a vowel" means followed by an unpronounced vowel, and "not preceded by a vowel" means preceded by an unpronounced vowel. It is this empty vocalic position that causes the lengthening of the vowel in the Finnish and the Greek data discussed here. Not only postconsonantal, but also word-initial consonants are assumed to be preceded by an empty vowel, a proposal first argued for by Lowenstamm (1999). Accordingly, the loss of a word-initial consonant may also cause compensatory lengthening, as shown in (28).

(28) *Word-initial consonant loss yielding compensatory lengthening*



Since consonant loss is not common in the coda mirror position, compensatory lengthening is also rare in this environment. The peculiarity of Samothraki Greek, then, is that it unexpectedly exhibits [r] loss in the coda mirror position and not in the expected coda position. The ensuing compensatory lengthening is a consequence predicted by the strict CV skeleton.

6 Conclusion

The phonological skeleton evolved as a result of the autosegmental idea taken to its logical conclusion: segments, after autosegmentalizion of all their melodic content, leave behind "traces" that encode their relative temporal order. The debates concerning the phonological skeleton are (i) whether skeletal slots specify any phonetic property (consonantalness *vs.* vocalicness) or none, i.e. whether the skeleton contains Cs and Vs or uniform Xs; and (ii) whether the mora can replace skeletal slots, with moraless consonants linked directly to the syllable node. This chapter has argued that skeletal slots are Cs and Vs, not merely Xs, but there is no further prosodic constituency (e.g. onsets, nuclei, or syllables). Furthermore, it has been claimed that the mora is not an independent element of the representation, but a consequence of parametrical settings on vocalic skeletal slots: pronounced V slots are universally moraic; unpronounced ones are moraic in some, but not in other languages. Consonants, on the other hand, are never moraic.

ACKNOWLEDGMENTS

I acknowledge useful comments and advice from two anonymous reviewers, Marc van Oostendorp, Ádám Nádasdy, Péter Siptár, and László Varga. I thank them all. Remaining errors are mine.

REFERENCES

Allen, W. Sidney. 1974. *Vox Graeca: A guide to the pronunciation of Classical Greek*. 2nd edn. Cambridge: Cambridge University Press.

Broselow, Ellen. 1995. Skeletal positions and moras. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 175–205. Cambridge, MA & Oxford: Blackwell.

Charette, Monik. 1991. *Conditions on phonological government.* Cambridge: Cambridge University Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. 1999. Affricates as noncontoured stops. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.) *Proceedings of LP '98: Item order in language and speech*, 271–299. Prague: Karolinum Press.

Clements, G. N. & Samuel, J. Keyser. 1983. *CV phonology: A generative theory of the syllable.* Cambridge, MA: MIT Press.

Davis, Stuart. 1988. Syllable onsets as a factor in stress rules. *Phonology* 5. 1–19.

E. Abaffy, Erzsébet. 2003. Az ómagyarkor: Hangtörténet [The Old Hungarian period: Sound changes]. In Jenő Kiss & Ferenc Pusztai (eds.) *Magyar nyelvtörténet* [*A history of Hungarian*], 301–351. Budapest: Osiris.

Everett, Daniel L. & Keren Everett. 1984. Syllable onsets and stress placement in Pirahã. *Proceedings of the West Coast Conference on Formal Linguistics* 3. 105–116.

Fudge, Erik C. 1969. Syllables. *Journal of Linguistics* 5. 253–286.

Gimson, A. C. 1989. *An introduction to the pronunciation of English.* 4th edn., revised by Susan Ramsaran. London: Edward Arnold.

Gimson, A. C. 2001. *Gimson's pronunciation of English.* 6th edn. London: Edward Arnold.

Goedemans, Rob. 1996. An optimality account of onset-sensitive stress in quantity-insensitive languages. *The Linguistic Review* 13. 33–47.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.

Gordon, Matthew. 2004. Syllable weight. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 277–312. Cambridge: Cambridge University Press.

Gussmann, Edmund. 2002. *Phonology: Analysis and theory.* Cambridge: Cambridge University Press.

Halle, Morris & Jean-Roger Vergnaud. 1980. Three-dimensional phonology. *Journal of Linguistic Research* 1. 83–105.

Harris, John. 1994. *English sound structure.* Oxford: Blackwell.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Hoberman, Robert D. 1988. Local and long-distance spreading in Semitic morphology. *Natural Language and Linguistic Theory* 6. 541–549.

Hock, Hans Henrich. 1986. Compensatory lengthening: In defense of the concept "mora." *Folia Linguistica* 20. 431–460.

Ingria, Robert. 1980. Compensatory lengthening as a metrical phenomenon. *Linguistic Inquiry* 11. 465–495.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge, MA: MIT Press.

Katsanis, Nikolaos. 1996. *Το γλωσσικό ιδίωμα της Σαμοθράκης* [*The dialect of Samothraki Greek*]. Δήμος Σαμοθράκης [Municipality of Samothraki].

Kavitskaya, Darya. 2002. *Compensatory lengthening: Phonetics, phonology, diachrony.* London & New York: Routledge.

Kaye, Jonathan. 1985. On the syllable structure of certain West African languages. In Didier L. Goyvaerts (ed.) *African linguistics: Essays in honour of M. W. K. Semikenke*, 285–308. Amsterdam: John Benjamins.

Kaye, Jonathan. 1989. *Phonology: A cognitive view.* Hillsdale, NJ: Lawrence Erlbaum.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1985. The internal structure of phonological elements: A theory of charm and government. *Phonology Yearbook* 2. 305–328.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology* 7. 193–231.

Kenstowicz, Michael & Jerzy Rubach. 1987. The phonology of syllabic nuclei in Slovak. *Language* 63. 463–497.

Kiparsky, Paul. 2008. Compensatory lengthening. Paper presented at the CUNY Conference on the Syllable.

Lahiri, Aditi & B. Elan Dresher. 1999. Open Syllable Lengthening in West Germanic. *Language* 75. 678–719.

Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT.

Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.

Lowenstamm, Jean. 1996. CV as the only syllable type. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, 419–441. Salford: ESRI.

Lowenstamm, Jean. 1999. The beginning of the word. In Klaus Kühnhammer & John Rennison (eds.) *Phonologica 1996*, 153–166. The Hague: Holland Academic Graphics.

Lowenstamm, Jean & Jonathan Kaye. 1986. Compensatory lengthening in Tiberian Hebrew. In W. Leo Wetzels & Engin Sezer (eds.) *Studies in compensatory lengthening*, 97–132. Dordrecht: Foris.

McCarthy, John J. 1979. On stress and syllabification. *Linguistic Inquiry* 10. 443–465.

McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12. 373–418.

McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17. 207–263.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

Minkova, Donka. 1982. The environment for open syllable lengthening in Middle English. *Folia Linguistica Historica* 3. 29–58.

Moravcsik, Edith A. 1978. Reduplicative constructions. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 3: *Word structure*, 297–334. Stanford: Stanford University Press.

Odden, David. 1986. On the role of the Obligatory Contour Principle in phonological theory. *Language* 62. 353–383.

Prince, Alan. 1984. Phonology with tiers. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure: Studies in phonology presented to Morris Halle by his teacher and students*, 234–244. Cambridge, MA: MIT Press.

Prunet, Jean-François. 1987. Liaison and nasalization in French. In Carol J. Neidle & Rafael Núñez Cedeño (eds.) *Studies in Romance languages*, 225–235. Dordrecht: Foris.

Raimy, Eric. 1999. Strong syllable reduplication in Mokilese. In Rebecca Daly & Anastasia Riehl (eds.) *Proceedings from ESCOL '99*, 191–202. Ithaca, NY: CLC Publications.

Roca, Iggy. 1994. *Generative phonology*. London & New York: Routledge.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.

Scheer, Tobias & Philippe Ségéral. 2001. La coda-miroir. *Bulletin de la Société de Linguistique de Paris* 96. 107–152.

Scheer, Tobias & Péter Szigetvári. 2005. Unified representations for the syllable and stress. *Phonology* 22. 37–75.

Siptár, Péter & Miklós Törkenczy. 2000. *The phonology of Hungarian*. Oxford: Oxford University Press.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT.

Takahashi, Toyomi. 1999. Constraint interaction in Aranda stress. In S. J. Hannahs & Mike Davenport (eds.) *Issues in phonological structure: Papers from an international workshop*, 151–181. Amsterdam & Philadelphia: John Benjamins.

Topintzi, Nina. 2006. Moraic onsets. Ph.D. dissertation, University College, London.

Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12. 85–129.

55 Onsets

NINA TOPINTZI

Onsets are obligatory in the most typical syllable found cross-linguistically, the consonant–vowel (CV) syllable, and as such, are found ubiquitously across languages. This chapter explores various aspects of onsets, covering much of their structural, segmental, and suprasegmental behavior. Using empirical data as a point of departure, various stances and theoretical views will be addressed on a number of issues. These include the presence of the onset in unmarked CV syllables (§1), onset clusters and the role of sonority in their formation (§2), and the structure and representation of the onset within the syllable (§3). The focus will then shift to the onset's often disregarded role in suprasegmental phonology with reference to several weight-based phenomena (§4). The chapter closes by briefly reviewing approaches that tackle the onset–coda asymmetry (§5).

1 Onsets in unmarked syllables

Most phonologists agree that the most unmarked syllable universally is a CV syllable (Jakobson 1962: 526; CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE), i.e. a syllable that consists of a nucleus and a preceding consonant, the onset. When the onset consists of a single segment then it is simplex; when it contains a consonant cluster then it is complex. The present section deals with the former.

Evidence for the unmarkedness of CV syllables comes from a variety of sources. First, CV syllables exist in all languages (unlike other syllable types, which only occur in some) and indeed there may be languages whose sole syllable type is CV, e.g. Hua (Blevins 1995) or Senoufo (Zec 2007). While it is the case that every language will have CV syllables, it is not equally true that every syllable in a language will have an onset. Unlike Totonak and Dakota (and of course Hua and Senoufo), where onsets are obligatory, in many other languages they are optional, e.g. Greek, English, and Fijian (Zec 2007).

The naturalness of CV syllables is also indicated by the fact that they are the first syllables produced by children during the initial stages of language acquisition (CHAPTER 101: THE INTERPRETATION OF PHONOLOGICAL PATTERNS IN FIRST LANGUAGE ACQUISITION).

(1) *CV outputs by a Dutch child at age 1;5,2* (Levelt *et al.* 2000)

/pus/	[pu]	'cat'
/klar/	[ka]	'finished'
/oto/	[toto]	'car'
/api/	[tapi]	'monkey'

As Buckley (2003) shows, however, children's initial productions may also involve VC syllables. Importantly though, these never seem to arise independently, i.e. without CV syllables also being present in the language.

The dominance of CV syllables is seemingly contradicted by Arrernte (also known as Aranda; Breen and Pensalfini 1999), Barra Gaelic, and Kunjen – especially its dialect Oykangand – whose syllables are claimed to be of the VC type (with extra codas if need be) and not of the CV type (Blevins 1995 and references therein). These cases are rather weak, however, since for the most part alternative explanations that actually make use of the CV syllable type have been proposed. For instance, Blevins (1995: 230–231) observes that in Kunjen, aspiration only appears prevocalically. In principle, this could be understood as occurring either syllable-initially or syllable-finally, but empirical facts suggest that only the former analysis is viable. If aspiration were to apply syllable-finally, then it should also emerge word-finally, something that never occurs. The facts are thus only compatible with syllabification in the onset. Perhaps the strongest argument in favor of the existence of CV syllables, though, comes from a rule of utterance-initial reduction that deletes initial onsetless syllables, presumably as a means to achieve more well-formed onsetful syllables, as in (2).

(2) *Oykangand reduction in utterance-initial position* (Sommer 1981: 240)

unreduced	*reduced*		*deleted material*
igigun	gigun	'keeps going'	[i]
amamaŋ	mamaŋ	'mother (voc)'	[a]
uŋgul	gul	'there'	[uŋ]

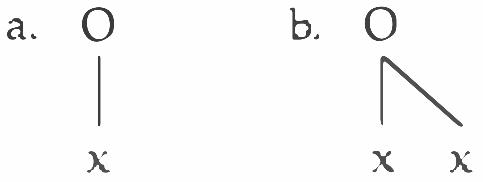
2 Complex onsets

As well as simplex onsets, onsets can also be complex, usually composed of two segments and hence considered maximally binary (Blevins 1995; Morelli 1999; Baertsch 2002; among many others), as in Greek ['tre.xo] 'I run', ['pe.tra] 'stone', ['vli.ma] 'missile', or ['tu.vlo] 'brick'. Longer sequences such as [str] or [spl] are also commonly allowed, as in English [streɪ] *stray* or [splɪt] *split*, but usually these are not considered to exceed the binarity maximum, as there is evidence that the [s] here is not part of the onset (see CHAPTER 38: THE REPRESENTATION OF SC CLUSTERS).

Yet in some work, the existence of complex clusters is denied altogether. For example, Lowenstamm (1996) and Scheer (2004) claim that all surface syllable types are subsumed under the CV matrix with the addition of empty positions, e.g. English [dØ][ri][mØ] *dream*. Duanmu (2008) interprets complex onsets such as *pl, fr, kl, kr* as complex sounds under a single timing slot, on the assumption that such sounds are possible if the articulatory gestures of two sounds can overlap (CHAPTER 54: THE SKELETON).

Most phonological models, however, allow complex onsets and provide relevant analyses to account for them. In government phonology (van der Hulst and Ritter 1999; Kaye 2000), for example, binarity is explicitly integrated within the model through the Binarity Theorem (Kaye 1990, 2000), which states that constituents cannot dominate more than two positions, so that onsets may either exhibit single association to a skeletal point (3a) or be maximally binary branching (3b).

(3) *Onsets within government phonology* (van der Hulst and Ritter 1999)



More commonly, the binarity of the onset and the combinatorial possibilities among segments within it are attributed to co-occurrence restrictions between adjacent segments (Clements 1990; Zec 2007: 164). In fact, a number of proposals subscribe to the idea that onset syllabification – like the other components of the syllable – is governed by sonority considerations (e.g. Hooper 1976; Steriade 1982; Selkirk 1984; Clements 1990; among others). Briefly, in this approach, more sonorous segments are preferred toward the center of the syllable, whereas less sonorous ones make better syllable margins, i.e. onsets and codas (Clements 1990).[1] Despite certain objections to sonority (see below; and also Parker 2002; CHAPTER 49: SONORITY), its importance for phonological theory is generally acknowledged (Steriade 1982; Selkirk 1984; Clements 1990; Rice 1992; Kenstowicz 1994; Zec 1995). One fairly standard version of the sonority hierarchy is shown below (after Clements 1990).

(4) *Sonority scale* (>= more sonorous than)

vowels > glides > liquids > nasals > obstruents[2]

One principle that makes use of this scale is the Sonority Sequencing Principle (SSP; Clements 1990), which states that the sonority profile of a syllable must be such that sonority rises sharply toward the peak and gradually lowers after it.

Evidence for the SSP comes from various sources. One example is Imdlawn Tashlhiyt Berber (e.g. Dell and Elmedlaoui 1985), known for its long sequences of consonants. Indeed, there may be words that consist of no vowel at all, e.g. [tftkt] 'you suffered a sprain'. These seemingly highly complicated strings can, however, be easily analyzed if one utilizes the SSP, plus a few other assumptions. Bearing in mind that in Imdlawn Tashlhiyt Berber: (i) any segment can be a syllable nucleus, (ii) onsetless syllables are only allowed word-initially, (iii) codas may appear word-finally, and (iv) complex onsets are banned, the following examples are syllabified in such a way that the nucleus of each syllable comprises a sonority peak.

1 For more detailed discussion on the Sonority Sequencing Principle and the Minimal Sonority Distance, see CHAPTER 49: SONORITY.

2 For a discussion of other variants see Parker (2002).

(5) *Imdlawn Tashlhiyt Berber syllabification*

/ut-x-k/	[u.tχ̩k]	'I struck you'
/rks-x/	[r̩.ks̩x]	'I hid'
/t-msx-t/	[tm̩.sχ̩ːt]	'you have transformed'

Additional evidence for the SSP comes from onset cluster simplification processes, as in Sanskrit (see Steriade 1988 and CHAPTER 119: REDUPLICATION IN SANSKRIT for relevant data) or Attic Greek (Steriade 1982), whereby C_1C_2 onset strings are reduced to simplex onsets in reduplication. Notably, the surviving C is the least sonorous one, resulting in a more abruptly rising slope toward the nucleus. Similar facts arise in child speech (CHAPTER 101: THE INTERPRETATION OF PHONOLOGICAL PATTERNS IN FIRST LANGUAGE ACQUISITION), as is evident in the outputs of an English-learning girl aged 2;9 reported on by Gnanadesikan (1995).

(6) *Cluster simplification to the least sonorous consonant*

clean	[kin]
snow	[so]
friend	[fɛn]
sky	[kaj][3]

Not all languages admit the same inventory of complex onsets. It is generally held to be true that the larger the distance in sonority between C_1 and C_2, the more well-formed the onset cluster. Thus, obstruent (O) + glide (G) clusters are highly favored, followed by O + liquid (L), O + nasal (N), and so on. Onset clusters preferably satisfy a *Minimal Sonority Distance* restriction in order to be allowed in a language (Vennemann 1972; Hooper 1976; Steriade 1982; Selkirk 1984; Baertsch 2002). In Bulgarian, no distance at all is necessary, thus all of OL, NL, ON, LL, NN, and OO clusters are admitted (Zec 2007); in other languages, different degrees of Minimal Sonority Distance are applicable: in Chuckchee, only OL, NL, and ON clusters are well-formed (Levin 1985); in Spanish, only OL onset clusters (Baertsch 2002); and in Huariapano (Parker 1994), only OG clusters.

In a sense, Minimal Sonority Distance generates the expectation that if a language allows C_1C_2 onset clusters where C_2 is of sonority X, then it should also admit onset clusters with a C_2 whose sonority is higher than X. But as we have just seen, this is not always the case: e.g. Spanish, which bans *OG clusters. To make things worse, many languages also allow sonority plateaus and even reversals. For example, Greek plateaus like [kt], [fθ], and [vɣ] are tolerated, as in [ktirio] 'building', [akti] 'coast', [fθiro] 'impair', [afθonos] 'abundant', [vɣazo] 'remove', [avɣo] 'egg'. Russian also permits reversals, e.g. [rtut] 'mercury' and [lvov] (city name) (Gouskova 2001), which, however, are often considered not to be complex

[3] Under the assumption that [sk] is a complex onset, the fricative [s] must be more sonorous than the stop [k] (cf. Dell and Eldmedlaoui 1985; de Lacy 2006; among others). In the sonority hierarchy I have adopted here, this distinction is not made. On the other hand, a difference in sonority of fricatives as opposed to stops would yield incorrect results in other accounts, e.g. Kreitman (2006). If [s], however, is not part of the onset (cf. CHAPTER 38: THE REPRESENTATION OF SC CLUSTERS), this issue does not arise in the first place.

onsets; rather, the segment(s) violating the SSP can be realized as syllabic, e.g. [r̩.tut], or even extrasyllabic, attaching to some higher level of prosodic structure, e.g. the foot or prosodic word (see CHAPTER 40: THE FOOT and CHAPTER 38: THE REPRESENTATION OF SC CLUSTERS for more discussion). Such data partly explain why the validity of sonority is sometimes contested.

Other objections to sonority include the lack of a clear way to phonetically define and measure it, and its inability to explain the frequent ban on sequences of the type *ji*, *wu*, *bw*, or *dl* (quite likely an Obligatory Contour Principle effect). Some researchers have therefore gone as far as to discard sonority. For example, Ohala (1990) and Harris (2006) claim that attested sequences in languages can be best captured through the perceptual distance between neighboring sounds in terms of a number of different acoustic properties, including amplitude, periodicity, spectral shape, and fundamental frequency (F0) (Ohala 1990: 334). As Ohala (1990: 334–335) admits, however, this view explains which sequences should be found in languages, but does not explain how and why they are grouped into syllables.

This is perhaps why – despite criticism – sonority still remains highly influential in current work on syllabification (cf. Baertsch 2002; Gouskova 2004; Zec 2007; among many others). But there is yet another possibility. Rather than completely endorsing or abandoning sonority, we can accept it, but loosen somewhat the predictions and generalizations it makes. Berent *et al.* (2007) put forward a proposal along these lines. In particular, they suggest a more flexible version of sonority-based generalizations regarding the profile of onset clusters. They state that:

> In any given language:
> (a) The presence of a small sonority rise in the onset implies that of a large one.
> (b) The presence of a sonority plateau in the onset implies that of some sonority rise.
> (c) The presence of a sonority fall in the onset implies that of a plateau. (Berent *et al.* 2007: 594)[4]

On this view, Spanish is no longer problematic (since OL clusters involve high sonority, there is no reason that there should be OG clusters too), and the plateaus of Greek are expected, given that it also has sonority rises, while Russian has falls only because it also has plateaus. More generally, Berent *et al.* (2007) test the statements above against the sample of Greenberg (1978) and find that they overwhelmingly hold true typologically.

Other typological surveys on onset clusters also tend to employ sonority, usually with some modification or enrichment of the theory. For instance, Morelli (2003) investigates the patterns of obstruent onset clusters and proposes implicational relationships between them, as schematized in (7), where fricative + stop (FT) clusters are the least marked, TT the most marked, and TF somewhere in between. FF clusters merely imply the existence of FT, without further implicational relationship with other clusters.

[4] Berent *et al.* (2007) seem to adopt Greenberg's (1978) characterization of small and high sonority. High-sonority rises are OL clusters; low-sonority rises are NL and ON; plateaus are OO, and falls are LN and NO clusters.

(7) *Implicational relationships between obstruent onset clusters* (Morelli 2003)[5]

$$\begin{array}{r} TT \\ \Downarrow \\ TF \\ \Downarrow \\ FF \Rightarrow FT \end{array}$$

Kreitman (2006) focuses on sonorant (S) and obstruent (O) clusters and proposes the implicational hierarchy SO ⇒ SS ⇒ OO ⇒ OS, with OS clusters being the most unmarked, and SO ones the most marked. These are respectively the most and least favored clusters as far as sonority is concerned. SS and OO clusters involve sonority plateaus, but do not randomly appear in languages as one would expect; instead, the presence of SS systematically implies OO. To account for this fact, Kreitman points to the increased salience of obstruents as opposed to sonorants (cf. Ohala 1983: 193). Since obstruents are considered to carry more information, due to their acoustic form, they are easier to distinguish from non-obstruents. Thus, combinations between obstruents should be perceptually favored over those between sonorants.

What all these studies highlight is that removing sonority from the equation is not useful; rather it seems that consideration of other factors, e.g. the role of perceptual salience, may enhance the role of sonority conceptually and improve its empirical coverage.

3 The status of the onset within the syllable

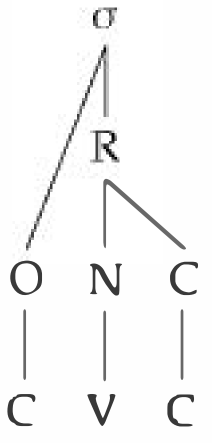
Moving away from the principles that regulate onset syllabification, let us consider the representation of the onset within the syllable. Various models of the syllable have been proposed throughout the years (see Blevins 1995; van der Hulst and Ritter 1999 for overviews; see also CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE), which due to lack of space will not be discussed here in detail. Nonetheless, reference will be made to those that are especially relevant to onsets. Broadly speaking, we can identify two major theories: (i) those that distinguish between onsets and rimes (Pike and Pike 1947; Kuryłowicz 1948; Fudge 1969; Selkirk 1982; Levin 1985; Kaye *et al.* 1990; Blevins 1995), and (ii) moraic models that do away with the rime, i.e. the nucleus + coda string, as a separate constituent (Hyman 1985; Hayes 1989; Morén 2001).

3.1 *Onset–rime models*

No single version of the onset–rime model is available, and there are significant divergences between models. For instance, Fudge (1969) accepts the syllable as a constituent, whereas Kaye *et al.* (1990) explicitly do away with it, but nonetheless treat the onset and rime as "an inseparable package" (van der Hulst and Ritter 1999: 23).

[5] Inclusion of sC clusters among the FT clusters and their treatment as onset clusters, at least word-initially, is quite problematic for Morelli, however, in light of evidence showing how sC clusters differ from true branching onsets in various ways (see CHAPTER 38: THE REPRESENTATION OF SC CLUSTERS).

(8) *A typical representation of the onset–rime model* (Blevins 1995)



Specific syllable models make different claims about constituenthood. For instance, Blevins (1995) essentially only recognizes the rimal constituent and sees no strong argument for an onset constituent – and for that matter, a coda constituent. For government phonology (van der Hulst and Ritter 1999; Kaye 2000), on the other hand, onsets, nuclei, and rimes are constituents.

The basic argument for the rime hinges on the idea that co-occurrence restrictions are always more likely to occur between nuclei and codas, rather than between either onsets and nuclei or onsets and codas. The strongest argument for the rime though comes from weight facts (Blevins 1995; van der Hulst and Ritter 1999: 23). Consider stress, for example. As is well known, in many languages heavy syllables attract stress in contrast to light syllables (e.g. Hopi; Jeanne 1982). Importantly, heaviness implies a binary rime, $[VV]_R$ or $[VC]_R$, or both, depending on the language. Since the presence of onsets is disregarded in such an evaluation, it must mean that rimes form a constituent that clearly excludes the onset.

Nonetheless, each of the arguments in support of the rime has been challenged. Davis (1985) attacks the reliability of co-occurrence and phonotactic restrictions, given that those are not exclusive to nuclei and codas, but are also found between onsets and nuclei or onsets and codas. For instance, in Korean (Cho 1967), fronted vowels do not appear after labial onsets, while in Yindjibarndi (Wordick 1982), the presence of /r/ in both the onset and a coda of a syllable is banned. Another objection to the onset–rime distinction is found in Yip (2003), who claims that if it were valid, then the boundary between the two constituents should be clear and consistent, and thus segments should uniformly belong to either the onset or the rime, but not to both. English and Mandarin pre-nuclear glides, however, behave sometimes like onsets and sometimes as rimes. As for the weight effects induced by the rime, it is possible to capture them in a different manner without reference to the rime per se. This is what moraic theory does, as we will see in a moment.

Before moving on, though, it is notable that the onset–rime debate is also predominant in psycholinguistic studies that explore the onset–rime boundary in terms of implicit and explicit, i.e. non-conscious *vs.* conscious, phonological awareness. Work by Treiman (1986 and references therein) on various segmentation and substitution tasks in both adults and children suggests that there is a closer connection between VC than CV, thus offering support for the onset–rime boundary. In the same vein, Uhry and Ehri (1999) show that English-speaking kindergarten children preferred to keep VC, rather than CV, intact during segmentation. The opposite result, however, was found by Lewkowicz and Low (1979).

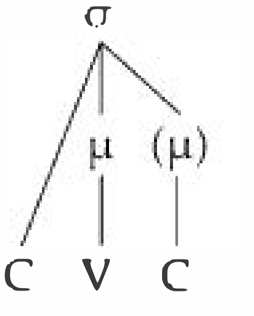
More recently, Geudens and Sandra (2003), in a series of four experiments on Dutch-speaking pre-readers and beginning readers, found no support for the onset–rime boundary. Importantly, they applied strict criteria regarding the selection of items under investigation, such that they could control for distributional and sonority effects. In particular, they used items of different sonority equally often and found that syllables with obstruents were easier to perceive and segment than syllables with sonorants (2003: 172); see also CHAPTER 8: SONORANTS. The influence of sonority may in fact explain some of the findings of previous studies, such as Schreuder and van Bon's (1989) finding that Dutch first-graders break up a CV string more easily than a VC one. In their study, sonorants were mainly used, but sonorants undergo more vocalization in coda rather than onset position, possibly explaining why children find it harder to break them up in a VC environment rather than a CV one.

All in all, psycholinguistic experimentation also reflects contradictory evidence with regard to the onset–rime boundary debate. What this absence of consensus at the very least suggests is that the boundary dispute is well grounded.

3.2 *Moraic model*

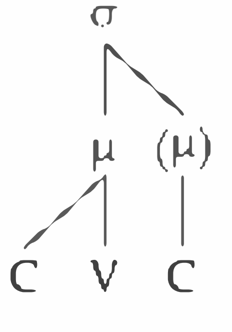
A common response to criticism against the rime has been to dispense with it as a constituent altogether and to replace it with the concept of mora. In moraic theory (Hyman 1985; Hayes 1989), only segments under – what used to be – the rime node may bear moras. Since the latter are needed independently to account for a number of phenomena related to syllable weight, the natural conclusion has been to structurally eliminate the rime from representations. The representation of a CVC syllable in this model is presented next (compare with (8)). Note that the bracket around the mora of the coda indicates that this may be moraic or not on a language-specific basis (cf. Weight-by-Position; Hayes 1989).

(9) *Moraic model* (Hayes 1989)



Within moraic theory, there is no definite agreement as to where exactly the onset associates to. According to Hayes, it directly adjoins to the syllable as in (9). For Hyman (1985), Itô (1989), and Buckley (1992), though, it attaches to the following mora, as in (10).

(10) *Onset association* (Hyman 1985)



In both these versions of moraic theory, the onset is not recognized as a constituent. This is much more clearly shown in (9), where it directly links to the syllable node, but it is visible even in (10), since the mora is shared between the onset and the nucleus.

While Hayes's representation is the most widely employed, there is nevertheless some evidence for (10). Katada (1990) describes the Japanese chain language game *shiritori*, in which players say a word that must begin with the final mora of the previous player's word. If the word ends in a CV syllable, as in [tu.ba.me] 'swallow', then the next word can be something like [me.da.ka] 'killfish'. If the word ends in a long vowel, then the last mora is the second half of the vowel, to the exclusion of the first half, as well as the onset. Thus [bu.doo] 'grapes' can be followed by [o.ri.ga.mi] 'folding paper' but not by *[doo.bu.tu] 'animal'. Importantly, a word like [riN.go] 'apple' (where N is a moraic nasal) cannot be followed by *[o.ri.ga.mi], but must begin with [go]. This is easily explained if the final mora in [go] also associates to the onset, as claimed by (10), rather than linking directly to the syllable (9). The game ends if the final mora cannot form a proper onset, as happens when it is a moraic nasal, e.g. [ki.riN] 'giraffe'.

Since the moraic model identifies no rime constituent, it bypasses the problems faced by the onset–rime model with regard to the extension of co-occurrence restrictions beyond the rimal node, as well as the absence of a clear boundary between the onset and the rime. Superficially, however, it does equally well as the onset–rime model in accounting for syllable weight, simply by stating or – more accurately, stipulating – that moras are strictly limited to nuclei and codas. But even this assertion has been contested. Work by Hajek and Goedemans (2003), Gordon (2005), and Topintzi (2006, 2010) has shown that there is good evidence for the existence of onset weight. We explore this issue next.

4 The suprasegmental phonology of onsets

Contrary to popular belief, onsets do seem to be prosodically active, albeit in a limited number of languages. Their effects become evident in a range of phenomena, including stress, compensatory lengthening, gemination, word minimality, and tone. This section examines the relevant data and theoretical issues that stem from them.

4.1 Stress

Of all these phenomena, onset-sensitive stress has received the most extensive attention. In brief, three patterns are attested: (i) onset effects due to the presence of an onset, (ii) onset effects due to the quality of an onset, and (iii) patterns (i) and (ii) combined.

Starting from (i), we find that in a number of languages onsetful syllables attract stress more than onsetless ones. Languages of this type include Arrernte (Strehlow 1944), Alyawarra (Yallop 1977), and other Australian languages, such as Lamalama, Mbabaram, Umbuygamu, Umbindhamu, Linngithig, Uradhi, Kuku-Thaypan, Kaytetj, and Agwamin (most of them are Cape York and Arandic languages; see Davis 1985, Goedemans 1998, and Blevins 2001 for more details). Beyond Australia, this pattern is attested in unrelated languages of North and

South America, Iowa-Oto (Robinson 1975), Banawá (Buller *et al.* 1993), and Juma (Abrahamson and Abrahamson 1984).[6]

In Arrernte, C-initial words receive stress on the first syllable (11a), but V-initial ones have stress on the second syllable (11b). One exception is disyllabic words, where stress is word-initial regardless of whether the word begins with a vowel or a consonant (11c). This is probably attributed to Arrernte's avoidance of final stress or preference for creating binary feet, as the lack of final secondary stress in words like *[a('ralka)(ˌma)] reveals.

(11) *Arrernte stress* (Strehlow 1944)

a. *consonant-initial words of three or more syllables*

'ra:tama	'to emerge'
'kutunˌgula	'ceremonial assistant'
'lelanˌtinama	'to walk along'

b. *vowel-initial words of three or more syllables*

er'guma	'to seize'
a'ralkama	'to yawn'
u'lambuˌlamba	'water-fowl'

c. *words of two syllables (C- or V- initial)*

'ilba	'ear'
'a:twa	'man'
'kala	'already'
'gura	'bandicoot'

A common denominator is that stress may shift – albeit very locally – to dock on a syllable with an onset. This is not the only possibility, however. In other languages, the stress location remains constant, but if it falls on an onsetless syllable, this acquires an onset. Consider Dutch (Booij 1995: 65). In instances of hiatus where the first vowel is /a/, a glottal stop is inserted before the second vowel, but only if this is stressed by the normal algorithm, e.g. /paɛlja/ → [pa.'ʔɛl.ja] 'paella', /aɔrta/ → [a.'ʔɔr.ta] 'aorta'. Otherwise, no insertion is applicable: /xaɔs/ → ['xa:.ɔs] 'chaos', /farao/ → ['fa:.ra.o:] 'Pharaoh'. Most analyses view this as a prominence (Smith 2005) or alignment (Goedemans 1998; Topintzi 2010) effect.

In yet other languages, the mere presence of an onset is not the issue (see Topintzi 2010: 48 for details on Karo); it is the quality of the onset that matters. This is the case in Karo (Gabas 1999) and possibly Arabela (Payne and Rich 1988). In the former, stress falls on the final syllable, except when the penultimate syllable is a better stress bearer. Better stress bearers are, in order of priority, a syllable with (i) a high tone, (ii) a nasal vowel, or (iii) a voiceless or sonorant onset. When (i) and (ii) are irrelevant, (iii) is taken into consideration and stress falls on the final syllable if the onset is a sonorant (12a) or voiceless (12b) or a voiced obstruent preceded by another voiced obstruent onset (12c).

[6] However, the case of Juma should be treated with caution, because only a handful of data are available and because it is possible to re-analyze it. In particular, words like [pe'jikɔ'pia] 'bird (sp.)' may be argued to contain a final diphthong, i.e. [pe.'ji.kɔ.'pia], rather than a sequence of heterosyllabic vowels, i.e. [pe.'ji.kɔ.'pi.a], which would lend support to the onset effect. Interestingly, Juma is the sole language where the effect appears at the right edge of the word and not at the left. This may perhaps be an additional indication that it is not truly onset-sensitive.

(12) *Karo final stress and onset voicing* (Gabas 1999: 14, 39–41)[7]

a. *final syllable with sonorant onset*
kɔ'jɔ 'crab'
ja'ʔ'mᵇɔ 'yam (sp.)'
kiɾi'wɛp˺ 'butterfly'

b. *final syllable with voiceless onset*
pa'kːɔ 'fontanel'
maʔ'pɛ 'gourd'
kuɾuʔ'cu 'saliva'

c. *final and prefinal syllables with voiced obstruent onsets*
kiɾi'bɔp˺ 'frog (sp.)'
miɾi'ɾij 'toad (sp.)'

Stress, however, falls on the penult if the final syllable has a voiced obstruent onset and the previous one does not, indicating the stress-attracting nature of the voiceless obstruents and the sonorants in this language.

(13) *Karo penult stress and onset voicing* (Gabas 1999: 14, 39–41)

'jaba	'rodent (sp.)'	'pibɛʔ	'foot'
'wɛɾɛ	'frog'	'kaɾo	'macaw'
'mɔga	'mouse'	iʔ'cɔgɔ	'quati (sp.)'

Nonetheless, other cases where stress is seemingly sensitive to the onset quality have been shown to be much less robust or even wrong. One example of the latter arises in Mathimathi, where stress is normally word-initial unless attracted by the second syllable when it begins with a coronal onset. Davis (1988) attributes this to genuine onset-sensitivity. Gahl (1996), on the other hand, shows that another account is more plausible, namely one that considers Mathimathi stress to be morphologically based. She claims that stress is located on the last stem syllable of the word (or better, last stem vowel). Stems are generally monosyllabic or bisyllabic. It so happens that apparent stress shift appears on stems of the type $C_1VC_2VC_3$, where the medial consonant is invariably coronal (Gahl 1996: 329). Evidence for Gahl's analysis comes from monosyllabic C_1VC_2 stems, where C_2 is again coronal. Addition of a suffix to such stems renders C_2 an onset of the second syllable. If Davis were right, then stress here should also be peninitial. However, it is initial, as predicted by Gahl's morphological account; cf. peninitial stress in bisyllabic stems such as [ˌgu.'ra.g+i] 'sand' *vs.* initial stress in monosyllabic stems such as ['wa.d̪+a.d̪+a] 'to come'. In both cases, C_2 is coronal. Thus, re-examination of the facts in light of morphological considerations may reveal the lack of true onset-sensitive effects (see also Nanni 1977 on the English suffix *-ative* or Davis *et al.* 1987 on Italian infinitives).

A final pattern that emerges involves the combination of true onset-presence and onset-quality effects. A well-known example is Pirahã (Everett and Everett 1984; Everett 1988), an Amazonian language where codas are banned. Onsetless light syllables (V) do not occur, and stress may only dock on one of the three final

[7] Note that [ɾ] in Karo behaves like [d], which is otherwise missing from the inventory (Gabas 1999: 12).

syllables of the word. The weight and stress hierarchy the language motivates is: PVV > BVV > VV > PV > BV (P = voiceless; B = voiced). In particular, VV nuclei attract stress more than V ones (14c), and voiceless onsets have the same effect as opposed to voiced ones (14a) and (14d). Crucially, and unlike Karo, Pirahã 'voiced' consonants also include sonorants, which appear as allophones of voiced stops, e.g. /b/ may surface as [b], [m], or the bilabial trill [ʙ]. Consequently, in this language, only voiceless obstruents attract stress. Between equally heavy syllables in terms of nucleic weight, onsetful ones attract stress over onsetless (14b). Finally, if there is more than one equal contender for stress, the rightmost one receives it (14e).

(14) *Pirahã examples* (Everett and Everett 1984; Everett 1988)

a.	PVV > BVV			
	'káo.bá.bai	'almost fell'		(1988: 239)
	pa.'hai.bií	'proper name'		(1984: 708)
b.	BVV > VV			
	'bii.oá.ii	'tired (lit.: being without blood)'		(Everett, p.c.)
	poo.'gáí.hi.aí	'banana'		(1984: 709)
c.	VV > PV			
	pia.hao.gi.so.'ai.pi	'cooking banana'		(1984: 710)
d.	PV > BV			
	ti.'po.gi	'species of bird'		(1984: 710)
	'ʔí.ʙo.gi	'milk'		(Everett, p.c.)
e.	*rightmost heaviest stress*			
	ho.áo.'íi	'shotgun'	*ho.'áo.íi	(1984: 710)
	ti.'po.gi	'species of bird'	*'ti.po.gi	(1984: 710)
	paó.hoa.'hai	'anaconda'	*paó.'hoa.hai *'paó.hoa.hai	(1984: 707)

What is common to all these examples is that the voiceless obstruent onsets systematically attract stress, contrary to the voiced obstruent ones. Various analyses have been offered to account for the Pirahã facts (and many fewer for Karo). These are examined in Topintzi (2010). Some make use of the increased prominence of onsetful syllables and voiceless onsets over onsetless syllables and voiced onsets respectively (Everett and Everett 1984; Hayes 1995; Goedemans 1998; Smith 2005). Some treat certain onsets as weightful and some as weightless (Topintzi 2006, 2010), and others offer a mixed system that utilizes weight but sees it as a function of prominence (Gordon 2005). Due to space limitations, these proposals will not be reviewed here. However, there is one important empirical argument that favors the onset weight approach, namely the existence of other phenomena beyond stress that are weight-related and influenced by onsets.

4.2 *Compensatory lengthening, geminates, and word minimality*

An explicit prediction of the onset–rime and the moraic models is that onsets will never participate in weight-related processes. For the former, this is because onsets are excluded from the prosodic hierarchy (van der Hulst and Ritter 1999: 31). For the latter, it is because onsets never bear moras (Hayes 1989). However, both assertions are entirely stipulative and subject to modifications given the existence of counterevidence.

First, consider compensatory lengthening (CHAPTER 64: COMPENSATORY LENGTHENING), a phenomenon widely utilized in support of standard moraic theory. In standard moraic theory (Hayes 1989), it is predicted that onsets will neither induce compensatory lengthening (through deletion) nor undergo it (through lengthening). Yet several cases of both types have been reported.

In Samothraki Greek, the onset /r/ deletes and generally leads to lengthening of the following vowel, e.g. /'rema/ > ['eːma] 'stream', /'ruxa/ > ['uːxa] 'clothes', /'ðedru/ > ['ðeduː] 'tree', /kra'to/ > [kaː'to] 'I hold' (Katsanis 1996: 50–51). Onondaga (Michelson 1988) is somewhat similar, although /r/-deletion leads to lengthening, whether it is in an onset or a coda originally. Numerous other examples have been reported (Rialland 1993; Beltzung 2007), all of which, however, are highly morphologized. For instance, in Romanesco Italian, the initial /l/ of the definite article and of the object clitic /lo la li le/ optionally deletes (Loporcaro 1991: 280), causing lengthening of the unstressed vowel that follows, e.g. [lo 'stupido] > [oː 'stupido] 'the stupid (MASC)' or [la 'bruːʃo] > [aː 'bruːʃo] 'I burn her'. Beyond this environment, such compensatory lengthening does not appear. Analogous effects are observed in Anuak/Anywa, Lango, Gyore, Turkana, and Ntcham (see Beltzung 2007; Topintzi 2010 and references therein).

Nonetheless, one could question the validity of this approach in terms of onset weight structure and instead provide a more phonetic explanation, as done by Kavitskaya (2002). She observes that vowels in CVC syllables are phonetically longer when followed by certain consonants whose transitions can be misheard as part of the vowel (i.e. sonorants, approximants). On deletion of such consonants, the 'excess' length of the preceding vowels can be phonologized, so that listeners reinterpret them as phonemically longer. Thus vowels are reinterpreted by listeners as phonemically longer. This approach also extends to compensatory lengthening induced by onsets, but only works when highly sonorous consonants are deleted. In principle, this is appropriate for some of the cases, e.g. Samothraki Greek or Romanesco Italian, but is nevertheless problematic. For instance, it cannot explain why the same phonologization of length has not occurred with regard to the Samothraki coda *r*, especially since this is the prototypical position for compensatory lengthening. More troublesome, though, is the inability to account for cases like Ntcham, where the onset that is lost is the highly *non*-sonorous /k/.

More strikingly, onsets also can serve as the target of compensatory lengthening. This means that a segment deletes and the preceding onset lengthens, i.e. geminates in order to compensate for its loss.[8] For instance, Pattani Malay (Yupho 1989; Topintzi 2008) contrasts singletons and geminates in onsets, but only word-initially (on initial geminates see CHAPTER 47: INITIAL GEMINATES), e.g. [ˌbu'wɔh] 'fruit' *vs.* ['bːuˌwɔh] 'to bear fruit', [ˌɟa'lɛ] 'road' *vs.* ['ɟːaˌlɛ] 'to walk' (Yupho 1989: 135). Moreover, it exemplifies a case of compensatory lengthening (Michael Kenstowicz, personal communication). In instances of free variation, one variant involves loss of the word-initial syllable and gemination of the second

[8] This characterization is unavoidably linked to a broader discussion of what exactly constitutes a geminate. Briefly, the debate relates to whether geminates are inherently moraic (i.e. heavy) or involve double linking to higher structure (i.e. long). This issue is thoroughly examined in CHAPTER 37: GEMINATES and CHAPTER 47: INITIAL GEMINATES.

onset, as in e.g. [buwi] ~ [wːi] 'give', [sidadu] ~ [dːadu] 'police', [pɨmatɔ] ~ [mːatɔ] 'jewelry' (Yupho 1989: 130).

That these geminates are moraic is supported by another fact of the language, namely stress. Primary stress is word-final, unless the word begins with a geminate, in which case it shifts to the initial syllable (the other syllables receive secondary stress). We can easily understand this effect by claiming that the syllable hosting a geminate is bimoraic and therefore heavy, and as such, attracts stress in preference to monomoraic syllables.

(15) *Stress in Pattani Malay* (Yupho 1989: 133–135)

a. *words lacking geminates*

ˌa'lɛ	'road path'
ˌda'lɛ	'in, deep'
ˌmãˌkɛ'nɛ̃	'food'

b. *words with initial geminates*

'mːaˌtɔ	'jewelry'
'ɟːaˌlɛ	'to walk'

As well as Pattani Malay, Trukese provides evidence that onset geminates are moraic (see also CHAPTER 37: GEMINATES for discussion). First, Trukese words are minimally (C)VV, e.g. [maa] 'behavior', [oo] 'omen', or C_iC_iV, i.e. a geminate plus a short vowel, e.g. [tto] 'clam (sp.)', [ʧʧa] 'blood' (Davis and Torretta 1998; Muller 1999).[9] CVC and CV words are not allowed (Davis 1999), thus singleton codas contribute no mora (Muller 1999). Presumably, minimality is satisfied by bimoraic words, provided of course that geminates add a mora to their syllable. An additional process of compensatory lengthening following the deletion of the final mora in a word corroborates the moraicity of onset geminates (CHAPTER 37: GEMINATES).

Various proposals within the standard moraic theory tradition have been put forward to account for initial moraic geminates (Davis 1999; Curtis 2003), common to which has been the lack of any association between the geminate's mora and the onset, in line with a major tenet of the theory, namely the ban on onset moraicity. Crucially, to achieve this effect, these approaches link the geminate's mora to some position other than the onset, which is made possible by the double linking commonly assigned to geminates (see CHAPTER 37: GEMINATES). But this solution is not available in cases of moraic initial consonants that are singletons rather than geminates. Such cases exist.

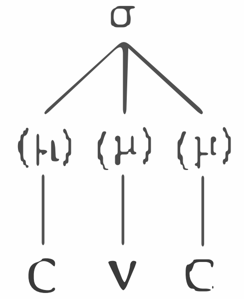
In Bella Coola (Bagemihl 1998) the minimality criterion is fulfilled by VV, VC, and CV words, but crucially not by V words.[10] Topintzi (2006, 2010) argues that the easiest way to uniformly understand these data and place them alongside the root-maximality facts of the language – that make reference to mora structure – is by stating a bimoraic word minimum and by allowing onsets to bear moras.

[9] Many languages impose a minimum size for words to be well-formed. Commonly, words are required to be at least bimoraic (C)VV as in Ket or Mocha, or (C)VV/(C)VC as in English or Evenki (Gordon 2006), or bisyllabic, e.g. Pitta-Pitta (Hayes 1995: 201).

[10] In fact, minimal words with two unsyllabified consonants CC are also allowed. Evidence for the existence of unsyllabified consonants would take us too far afield; see Bagemihl (1998) and Topintzi (2006) for details.

To capture these facts, Topintzi (2006, 2010) puts forward a flatter syllable structure (reminiscent of Davis 1985), whereby all syllable constituents come in either moraic or non-moraic versions. This is hardly surprising for codas; cf. moraic codas in Latin, Delaware, English, Kiowa, and Turkish *vs.* non-moraic ones in Wargamay, Lenakel, Eastern Ojibwa, and Khalkha Mongolian (e.g. Hayes 1995; Zec 1995, 2007; Morén 2001). The claim extends to onsets too, e.g. moraic for onset geminates in Trukese or voiceless obstruents in Pirahã *vs.* non-moraic in a host of other languages. Applying the same distinction to nuclei is also not too far-fetched, as it has been suggested that they can occasionally be weightless, for example in Malagasy (Erwin 1996), Kabardian (Peterson 2007), Alamblak (Mellander 2003), and Chuvash and Mari (Hyman 1985). The following representation illustrates the proposal outlined by Topintzi (2006, 2010).[11]

(16)



Even with this modification, though, moraic theory faces problems when it encounters data such as those in Seri and Kikamba (Roberts-Kohno 1995) and Onondaga and Alabama (Broselow 1995 and references therein), and French *h-aspiré* (Boersma 2007 and references therein). In Seri (Marlett and Stemberger 1983; Crowhurst 1988; Broselow 1995), the distal prefix [jo-] attaches to either C- or V-initial stems. In the former, nothing remarkable occurs (17a), but in V-initial stems, things become more complex. In general, when the first vowel of the stem is low back /a/ or low front /æ/ the prefix vowel deletes and compensatory lengthening results (17b). But in some specific stems, no deletion (and consequently no compensatory lengthening) occurs. Instead, a hiatus context is created (17c).

(17) *Seri distal forms*

	stem		*distal*	
a.	*C-initial stems*			
	-mækæ	'be lukewarm'	jo-meke	
	-pokt	'be full'	jo-pokt	
b.	*general pattern of /a, e/-initial stems*			
	-ataχ	'go'	joː-taχ	
	-æmæ	'be used up'	joː-me	
c.	*exceptional pattern of /a, e/-initial stems*			
	-amwx	'be brilliant'	jo-amwx	*joː-mwx
	-ænx	'play stringed instrument'	i-jo-enx	*i-joː-nx

According to Crowhurst (1988), these data support a mixed representation that includes both X slots and moras (CHAPTER 54: THE SKELETON). The idea is that the stems in (17c) are underlyingly specified with an empty slot in the onset, whose

[11] Simultaneous moraicity on all three positions is presumably attested in Karo (see Topintzi 2010: 49).

net effect is to block deletion (and compensatory lengthening), because of its intervening position between the two vowels. Effectively, then, (17c) acts as if it were a C-initial stem (17a). Data of this type can also be easily accommodated in government phonology (Kaye *et al.* 1990), which by its nature allows reference to empty positions.

It is, however, not entirely clear that Seri cannot be accommodated by moraic theory alone (especially if onsets may bear moras). Unlinked moras appear in numerous works (cf. van Oostendorp 2005; Topintzi 2007) and are in fact suggested by Crowhurst herself. We could assume then that the input for [jo-amwx] is /jo-^{M}amwx/, where ^{M} indicates a floating mora. If on the surface this mora remains unassociated but anchored at the left edge of the stem, then it can produce the same blocking effect of deletion that Crowhurst achieves by means of an unassociated x-slot.

Even if this is feasible, it is unlikely that all similar kinds of facts will be subject to reanalysis. One solution would be to reconsider representations that simultaneously use x-slots and moras, as Crowhurst does. This idea has reappeared in Muller's (2001) Composite Model with respect to geminates, and in Vaux (2009) as a more complete model of timing. Whether such enrichment of the theory is justified remains to be seen. Alternatively, one could entertain Itô's suggestion (1989: 255 and references therein) that "the role previously played by lexically empty skeletal slots can be taken over, wholly or in part, by bare melodic root nodes."

4.3 Tone

Another phenomenon where onsets seem to be involved, albeit rarely, is tone (CHAPTER 45: THE REPRESENTATION OF TONE). Relevant cases reported include Musey (Shryock 1995) and Kpelle (Welmers 1962; Hyman 1985).

In Musey, consonants are divided into Type A (or High consonants) and Type B (or Low consonants). Type A consonants include the sonorants and the historically voiceless obstruents. Type B ones correspond to the historically voiced consonants. Both Type A and B obstruents are basically voiceless (Shryock 1995: 68–69), with Type A stops presenting longer positive voice onset time (VOT), less closure voicing, and higher F0 at the onset of the following vowel than the Type B ones.

The rightward displacement of lexical L tone when a suffix is added in (18) shows the genuine contrast between the two types of consonants as well as their tonal effects. When the lexical L tone shifts, the vowel that hosted the tone is interpreted as mid or high if the onset is Type A, but as low if the onset is Type B.

(18) *Rightward displacement of lexical L tone in Musey*

a. *cliticization of* /-na/

Type A	sà → sanà → sānà		'person'
Type B	fìù → fììnà		'goat'

b.

	subjunctive		*subjunctive with affixation*	
Type A	tò	'sweep'	tōm̀	'sweep it'
Type B	dò	'pick'	dòm̀	'pick it'

Thus, at some level of representation, the onset consonants above seem to bear tone – be it by conditioning it or by having it floating in the input – which subsequently surfaces on the neighboring vowel to the right. What is more interesting is that the tone induced depends on the quality of the consonant involved: voiceless obstruents (and sonorants, which I will come back to in a moment) cause M tone, voiced obstruents cause L tone. This fact correlates precisely with data we find in tonogenesis (cf. Vietnamese (Haudricourt 1954) or synchronically in Kammu dialects (Svantesson 1983); see also CHAPTER 97: TONOGENESIS), where the historical contrast between voiceless and voiced obstruents is neutralized in favor of voiceless obstruents and is reinterpreted by means of tone, as shown below.

(19) *Common pattern in tonogenesis*

voicing contrast; no tone		*no voicing contrast; presence of tone*
pa	>	pá
ba	>	pà

This pattern is phonetically grounded: in voiceless obstruents, the cricothyroid muscle stretches the vocal folds to obstruct vocal fold vibration resulting in vocal fold tensing, which in turn leads to a higher F0. In voiced obstruents the larynx and hyoid bone are lower and a lowered larynx results in a lower F0 (Yip 2002: 6–7; Honda 2004). In fact, depression of F0 after voiced stops is very likely universal, as Kingston and Solnit (1988b) state (CHAPTER 114: BANTU TONE). Sonorants, on the other hand, do not automatically perturb the F0 of adjacent vowels, and thus may cause either elevation or depression of the F0 (Kingston and Solnit 1988a: 276). This finding is also in line with the behavior of sonorants in onset-sensitive stress discussed above. Recall that in Karo sonorants act like voiceless obstruents in attracting stress, but in Pirahã like voiced obstruents in avoiding it.

Reviewing the vast literature on the phonological effects of the onset/tone interaction phenomenon is well beyond the goals of the present chapter (see Yip 2002; Gordon 2006; van Oostendorp 2006; Tang 2008 for relevant overviews). For our purposes and in light of the data above, it suffices to say that Musey exhibits mixed behavior. On the one hand, it has not entirely lost the voicing contrast between stops (see the discussion on Type A and B consonants) – since it retains phonetic voicing by means of short *vs.* long VOT – but is moving in that direction, as the facts above reveal; on the other hand, it has introduced tone, which is commonly associated to specific onset quality, but has not (yet?) extended this pattern throughout the system. One thing seems quite clear: onsets in Musey may act as phonological tone bearers. And as expected, voiceless obstruents produce tone raising and voiced ones tone lowering. The more neutral sonorants here pattern with the voiceless obstruents.

Along similar lines, we can understand the data in Kpelle. However, unlike Musey, Kpelle onsets act as surface tone-bearing units (TBUs). First, consider minimal pairs such as (20), where a sonorant onset can appear toneless, L-toned, or H-toned. This is hardly surprising, given the capacity of sonorants to bear any type of tone.

(20) mare-kɛ̂i 'a question'
m̀are kɛ́ 'ask him'
ḿare kɛ́ 'ask me'

Moreover, the possessive form involves an underlyingly H-toned nasal prefix for the 1st singular or a floating L tone for the 3rd singular (plus the independent processes of voicing assimilation in obstruent-initial stems and total assimilation and nasal simplification in sonorant-initial stems),[12] both of which surface on onset positions.

(21) *Kpelle onsets as TBUs* (Hyman 1985: 44)

	stem	'my'	'his/her'	
a.	*initial obstruent*			
	pólù	ḿbólù	b̀ólù	'back'
	túɛ	ńdúɛ́	d̀úɛ́	'front'
	kɔ́ɔ́	ŋ́gɔ́ɔ́	g̀ɔ́ɔ́	'foot, leg'
	fíí	ɱ́víí	v̀íí	'hard breathing'
b.	*initial sonorant*			
	lēē	ńēē	ǹēē	'mother'
	jéé	ɲ́éé	ɲ̀éé	'hand, arm'
	ma̱lóŋ	ḿa̱lóŋ	m̀a̱lóŋ	'misery'
	ɲí̱ŋ	ɲ́í̱ŋ	ɲ̀í̱ŋ	'tooth'

These examples show that sonorants and voiced obstruents may appear as surface TBUs, but the same does not hold for voiceless obstruents. This is entirely expected, given that the physical correlate of tone is F0, thus only voiced segments should be able to present it, i.e. vowels, sonorants, and voiced obstruents (Gordon 2006). The Musey data nonetheless have suggested that voiceless onsets should be allowed to be input phonological TBUs (a similar claim for Kpelle appears in Topintzi 2010); if this view is along the right lines, future investigation should focus on how the phonology–phonetics mapping of onset–tone association is accomplished.

5 Onset–coda weight asymmetry

Finishing this chapter, it should by now be obvious that while there is evidence that onsets participate in at least some of the phenomena that codas do, the frequency with which they do so is indisputably much lower and in some cases exceedingly rare. This issue has been mentioned but barely dealt with in the literature; nevertheless, it deserves some brief discussion. Of course, for those who deny any role for onsets in prosody (cf. the standard moraic theory of Hayes 1989), there is not much to explain in the first place. The asymmetry in behavior is the outcome of the more restricted – moraically speaking – structural representation of onsets, compared to that of codas. However, as we have just seen, this approach is too restrictive when it encounters many of the empirical data presented previously.

[12] A reviewer points out that the input for the 3rd singular could instead include a low-toned nasal that on the surface fuses with the onset consonant, similarly to what happens in sonorant-initial stems. This is certainly a possibility, but not one Hyman seems to assume. In any case, this issue is orthogonal to the point made here.

To my knowledge, the first explicit attempt to account for the rarity of onset weight and hence of the onset–coda prosodic asymmetry was offered by Goedemans (1998). Through a set of perception experiments using synthetic stimuli, Goedemans found that Dutch listeners are more attuned to perceive fluctuations in vowel or coda duration rather than onset duration. He next devised an additional experiment to check for the possibility that there is inherently a human bias against perceiving onset duration, but found no evidence in support of this. He therefore concluded that the effect described above must genuinely be due to the weightlessness of onsets. One problem posed by this account is that Goedemans found that listeners recognize duration shifts in onset sonorants better than obstruents. This implies that the former should be preferred as weight bearers to the latter, *contra* the empirical data, which suggest that in onsets the real difference is between voiced and voiceless obstruents (and that sonorants may pattern with either; cf. Karo *vs.* Pirahã). More troublesome for this proposal is how to accommodate later work by the same author (cf. Hajek and Goedemans 2003), where onset weight is emphatically argued for, albeit for geminates only (Rob Goedemans, personal communication).

Other, more functional accounts of the onset-coda weight asymmetry include Smith (2005) and Gordon (2005), both of which accept onset-sensitivity, but only with regard to stress. To explain why onsets may have a stress-attracting effect, they offer variants of the same idea relating phonological considerations to more general cognitive abilities, such as the sensitivity to auditory stimuli (Viemeister 1980; Delgutte 1982). More specifically, they allude to the evidence of "neural response patterns that the presence of an onset, and specifically a low-sonority onset, does in fact enhance the perceptual response to a syllable" (Smith 2005: 50). Empirically, though, as we know, sonorant onsets may also contribute to weight (or prominence), a fact that both functional accounts fail to capture. Despite this problem, Gordon (2005) claims that in most cases, i.e. most languages, the onset effect is subordinated to the perceptual energy of the rime itself, which is why rimal weight is prioritized over onset weight.

Finally, Topintzi (2006, 2010) does not confront this issue in much detail, but nonetheless claims that instead of a single property, it is a constellation of phonological factors, perhaps complemented by the functional accounts above, which may prove enlightening (for details, see Topintzi 2010: §3.3.3, §5.4.1, §6.2.3). For example, the rarity of onset-sensitive tone is attributed to the fact that tone and onset-weight requirements are incompatible with one another. Tone requires the presence of F0, whereas moras that can bear tone in the onset are best assigned to voiceless onsets, which by nature lack F0.

In spite of the virtues of each approach, none simultaneously manages to combine accurate empirical coverage with a convincing account that acknowledges the onset-coda asymmetry in its correct perspective and offers a plausible explanation. Future research may fill in the missing pieces of the puzzle.

ACKNOWLEDGMENTS

Thanks to Beth Hume, Marc van Oostendorp, and two anonymous reviewers for their instructive comments on various aspects of this chapter. All errors are my own.

REFERENCES

Abrahamson, Arno & Joyce Abrahamson. 1984. Os fonemas da língua Júma. In Robert A. Dooley (ed.) *Estudos sobre línguas tupí do brasil*, 157–174. Brasília: Summer Institute of Linguistics.

Baertsch, Karen. 2002. An optimality theoretic approach to syllable structure: The split margin hierarchy. Ph.D. dissertation, Indiana University.

Bagemihl, Bruce. 1998. Maximality in Bella Coola (Nuxalk). In Ewa Czaykowska-Higgins & M. Dale Kinkade (eds.) *Salish languages and linguistics: Theoretical and descriptive perspectives*, 71–98. Berlin & New York: Mouton de Gruyter.

Beltzung, Jean-Marc. 2007. Allongements compensatoires: Une typologie. Paper presented at the 7th Biennial Meeting of the Association for Linguistic Typology, Paris. Handout available (June 2010) at http://jeanmarc.beltzung.free.fr/?page_id=3.

Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104. 591–630.

Blevins, Juliette. 1995. The syllable in phonological theory. In Goldsmith (1995), 206–244.

Blevins, Juliette. 2001. Where have all the onsets gone? Initial consonant loss in Australian Aboriginal languages. In Jane Simpson, David Nash, Mary Laughren, Peter Austin & Barry Alpher (eds.) *Forty years on: Ken Hale and Australian languages*, 481–492. (*Pacific Linguistics* 512.) Canberra: Australian National University.

Boersma, Paul. 2007. Some listener-oriented accounts of *h-aspiré* in French. *Lingua* 117. 1989–2054.

Booij, Geert. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.

Breen, Gavan & Rob Pensalfini. 1999. Arrernte: A language with no syllable onsets. *Linguistic Inquiry* 30. 1–25.

Broselow, Ellen. 1995. Skeletal positions and moras. In Goldsmith (1995), 175–205.

Buckley, Eugene. 1992. Kashaya laryngeal increments, contour segments, and the moraic tier. *Linguistic Inquiry* 23. 487–496.

Buckley, Eugene. 2003. Children's unnatural phonology. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 29. 523–534.

Buller, Barbara, Ernest Buller & Daniel L. Everett. 1993. Stress placement, syllable structure, and minimality in Banawá. *International Journal of American Linguistics* 59. 280–293.

Cho, Seung-bog. 1967. *A phonological study of Korean with a historical analysis*. Uppsala: Almqvist & Wiksell.

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 283–333. Cambridge: Cambridge University Press.

Crowhurst, Megan J. 1988. Empty consonants and direct prosody. *Proceedings of the West Coast Conference on Formal Linguistics* 7. 67–79.

Curtis, Emily. 2003. Geminate weight: Case studies and formal models. Ph.D. dissertation, University of Washington.

Davis, Stuart. 1985. Topics in syllable geometry. Ph.D. dissertation, University of Arizona.

Davis, Stuart. 1988. Syllable onsets as a factor in stress rules. *Phonology* 5. 1–19.

Davis, Stuart. 1999. On the representation of initial geminates. *Phonology* 16. 93–104.

Davis, Stuart & Gina Torretta. 1998. An optimality-theoretic account of compensatory lengthening and geminate throwback in Trukese. *Papers from the Annual Meeting of the North East Linguistic Society* 28. 111–125.

Davis, Stuart, Linda Manganaro & Donna Jo Napoli. 1987. Stress on second conjugation infinitives in Italian. *Italica* 64. 477–498.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.

Delgutte, Bertrand. 1982. Some correlates of phonetic distinctions at the level of the auditory nerve. In Rolf Carlson & Björn Granström (eds.) *The representation of speech in the peripheral auditory system*, 131–150. Amsterdam: Elsevier Biomedical.

Dell, François & Mohamed Elmedlaoui. 1985. Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7. 105–130.

Duanmu, San. 2008. *Syllable structure: The limits of variation*. Oxford: Oxford University Press.

Erwin, Sean. 1996. Quantity and moras: An amicable separation. *UCLA Occasional Papers in Linguistics* 17. 2–30.

Everett, Daniel L. 1988. On metrical constituent structure in Pirahã phonology. *Natural Language and Linguistic Theory* 6. 207–246.

Everett, Daniel L. & Keren Everett. 1984. On the relevance of syllable onsets to stress placement. *Linguistic Inquiry* 15. 705–711.

Fudge, Erik C. 1969. Syllables. *Journal of Linguistics* 5. 253–286.

Gabas, Nilson, Jr. 1999. A grammar of Karo (Tupi). Ph.D. dissertation, University of California, Santa Barbara.

Gahl, Susanne. 1996. Syllable onsets as a factor in stress rules: The case of Mathimathi revisited. *Phonology* 13. 329–344.

Geudens, Astrid & Dominiek Sandra. 2003. Beyond implicit phonological knowledge: No support for an onset–rime structure in children's explicit phonological awareness. *Journal of Memory and Language* 49. 157–182.

Gnanadesikan, Amalia E. 1995. Markedness and faithfulness constraints in child phonology. Unpublished ms., University of Massachusetts, Amherst (ROA-67).

Goedemans, Rob. 1998. Weightless segments. Ph.D. dissertation, University of Leiden.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.

Gordon, Matthew. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory* 23. 595–653.

Gordon, Matthew. 2006. *Syllable weight: Phonetics, phonology, typology*. London: Routledge.

Gouskova, Maria. 2001. Falling sonority onsets, loanwords and syllable contact. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 37. 175–186.

Gouskova, Maria. 2004. Relational hierarchies in Optimality Theory: The case of syllable contact. *Phonology* 21. 201–250.

Greenberg, Joseph H. 1978. Some generalizations concerning initial and final consonant clusters. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 2: *Phonology*, 243–279. Stanford: Stanford University Press.

Hajek, John & Rob Goedemans. 2003. Word-initial geminates and stress in Pattani Malay. *The Linguistic Review* 20. 79–94.

Harris, John. 2006. The phonology of being understood: Further arguments against sonority. *Lingua* 116. 1483–1494.

Haudricourt, André-George. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 242. 69–82.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Honda, Kiyoshi. 2004. Physiological factors causing tonal characteristics of speech: From global to local prosody. In Bernard Bel & Isabelle Marlien (eds.) *Speech prosody 2004*. Nara, Japan. Available (June 2010) at www.isca-speech.org/archive/sp2004/sp04_739.pdf.

Hooper, Joan B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.

Hulst, Harry van der & Nancy Ritter. 1999. Theories of the syllable. In Harry van der Hulst & Nancy Ritter (eds.) *The syllable: Views and facts*, 13–52. Berlin & New York: Mouton de Gruyter.

Hyman, Larry M. 1985. *A theory of phonological weight*. Dordrecht: Foris.

Itô, Junko. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7. 217–259.

Jakobson, Roman. 1962. *Selected writings*, vol. 1: *Phonological studies*. The Hague: Mouton.

Jeanne, LaVerne Masayesva. 1982. Some phonological rules of Hopi. *International Journal of American Linguistics* 48. 245–270.

Katada, Fusa. 1990. On the representation of moras: Evidence from a language game. *Linguistic Inquiry* 21. 641–646.

Katsanis, Nikolaos. 1996. *Το γλωσσικό ιδίωμα της Σαμοθράκης* [*The dialect of Samothraki Greek*]. Δήμος Σαμοθράκης [Municipality of Samothraki].

Kavitskaya, Darya. 2002. *Compensatory lengthening: Phonetics, phonology, diachrony*. London & New York: Routledge.

Kaye, Jonathan. 1990. "Coda" licensing. *Phonology* 7. 301–330.

Kaye, Jonathan. 2000. A users' guide to Government Phonology. Unpublished ms., University of Ulster.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology* 7. 193–231.

Kenstowicz, Michael. 1994. Sonority-driven stress. Unpublished ms., MIT (ROA-33).

Kingston, John & David Solnit. 1988a. The inadequacy of underspecification. *Papers from the Annual Meeting of the North East Linguistic Society* 19. 264–278.

Kingston, John & David Solnit. 1988b. The tones of consonants. Unpublished ms., Cornell University & University of Michigan.

Kreitman, Rina. 2006. Cluster buster: A typology of onset clusters. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 42. 163–179.

Kuryłowicz, Jerzy. 1948. Contribution à la théorie de la syllabe. *Bulletin de la Société Polonaise de Linguistique* 8. 80–113.

Levelt, Clara C., Niels O. Schiller & Willem J. Levelt. 2000. The acquisition of syllable types. *Language Acquisition* 8. 237–264.

Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.

Lewkowicz, Nancy K. & Leone Y. Low. 1979. Effects of visual aids and word structure on phonemic segmentation. *Contemporary Educational Psychology* 4. 238–252.

Loporcaro, Michele. 1991. Compensatory lengthening in Romanesco. In Pier Marco Bertinetto, Michael Kenstowicz & Michele Loporcaro (eds.) *Certamen phonologicum II: Papers from the 1990 Cortona Phonology Meeting*, 279–307. Turin: Rosenberg & Sellier.

Lowenstamm, Jean. 1996. CV as the only syllable type. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, 419–441. Salford: ESRI.

Marlett, Stephen A. & Joseph Paul Stemberger. 1983. Empty consonants in Seri. *Linguistic Inquiry* 14. 617–639.

Mellander, Evan W. 2003. Weightless syllables in Alamblak and the case against Head Dependence. Poster presented at the CASTL Kick-Off Conference, University of Tromsø.

Michelson, Karin. 1988. *A comparative study of Lake-Iroquoian accent*. Dordrecht: Kluwer.

Morelli, Frida. 1999. The phonotactics and phonology of obstruent clusters in Optimality Theory. Ph.D. dissertation, University of Maryland at College Park.

Morelli, Frida. 2003. The relative harmony of /s+stop/ onsets: Obstruent clusters and the Sonority Sequencing Principle. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*, 356–371. Cambridge: Cambridge University Press.

Morén, Bruce. 2001. *Distinctiveness, coercion and sonority: A unified theory of weight*. New York & London: Routledge.

Muller, Jennifer. 1999. A unified mora account of Chuukese. *Proceedings of the West Coast Conference on Formal Linguistics* 18. 393-405.

Muller, Jennifer. 2001. The phonology and phonetics of word-initial geminates. Ph.D. dissertation, Ohio State University.

Nanni, Debbie. 1977. Stressing words in *-Ative*. *Linguistic Inquiry* 8. 752–763.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.) *The production of speech*, 189–216. New York: Springer.

Ohala, John J. 1990. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26(2). 319–338.

Oostendorp, Marc van. 2005. Expressing inflection tonally. *Catalan Journal of Linguistics* 4. 107–126.

Oostendorp, Marc van. 2006. Tone and segmental structure. Handout for a course given at the CLOW/DGfS Linguistics Summer School on Micro- and Macrovariation, Stuttgart. Available (June 2010) at www.vanoostendorp.nl/pdf/tone.pdf.

Parker, Steve. 1994. Coda epenthesis in Huariapano. *International Journal of American Linguistics* 60. 95–119.

Parker, Steve. 2002. Quantifying the sonority hierarchy. Ph.D. dissertation, University of Massachusetts, Amherst.

Payne, David & Furne Rich. 1988. Sensitivity to onset in Arabela stress. Unpublished ms., Instituto Lingüístico de Verano.

Peterson, Tyler. 2007. Issues of homophony and the minimal word in the Adyghan languages. Paper presented at the Conference on the Languages of the Caucasus, Max Planck Institute for Evolutionary Anthropology, Leipzig.

Pike, Kenneth L. & Eunice V. Pike. 1947. Immediate constituents of Mazateco syllables. *International Journal of American Linguistics* 13. 78–91.

Rialland, Annie. 1993. L'allongement compensatoire: Nature et modèles. In Bernard Laks & Annie Rialland (eds.) *L'architecture et la géometrie des représentations phonologiques*, 59–92. Paris: CNRS.

Rice, Keren. 1992. On deriving sonority: A structural account of sonority relationships. *Phonology* 9. 61–99.

Roberts-Kohno, R. Ruth. 1995. Vowel coalescence and hiatus in Kikamba. In Akinbiyi Akinlabi (ed.) *Theoretical approaches to African linguistics*, 314–328. Trenton, NJ: Africa World Press.

Robinson, Lila W. 1975. Some phonological rules of Iowa-Oto (Siouan). *Proceedings of the Mid-America Linguistics Conference*, 439–448. Lawrence: University of Kansas.

Scheer, Tobias. 2004. *A lateral theory of phonology*, vol. 1: *What is CVCV, and why should it be?* Berlin & New York: Mouton de Gruyter.

Schreuder, Robert & Wim H. J. van Bon. 1989. Phonemic analysis: Effects of word properties. *Journal of Research in Reading* 12. 59–78.

Selkirk, Elisabeth. 1982. The syllable. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*. Part II, 337–383. Dordrecht: Foris.

Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*, 107–136. Cambridge, MA: MIT Press.

Shryock, Aaron M. 1995. Investigating laryngeal contrasts: An acoustic study of the consonants of Musey. *UCLA Working Papers in Phonetics* 89. Available at http://escholarship.org/uc/item/7jb1r066.

Smith, Jennifer L. 2005. *Phonological augmentation in prominent positions*. New York & London: Routledge.

Sommer, Bruce. 1981. The shape of Kunjen syllables. In Didier L. Goyvaerts (ed.) *Phonology in the 1980s*, 231–244. Ghent: Story-Scientia.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT.
Steriade, Donca. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5. 73–155.
Strehlow, Theodor G. H. 1944. *Aranda phonetics and grammar*. Sydney: University Press.
Svantesson, Jan-Olof. 1983. *Kammu phonology and morphology*. Malmö: CWK Gleerup.
Tang, Katrina Elizabeth. 2008. The phonology and phonetics of consonant–tone interaction. Ph.D. dissertation, University of California, Los Angeles.
Topintzi, Nina. 2006. Moraic onsets. Ph.D. dissertation, University College London.
Topintzi, Nina. 2007. Weight polarity in Ancient Greek and other languages. In Mary Baltazani, George K. Giannakis, Tasos Tsangalidis & George J. Xydopoulos (eds.) *Proceedings of the 8th International Conference on Greek Linguistics*, 503–517. Ioannina: Department of Linguistics, University of Ioannina.
Topintzi, Nina. 2008. On the existence of moraic onsets. *Natural Language and Linguistic Theory* 26. 147–184.
Topintzi, Nina. 2010. *Onsets: Suprasegmental and prosodic behaviour*. Cambridge: Cambridge University Press.
Treiman, Rebecca. 1986. The division between onsets and rimes in English syllables. *Journal of Memory and Language* 25. 476–491.
Uhry, Joanna K. & Linnea C. Ehri. 1999. Ease of segmenting two- and three-phoneme words in kindergarten: Rime cohesion or vowel salience. *Journal of Educational Psychology* 91. 594–603.
Vaux, Bert. 2009. On the phonological representation of timing and prosody. Paper presented at the CUNY Conference on the Foot, City University of New York. Handout available (June 2010) at www.cunyphonologyforum.net/FOOTPAPERS/vauxhandout.pdf.
Vennemann, Theo. 1972. On the theory of syllabic phonology. *Linguistische Berichte* 18. 1–18.
Viemeister, Neal. 1980. Adaptation of masking. In G. van den Brink & F. A. Bilsen (eds.) *Psychophysical, physiological and behavioural studies in hearing*, 190–198. Delft: Delft University Press.
Welmers, William E. 1962. The phonology of Kpelle. *Journal of African Languages* 1. 69–93.
Wordick, Frank. 1982. *The Yindjibarndi language*. (*Pacific Linguistics* C71.) Canberra: Australian National University.
Yallop, Colin. 1977. *Alyawarra: An Aboriginal language of Central Australia*. Canberra: Australian Institute of Aboriginal Studies.
Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.
Yip, Moira. 2003. Some real and not-so-real consequences of comparative markedness. *Theoretical Linguistics* 29. 53–64.
Yupho, Nawanit. 1989. Consonant clusters and stress rules in Pattani Malay. *Mon-Khmer Studies* 15. 125–137.
Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12. 85–129.
Zec, Draga. 2007. The syllable. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 161–194. Cambridge: Cambridge University Press.

56 Sign Syllables

Ronnie Wilbur

1 Introduction

This chapter focuses on the notion of syllable in sign languages. Although there is now a consensus on the defining feature of a syllable in sign languages, i.e. that there must be a movement, the initial idea of having "syllables" in sign languages met with considerable resistance on its introduction in the early 1980s, in large part because sign languages are fundamentally monosyllabic languages (see Coulter 1982 on American Sign Language (ASL); similar evidence has been provided for Finnish Sign Language by Jantunen 2007). There was also a strong undercurrent that using concepts borrowed from spoken language linguistics for sign language phenomena would be problematic, if not inappropriate. However, I had an opportunity to ask my colleague Ray Kent, a speech researcher who was then editor of the *Journal of Speech and Hearing Research*, what he would require if I were to send his journal a paper on syllables in sign languages. He said he would look for evidence that the concept was linguistically meaningful and could be reliably measured, and his response will begin our tour on this topic. Actually, proving to spoken language researchers that there are syllables in sign languages on those two criteria was remarkably easy (Wilbur and Nolen 1986; Wilbur and Allen 1991). In contrast, the phonological representation of sub-syllabic structure has been under constant and lively debate. Accordingly, the presentation of evidence for the existence of syllables will be brief, and the bulk of the chapter will focus on describing the issues of disagreement and providing evidence for an answer to the question of the representation of syllables in sign languages.

2 Historical background

Early sign language research treated the sign as the unit of analysis. This is best observed in Stokoe (1960), where each sign was treated as a unit. Then, when this unit was analyzed further, it was observed that a sign was composed of a "simultaneous" bundle of aspects/primes/parameters, including the "big four": handshape, location, movement, and orientation (Stokoe 1960; Friedman 1974, 1977; Battison 1978; Siple 1978; Klima and Bellugi 1979; Wilbur 1979; see also

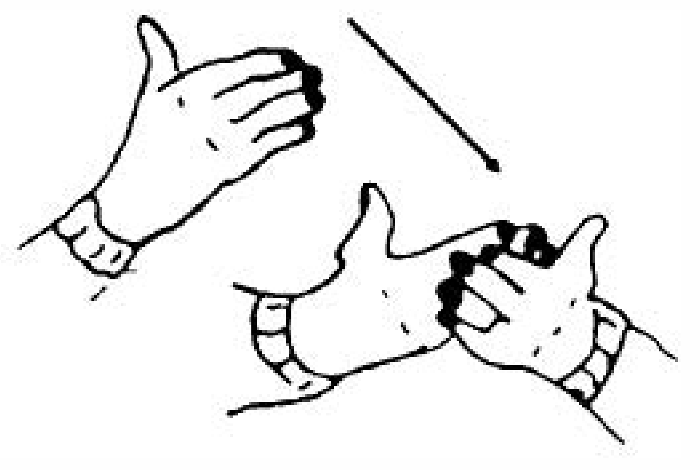


Figure 56.1 ASL ARRIVE is a monosyllabic form (arrow shows direction of movement)

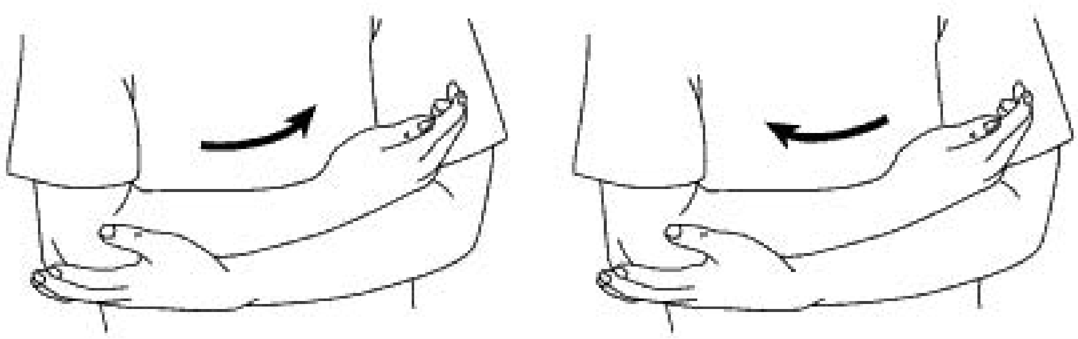


Figure 56.2 ASL BABY has two syllables, i.e. two movements (arrows show direction of movement)

CHAPTER 9: HANDSHAPE IN SIGN LANGUAGE PHONOLOGY; CHAPTER 10: THE OTHER HAND IN SIGN LANGUAGE PHONOLOGY; CHAPTER 24: THE PHONOLOGY OF MOVEMENT IN SIGN LANGUAGE). Several studies suggested that there are syllables in ASL and that these syllables have internal "sequential" organization (Kegl and Wilbur 1976; Chinchor 1978, 1981; Newkirk 1979, 1980, 1981; Liddell 1984). Before addressing the question of internal organization of sign syllables, some clarification of the notion "syllable" in sign languages is necessary.

To better appreciate the status of a syllable in sign languages, we must consider the difference between a syllable and a sign. In many cases, the sign is a single syllable and the boundaries of the two, i.e. a syllable or a sign, coincide (Coulter 1982, 1990). Figure 56.1 is an example of a single-syllable sign.

In other signs, such as BABY in Figure 56.2, there are two syllables in a single sign. The sign BABY is a larger unit than ARRIVE (Figure 56.1). In still other signs, such as MOTHER (Figure 56.3), there are specifications for handshapes, locations, and hand orientations, but, critically, these signs do not have their own movement specifications.

Since every syllable must have a movement, the phonological specification for these lexical items is smaller than a syllable; it has been proposed that the epenthetic/transition movement to, or away from, the target location provides the prosodic feature needed to produce a full syllable (Wilbur 1985; Brentari 1990b, 1998; Geraci 2009).

We have so far sketched sign syllables by providing three groups of signs with different movement specifications, i.e. with single movement, with two movements,

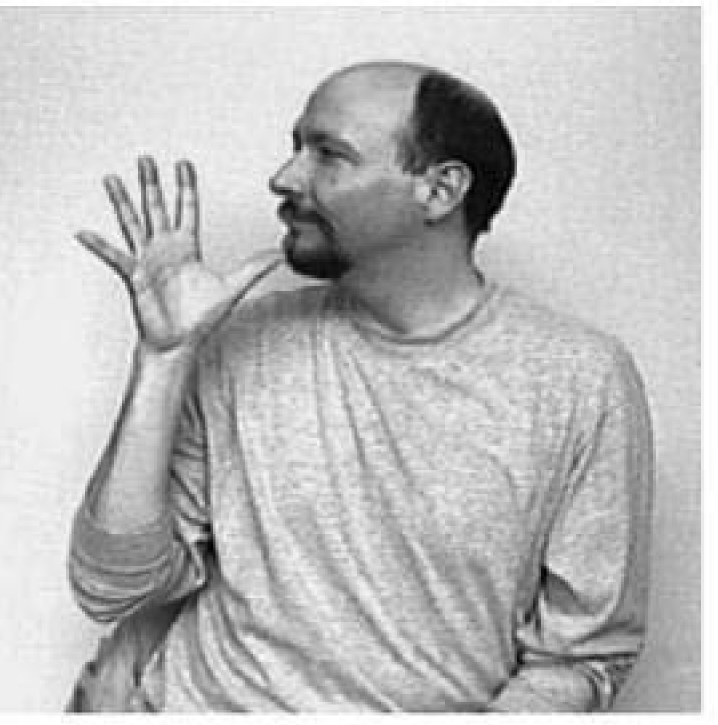


Figure 56.3 ASL MOTHER is a sign without a movement specification. By permission of Dr. Bill Vicars

and without lexical movement. We now consider the difference between a syllable and a morpheme in order to distinguish their functions. A morpheme is defined as the smallest possible unit of meaning. In Figures 56.1–3, each sign is also a single morpheme. In Figure 56.1, the morpheme and the syllable are the same size. In Figure 56.2, the morpheme is larger than the syllable (it has two syllables), and in Figure 56.3, the morpheme is smaller than the syllable (it is missing movement). A morpheme may also be as small as the feature specification for a single handshape, as in classifier constructions (see CHAPTER 9: HANDSHAPE IN SIGN LANGUAGE PHONOLOGY), or location, as in verb agreement.

It should be noted that two-syllable lexical items are highly constrained with respect to their movement specifications: the movement in the second syllable is either the exact opposite (180° rotation) of the movement in the first syllable (as in BABY, Figure 56.2) or it is 90° rotated (Figure 56.4; note transitional movement inserted when the end of the first movement is not the starting position of the second movement) (Wilbur and Petersen 1997).

Clearly movement is central to syllable structure (see CHAPTER 24: THE PHONOLOGY OF MOVEMENT IN SIGN LANGUAGE). The first attempt to break sign movement into smaller sequential pieces was Newkirk (1979, 1980, 1981). Considering rhythmic features of movement, he analyzed them into [onset] [movement] [offset]. Subsequently, a number of sequential and simultaneous proposals were offered. This brings us to the ongoing debate – what is the internal structure of a syllable with respect to sequentiality and simultaneity? In the following section, I provide an overview of what everyone does agree upon, i.e. that syllables in sign languages exist. In the subsequent section, I provide evidence of the behavior of syllables with respect to higher phonological organization. Finally, we dive inside the sign syllable and consider the evidence for the two theoretical options – sequential and simultaneous organization of syllable structure.

Figure 56.4 ASL CANCEL/CORRECT/CRITICIZE and schematic of movement showing second syllable (3 to 4) perpendicular to first (1 to 2)

3 Sign syllables exist

Syllables have been reliably measured, and in conversational contexts they have roughly the same duration as spoken language syllables (Wilbur and Nolen 1986). As will become clear below, there has been no shortage of linguistic uses for the syllable in sign language (morpho-)phonology, hence it is linguistically meaningful. It is fair to say that sign language phonologists now take the notion of sign syllable as a given, and that movement is its nucleus (the carrier of its perceptual salience; Jantunen and Takkinen 2010).

3.1 *Syllable measurements*

Investigators have measured sign duration, including signs that are clearly single syllables (Bellugi and Fischer 1972; Friedman 1976; Liddell 1978, 1984). For those signs which are monosyllabic, the measured duration means range from 233 to 835 msecs as a function of context. Liddell (1978) reported the effects of sentence position and syntactic function on duration of the monosyllabic signs DOG and CAT. His measurements show phrase-final lengthening, as the durations were longest in sentence-final position. Duration in sentence-initial position was next longest, and medial position in relative clauses had the shortest duration. His measurements also show a syntactic effect: objects were shorter than subjects or heads of relative clauses.

My investigation of syllable duration began in 1984, when videotape was "reel to reel," which meant that the tapes could be moved by hand, forward and backward, and measurement was in "fields" – 60 per second. To measure syllables, movements, and holds, we had to provide our own mechanical guidelines and demonstrate that they could be reliably used (Wilbur and Nolen 1986; Wilbur and

Schick 1987). We started with the cues identified by Green (1984) for beginnings and ends of signs, which worked well for signs that were perceptually monosyllabic. These cues included points of contact and changes in facial expressions and eye gaze. However, Green's procedures were not sufficient for determining syllable boundaries when the sign and the syllable are not coterminous, i.e. when we have more than one syllable in a sign. To capture the behavior of multisyllabic signs such as bidirectional signs (Figure 56.2 above), reduplicated forms, and compounds, additional cues were needed. With the aid of native signers, we determined that a change in the direction of movement marked a boundary between two adjacent syllables. For elliptical movements, we accepted Newkirk's (1979, 1980, 1981) argument that they were segmentable into two parts, and then we used the change in direction of movement as the boundary between the two parts. By contrast, circular movements, which show no internal structure, were treated as one syllable per circle. For holds, we established a procedure in which the end of a hold would be marked by one or more of the following cues: start of the next movement, loss of tension in the signing hand(s), change of eye gaze, initiation of signing by the other signer, or change of eye gaze by the other signer (Wilbur and Nolen 1986). Syllables were measured by two people at a time; if they could not agree, a third person was consulted. Over three thousand syllables were measured in four situations – natural conversation, elicited paragraphs, lists of signs, and phrases and compounds.

In conversations, 889 syllables from three signers had a mean of 248 msecs, comparable to the estimated 250 msecs for spoken English (Adams 1979; Hoequist 1983). This similarity may be a reflection of an underlying timing mechanism for motion that may surface not only in speech and signing, but also in non-linguistic motor behaviors. For example, in baseball, a bat swing takes about 200 msecs (Schmidt and Lee 2005).

For the lists, mean syllable duration was 299 msecs for the first production, and 417 msecs for the second. Thus, signers can have different durations at different times. For paragraphs, 14 signers produced paragraphs with either a stressed or unstressed target sign. The stressed target mean was 317 msecs, and the unstressed target mean 299 msecs. There were more syllables in the stressed condition, i.e. repeated syllables and/or resyllabification (similar to English *please* pronounced as *puh-leeze*).

In the last condition, compounds may have two syllables or can be reduced to one (Coulter 1982; Liddell 1984). Eighteen sets of compounds and their two component signs were provided by Ursula Bellugi. In each set, the two signs appeared in a phrase in isolation (e.g. FACE CLEAN) and in context (HE HAS FACE CLEAN 'He has a clean face'). The same morphemes also appeared in a compound (FACE-CLEAN 'handsome') in isolation and in context (HE FACE-CLEAN 'He is handsome'). The compounds had significantly more syllables per sign than simple lexical items (Wilbur and Nolen 1986). Also, the signs in isolation (whether simple lexical items or compounds) had significantly more syllables than in context, reflecting prosodic effects. Thus signers manipulate both syllable duration and number of syllables in their sign productions.

The evidence so far lends support to one of the two criteria that started our discussion, i.e. "Can syllables be reliably measured in sign languages?" We turn now to the other criterion, i.e. "Are syllables linguistically useful?"

4 Phonological applications identified for "syllable"

In this section, I briefly review the arguments that have been offered to show that syllables contribute to the statement of phonological processes and to our understanding of why some processes behave the way they do. The notion of syllables has proven to be useful in the statement of a variety of historical changes (Battison 1978; Frishberg 1978), synchronic morphological processes (Chinchor 1981), and phonological processes (Coulter 1982; Wilbur 1990b, 1993). Blevins (1993), Padden (1993) and the work of Brentari (1990a, 1990b, 1993, 1996, 1998) provide further arguments in favor of the role of syllable structure in ASL phonology. I review only a few, and refer the reader to the original authors for further evidence and discussions.

4.1 *Fingerspelling and fingerspelled loan signs*

Battison's (1978) discussion of the creation of new signs from fingerspelled words provides data to support the notion of syllables. Theoretically, each fingerspelled letter consists of a handshape and, when produced in slow sequences, a transition movement (handshape change) to reach each handshape. Thus, there could potentially be as many syllables as there are letters in the word being fingerspelled, because there could be one movement to make each handshape and each could therefore be a syllable. In actuality, fluent fingerspelling is performed with a phrasal rhythm that smoothes the transition handshape changes and reduces the prominence of certain handshapes while increasing the prominence of others (Akamatsu 1982, 1985; Wilcox 1992). In the process of becoming a lexicalized fingerspelled loan sign, some letters in a word are dropped, and remaining handshapes are associated to syllabic nodes, reducing the number of syllables produced. Fingerspelling the word "sick" (Figure 56.5) involves handshape changes from each letter to the following letter. Since fingerspelling is based on English spelling, each English word will have a different set of handshapes and a corresponding different set of transitional handshape changes. In contrast, in the fingerspelled loan sign #SICK (where # denotes a fingerspelled loan sign (Figure 56.6), the middle letters have been dropped, and the handshape change from S to K has created the movement nucleus of the syllable, to which a slight directional path movement has been added (the arrow does not do justice to this – the middle finger can appear to flick forward from the fist while the index finger straightens up). At the syllable-internal level, the features for

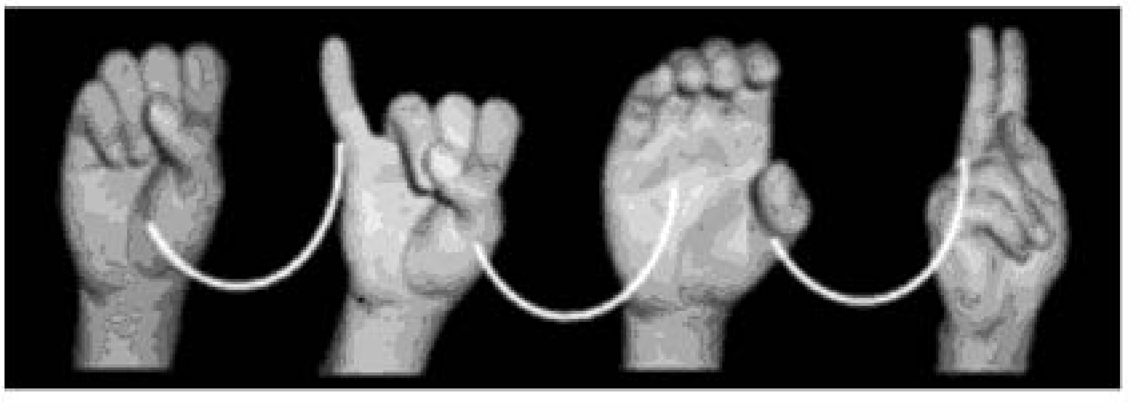


Figure 56.5 Handshapes S, I, C, K, with three transition movements between them

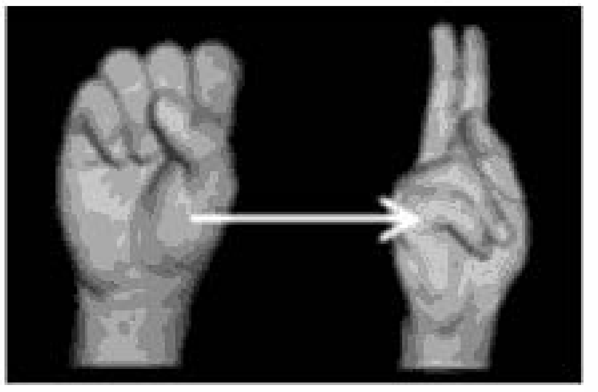


Figure 56.6 Fingerspelled loan sign #SICK, with one lexical movement

the handshapes S and K are associated with the same syllabic node, i.e. the only syllabic node. The handshape change from S to K is not permitted in core lexical items as reflected in Brentari's model, thus this form is clearly identified as of foreign origin from English. Brentari (1994) used the fingerspelled loan signs from published ASL lectures (Valli and Lucas 1992) to determine the processes involved in lexicalization. She found that long fingerspelled words with as many as eight or more letters reduced to fewer handshapes and just two movements. The result is that the newly lexicalized forms fit the phonotactics of ASL, having a maximum of two syllables.

4.2 *Evidence for a sonority hierarchy*

Several researchers have suggested a relationship between visibility and syllable sonority (Corina 1990; Perlmutter 1992; Sandler 1993; Brentari 1998). The sonority hierarchy treats movements made with joints closer/proximal to the body/trunk, such as elbows and shoulders, as more sonorous, because of their visibility in motion when compared to those lower down and more distal, such as hands and fingers, which are considered less visible and hence less sonorous (1) (see CHAPTER 49: SONORITY for an overview of issues surrounding the sonority hierarchy in spoken language):

(1) *Sonority hierarchy with respect to the relevant joints* (from Brentari 1998)

Most sonorous								*Least sonorous*
shoulder joint	>	elbow joint	>	wrist joint	>	base finger joints	>	non-base finger joints

Movements made with the wrist joint have a higher sonority value, i.e. are more visible due to larger movements, than movements made with the use of finger joints. With respect to the development of fingerspelled loan signs above, the addition of the slight path movement in the loan sign #SICK may be viewed as adding a higher sonority value to the less sonorous handshape change from S to K, thereby producing a more sonorous syllable and hence more acceptable lexical item. Thus what we observe is that the proposed prosodic hierarchy provides a rationale for why finger positions are dropped and wrist or elbow movements (depending on how the small path is made) are retained in fingerspelled loan words (CHAPTER 95: LOANWORD PHONOLOGY).

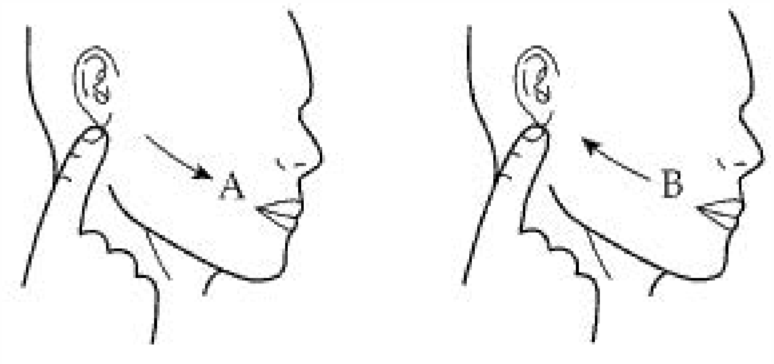


Figure 56.7 Metathesis for the ASL sign DEAF. (a) A-to-B. (b) B-to-A

4.3 Contact metathesis

Another situation where the notion of syllable is phonologically relevant is found with two-location contacting signs, which in certain circumstances can undergo a change that causes the two locations to switch their order (i.e. metathesis; Kegl and Wilbur 1976; Johnson 1986; Sandler 1986; Wilbur 1987; CHAPTER 59: METATHESIS). For example, the location points in signs such as DEAF, PARENTS, and FLOWER may switch from A-to-B to B-to-A (Figure 56.7), depending on preceding phonological context:

The process of metathesis is limited to signs that are both single morphemes and single syllables. Some were originally compounds and may be articulated in a way which reflects their origins. For example, the sign PARENTS might be deliberately made to emphasize its origin from FATHER + MOTHER: movement to contact at the forehead (FATHER) and then movement to contact at the chin (MOTHER). This form would be two syllables, and not subject to metathesis. In the monosyllabic version of PARENTS, the forehead contact is preceded by a transition movement, and the lexical syllable consists of the change in location from the forehead contact to the chin contact. This latter form can undergo metathesis, i.e. the hand can touch the chin first and the forehead second.

The statement of metathesis in terms of syllables greatly simplifies the formalization of the rule. We can see that the statement in terms of syllables is the correct one, as opposed to morphemes or signs. PARENTS can undergo metathesis in one form of production but not in another, so the rule about which signs can undergo metathesis could not refer to "contacts in the same sign" or "contacts in the same morpheme," but only to "contacts in the same syllable."

4.4 Handshape change (change in aperture)

Brentari and Poizner (1994) show that handshape change timing is different within syllables than it is between syllables (CHAPTER 9: HANDSHAPE IN SIGN LANGUAGE PHONOLOGY). Within syllables, if there is path movement and handshape change, the handshape change coordinates with the beginning and end of the path. However, when two signs in sequence have different handshapes, there must be a transitional handshape between the end of the first sign and the beginning of the second (as discussed for fingerspelling, above). In such conditions, the change in handshape is not coordinated with the path movement in the same way as within signs, that is, the timing of the change does not distribute evenly over the transitional path movement.

4.5 *Consistency in movement*

Likewise, Tyrone *et al.* (2007) compare monosyllabic sign movements toward the body to the same location (e.g. forehead) in two conditions: (i) the movement is part of the sign (THINK), and (ii) the movement is transitional prior to the sign (SMART). They report that within-syllable within-sign movements show typical bell-shaped velocity curves for targeted movement, whereas transitional movement between signs is less regular. These findings, along with the handshape change findings above, converge on the necessity of separating phonological syllable movement from phonetic epenthetic/transitional movement.

4.6 *Minimal word*

Simply put, the syllable is the smallest possible well-formed sign/word. Furthermore, two syllables is the maximum for well-formed core lexical signs (Perlmutter 1992; Sandler 1993; Brentari 1998; Jantunen 2007). Alternative formulations without mention of syllables would necessarily be more complex.

4.7 *Prosodic constraints*

Miller (1997) argued, on the basis of Quebec Sign Language (LSQ), that Phonological Phrases require a disyllabic foot. Similarly, van der Kooij and Crasborn (2008) suggest that in Sign Language of the Netherlands (NGT) the phonological constraint on the addition of sentence-final pointing must be stated in syllabic terms: sentence-final pointing is permitted only if the outcome is a disyllabic foot. Wilbur (1999a) observes that ASL pronouns in sentence-final position tend to be extrametrical with respect to stress assignment at the phrase level, supporting prosodic constraints proposed in Halle and Vergnaud (1987) for spoken languages.

5 Syllables and prosody

The last two arguments in favor of syllables also provide evidence for the prosodic hierarchy in sign languages. That is, metrical structure (lexical, phrasal, and clausal stress assignment), rhythmic structure, and intonational phrasing are dependent to some degree on the syllable level. (2) shows the prosodic hierarchy we adopt for further discussions (see CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE; CHAPTER 40: THE FOOT; CHAPTER 51: THE PHONOLOGICAL WORD; CHAPTER 57: QUANTITY-SENSITIVITY; CHAPTER 84: CLITICS for more discussion of the prosodic hierarchy). Prosodic words will be discussed in this chapter; for discussion of Intonational Phrases see Wilbur (1994), Sandler and Lillo-Martin (2006), and Weast (2008).

(2) *Prosodic hierarchy*

syllable < prosodic word < prosodic phrase < intonational phrase

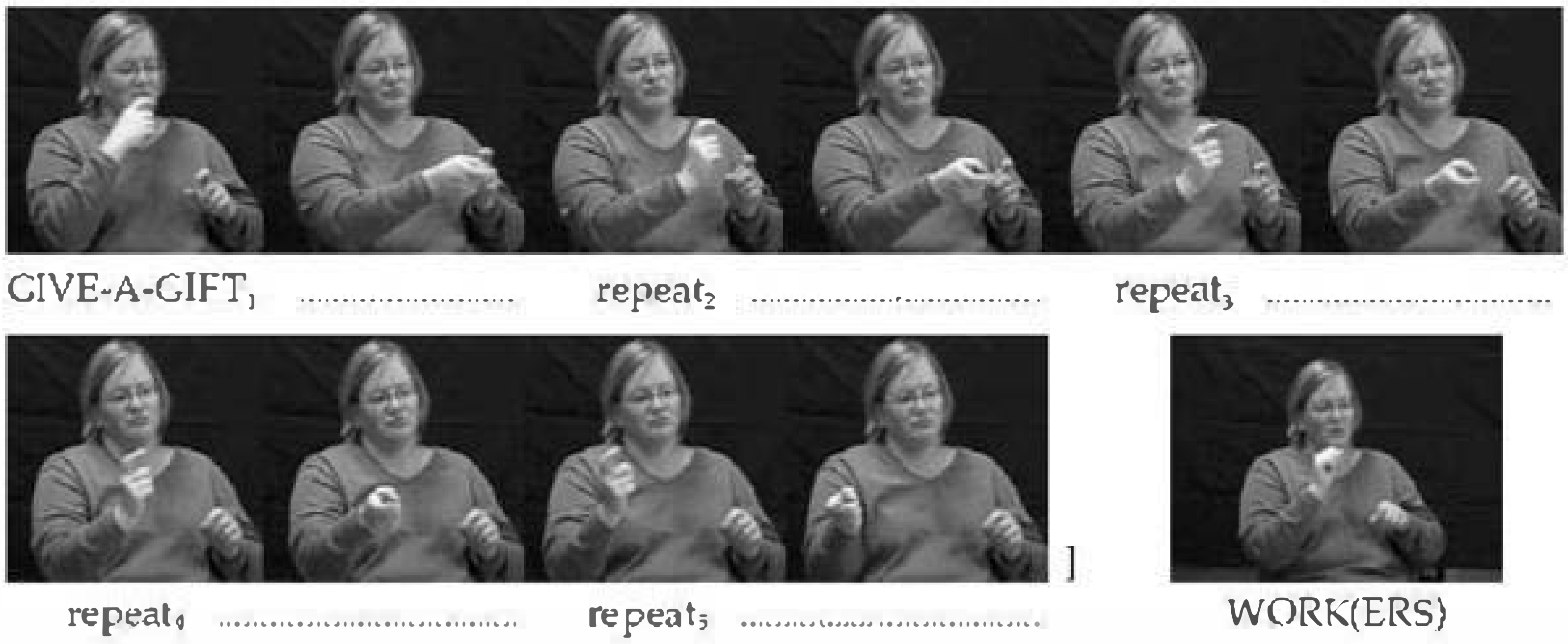


Figure 56.8 One prosodic word, composed of GIVE + distributive aspect, repeated five times, accompanied by one Posture-NM, followed by the next prosodic word, containing WORK

5.1 *Prosodic words*

As indicated, the minimal prosodic word is at least one syllable, and the prosodic constraint on well-formed lexical items is a maximum of two syllables. Brentari and Crossley (2002) demonstrated that changes in lower face tension (mouth and cheeks) mark the end of a prosodic word (PW), which is above the syllable in the prosodic hierarchy. Figure 56.8 shows a single lower face position, i.e. closed mouth with lip corners slightly down, referred to as posture non-manuals (P-NM), which stretches over one long PW, followed by the sign WORK, which has a round mouth and is in a different prosodic word. The context was "every year at Christmas time, the boss gives each of the employees a gift." Note that the single PW contains five syllables (five repetitions of the lexical item GIVE-A-GIFT).

(3) represents the marking of the relevant prosodic words through sign language glossing conventions. The tier above the glosses represents non-manual marking, and the line indicates the spread of the non-manual marker:

(3) *Prosodic grouping for Figure 56.8*

P-NM	P-NM
{GIVE-A-GIFT [Repeat × 5]}$_{PW}$	{WORK(ERS)}$_{PW}$

In contrast, Figure 56.9 shows "one car hits another car three times." The signer produces three mouth changes (Transition-NMs), once for each repetition. These changes result in three PWs, as represented in (4):

(4) *Prosodic grouping for Figure 56.9*

T-NM	T-NM	T-NM
{HIT-CAR}$_{PW}$	{HIT-CAR}$_{PW}$	{HIT-CAR}$_{PW}$

Figure 56.9 Sequence of three CAR-HITs, involving three syllables, three mouth changes, and three PWs

5.2 *Stress assignment*

Stress assignment is a prosodic process, and may occur on lexical items, compounds, and phrases. Early research on ASL stress focused on marking stress on lexical items (Covington 1973; Friedman 1976; Wilbur and Schick 1987; Coulter 1990). Stressed signs can be set off from unstressed signs by several cues: (i) faster/shorter transition movement than between unstressed signs, breaking the rhythmic pattern; (ii) higher in the signing space compared to their unstressed counterparts; (iii) increased repetitions compared to their unstressed counterparts, changing the duration; (iv) increased speed (higher peak instantaneous velocity) compared to their unstressed counterparts; (v) increased muscle tension compared to their unstressed counterparts; and (vi) stressed signs have a following pause (Wilbur and Schick 1987; Wilbur 1990a, 1990b, 1999b, 2009; Allen *et al.* 1991).

5.2.1 *Lexical items*

So far, no sign language has been shown to have distinctive lexical stress, comparable to English *'permit* and *per'mit* (Jantunen and Takkinen 2010). The predominance of monosyllabic lexical items is partly responsible for this absence. Another reason is that polysyllabic signs are restricted to three possibilities:

Lexicalization of repetition: A sign may have more than one syllable if it is formed as a result of lexicalization of a repeated form (e.g. ASL FINGERSPELL; Brentari 1998: 169). The result is a two-movement sign with a Return transition in the middle: A-Return-A. In these forms, only the first syllable is prominent/stressed (Supalla and Newport 1978; Coulter 1990).

Lexical disyllables: A sign may have two syllables if it is a lexical disyllable, i.e. if the morpheme itself requires two syllables. There are two types of disyllables, both of which are subject to constraints on the nature of the movements in each syllable (Wilbur 1990b). In the first type, the movement of the second syllable must be rotated in direction 180° from that of the first, returning the hands to their original location (BABY; Figure 56.2). In the second type, the movement of the second syllable is rotated 90° from the first (creating a crossing movement) (CANCEL; Figure 56.4). Supalla and Newport (1978) discuss the first type, and note that prominence is equal on both syllables. It is also the case that prominence is equal on both syllables in the second type. Thus all lexical disyllables have equal stress on both syllables. Similarly, van der Kooij and Crasborn (2008) show that NGT has both trochaic and iambic stress patterns for disyllabic signs, the type being predictable on the basis of the phonotactics of the rest of the sign. Thus, as for ASL, stress is not distinctive in NGT.

Lexical items have stress on the first, and perhaps only, syllable. Lexical disyllables are exceptional in being specified at the morphemic level for two syllables and in requiring equal prominence on both syllables.

Compounds and phrases: A sign may have two syllables if it is a compound, with the first weaker than the second. Unlike lexical items, the assignment of stress to ASL compounds and syntactic phrases follows a very general pattern. In a compound or phrase, a single stress is assigned to the *most prominent syllable of the rightmost lexical item.*

6 The internal structure of syllables

6.1 *The debate*

Historically, models of the internal structure of syllables (CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE) have taken one of two views. The first view is that sign syllables, like spoken syllables, are composed of sequences of segments (CHAPTER 54: THE SKELETON). These segments are of two types (like consonant and vowel in spoken language), namely Movement (M) (the hands are in motion) and a contrasting type with the hands not in motion. Distinctive features are distributed among these phonological segments parallel to spoken language C and V.

Liddell (1984) argued for two types of segments, movements (M) and holds (H). The remaining information – handshape, contact, orientation, location, and facial expression – is represented as features occurring simultaneously with each segment, thus there is a sequence of feature matrices within each sign. Signed syllables could then be of several types, e.g. M, MH, HM, HMH. Sandler (1986, 1989, 2008) proposed a different model, in which the segment opposition is between movement (M) and location (L), with handshape configuration on a separate autosegmental tier. The presence or absence of holds would be characterized by a binary feature in the location feature matrix; rather than having holds underlyingly, there will be some phonetic holds (list rhythm), some phonological holds (at utterance boundary), some morphological holds (ASL aspectual inflections may include final hold as part of their pattern), and some pragmatic holds (end of conversational turn, waiting for back-channel nod). For these models (e.g. Liddell and Johnson 1989; Sandler 1989, 2008; Sandler and Lillo-Martin 2006), the segments are at the top of the phonological trees containing the distinctive features, i.e. the mother nodes in a feature geometry model. That is, the syllable is composed of segments, which are characterized by relevant phonological features.

In the other view, supported in Brentari (1998) for ASL and van der Kooij (2002) for NGT, movements are dynamic prosodic units with similar autosegmental status as tones in contrastive tonal languages (e.g. Mandarin, Cantonese; CHAPTER 45: THE REPRESENTATION OF TONE; CHAPTER 107: CHINESE TONE SANDHI). An important step leading to this alternative view was van der Hulst's (1993) Head–Dependency Model, in which features that did not change during the sign were considered to be heads, with changing features treated as dependents (for detailed discussion see CHAPTER 24: THE PHONOLOGY OF MOVEMENT IN SIGN LANGUAGE). Head features could be location, orientation, the active (selected) fingers, or their configuration. Movement itself was dependent on change of location (path movement) or hand configuration or orientation (local movement) (see similar arguments in Wilbur 1987). Brentari (1998) provides arguments against the notion of dependent/emerging movement, and instead identifies those features that do not change within the syllable as Inherent Features (IF) and those that do change as Prosodic Features (PF). From this perspective, ASL syllables contain distinctive features which may be accessed by phonological rules only in terms of their tiers and syllabic positions (e.g. syllable-initial, syllable-final), without further subdivision or organization. The segments are abstract timing slots at the bottom of the tree, onto which the phonological features are mapped, i.e. the terminal nodes in a feature geometry approach. Thus the question arises of how these two models should be distinguished.

6.2 *Evidence related to syllable structure*

Jantunen and Takkinen (2010) observe that there is no "direct phonetic evidence" to support the sequential segmental models. In fact, evidence against the segmental arrangement of internal syllable structure comes from a variety of experimental sources: tapping, slips of the hand, and backwards signing.

6.2.1 Tapping

Spoken syllables have a rhythmic focus at the onset of the nuclear vowel (Allen 1972). That is, native English speakers who tap in time to speech cluster their taps at the stressed vowel onset. In a comparable study of ASL, native Deaf signers, native hearing signers and sign-naive hearing subjects were asked to "tap the rhythm" of five different three-sentence signed stories. Each story was presented 30 times. One story was repeated as the sixth condition (30 repetitions) for reliability; these conditions represent "tap the rhythm" (Allen *et al.* 1991). Finally, another one of the stories was repeated (30 repetitions) with new instructions to "tap the syllables." Analysis of the tap responses in this condition showed that for all groups, taps are evenly distributed within syllables and do not differ from chance distribution. That is, no syllable-internal rhythmic focus is apparent (Wilbur and Allen 1991). This result is very crucial, and can only be predicted if the sign syllable is composed of constantly changing movement (smoothly changing muscular activity), meaning that there is no single point in time which attracts perceptual attention in the way that the onset of a spoken stressed vowel does, with large changes in muscular and acoustic energy (Allen 1972). The absence of such peaks is consistent with the proposal in the Prosodic Model that there is no further segmentation inside the sign syllable.

6.2.2 Slips of the hand

Additional arguments against segmental models come from sign errors (Meier 1993; Wilbur 1993). English slips of the tongue tend to involve all the features of the segments involved (Fromkin 1971, 1973). If sign phonological features are distributed across segments, as suggested in segmental models, all features associated to each segment should be able to behave as a group. Therefore, parallel to speech, we might expect that the initial segments of two signs could switch with everything else remaining the same. In the corpus of 131 slips of the hand (Klima and Bellugi 1979), the predicted segmental switch did not occur. Instead, observed slips involved handshape, location, orientation, or handedness (one *vs.* two hands) features, with handshape involvement being the most common.

In one slip involving BLACK and WHITE, the handshape sequence in WHITE (open fingers and thumb changing to closed fingers and thumb touching at tips) is anticipated in BLACK, with its regular handshape completely replaced by the handshape change from WHITE, whereas its location (at the forehead) and movement direction (brushing across) remained unaffected (Klima and Bellugi 1979: 139). What did not happen was a complete replacement of the initial handshape, location, and orientation of BLACK with those of WHITE, which would have created a form that started at the chest, not at the forehead. In none of their examples did the features act together as a group, as would be predicted from segmental models.

6.2.3 *Backwards signing*

Backwards signing demonstrates that signers have access both to syllable sequences and to individual features within syllables which can be exchanged in temporal sequence, but not to units corresponding to segments as defined by the segmental models. This contrasts with the evidence from spoken language games and backwards speaking (Sherzer 1970; Cowan and Leavitt 1981, 1990; Cowan *et al.* 1982; Treiman 1983; Cowan *et al.* 1985; Cowan *et al.* 1987; Cowan 1989). Cowan *et al.* (1985: 680) report that fluent backwards talkers segment speech into "phonemic or syllabic" units, and then reverse their order. Their subjects fall into two groups, using either orthography or phonology as the basis for reversal. For example, for 'terrace', orthographic reversers would say /ekaret/, including the final "silent e" and adjusting the pronunciation of the letter "c" followed by back vowels to /k/. Phonological reversers would say /sıret/, simply reversing phonological segment order.

Data from backwards signing (Wilbur and Petersen 1997) provide evidence that signers treat monosyllabic signs in ways that are not compatible with segmental models. For example, Liddell's (1984) representation for THINK in (5) consists of two segments MH, i.e. M-[approach](AP) followed by H with contact. The exchange of these segments should yield HM, i.e. H with contact followed by M-[approach]. Signers actually produce "contact" followed by "move *away*," a result which would be predicted if the starting and ending locations of the movement are exchanged. That is, if movement is represented as a sequence of features, say [–contact] [+contact] (5b) (or [neutral] [forehead] rather than [approach]), then exchanging those features within-syllable to [+contact] [–contact] will result in the correct prediction of movement away from the forehead. Note that in Liddell's model in (5a), there is a sequence of [–contact] [+contact], but these features cannot switch independently of the segments AP and H to which they are assigned. This is what we mean by having the segments at the *top* of the model.

(5) a. *Liddell's model of* THINK

	THINK	
Segment	AP	H
Handshape	1	1
Orientation	TI	TI
Location	FH	FH
Contact	–	+
Non-manual markings	–	–

b. *Feature representation of change seen in backwards signing of* THINK

```
        σ                              σ
      /   \                          /   \
[−contact] [+contact]   →   [+contact] [−contact]
```

Incorrect predictions from segmental models are more obvious with the sign FLY, represented as a single M segment. The predicted backwards version should be the same as the original, because there is nothing available to exchange. Backwards signing shows that the direction of movement of FLY is reversed (Figure 56.10), comparable to the movement reversal in THINK. This remains an inexplicable fact in segmental models which treat movement as a single M segment with its own feature matrix. The only recourse is to change the representation to H_1MH_2 and then reverse the two Hs, but evidence for the presence of those H segments would need to be provided. In any case, the lack of analogy with spoken segment sequences can be seen: the backwards form of *cat* /kæt/ is /tæk/, with the vowel unchanged. But, clearly, in the backwards form of FLY the movement has changed.

Any segmental model containing M segments will have the same problem, because it is the phonological features associated with the movement that must be available to signers to be exchanged. In backwards signing, movements are consistently reversed by exchanging end specification with start specification, as though initial and final features are exchanged on their own tiers: end location with beginning location; end handshape with beginning handshape; end orientation with beginning orientation. Wilbur and Petersen (1997) argue that movement is not *inside* the syllable, but rather that movement *is* the syllable, a conception of "syllable" that takes movement as a dynamic gesture with only starting and ending specifications for the movement trajectory and no further linguistically meaningful internal specifications (see current arguments from gestural phonology approaches, e.g. Mauk and Tyrone 2008; Tyrone and Mauk 2008). For speech, Bagemihl (1989) and Pierrehumbert and Nair (1995) argue that sub-syllabic constituents, such as onset and rime or coda, do not participate in language game behavior, with Pierrehumbert and Nair extending this observation to phonological theory in general, claiming that these sub-syllabic constituents do not exist and that a flat syllable model, such as that proposed by Clements and Keyser (1983), is adequate to account for the facts. Bagemihl (1989: 485f.) notes that language

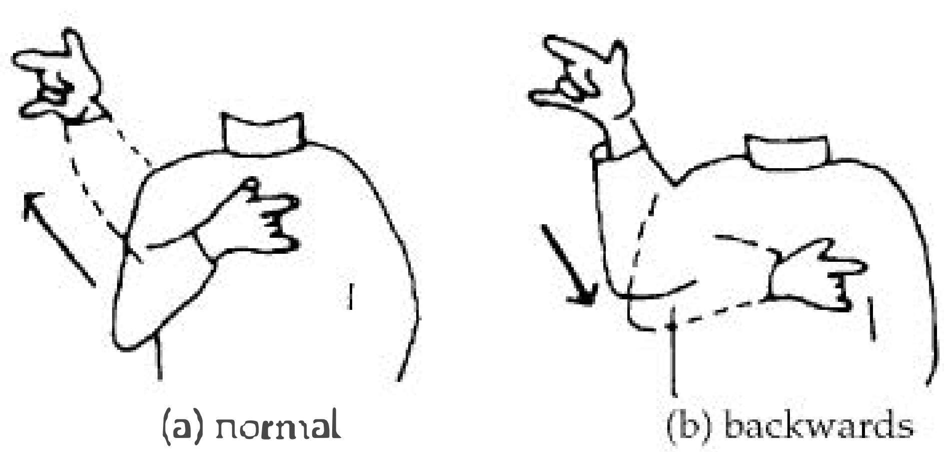


Figure 56.10 The sign FLY, made (a) with normal movement and (b) in backwards signing

games in cultures lacking phonemic alphabet writing systems do not use segments, only syllables. Furthermore, children do not use segmental language games until they are exposed to those writing systems. He suggests that alphabetic writing systems may be necessary for the development of metalinguistic awareness of segments as opposed to syllables.

Brentari (1998) capitalizes on these and other observations by separating the specifications that do not change during a syllable (Inherent Features) from those that do (Prosodic Features). Each syllable has two timing slots, one after the other, representing sequentiality, and the Prosodic Features are associated accordingly. The Inherent Features spread across both slots. Thus, her timing slots (sequentiality) are at the bottom of the tree, whereas for Sandler and Liddell, the sequentiality is at the top of their models. Jantunen and Takkinen (2010) review the sign language studies, and note that there is no evidence for internally structured sequential segmental syllables of the kind found in spoken languages (such as an onset–nucleus-coda distinction). Hence there is no justification for positing an intermediate level between segment (referring here to the timing slots in Brentari's model) and syllable, or more than two segments/slots per syllable. Finally, another benefit of the Prosodic Model for sign languages is that it provides seamless access to the prosodic hierarchy above the syllable.

7 What does simultaneity in syllable structure buy us?

It is time to turn our attention to the benefits of the notion of simultaneity in sign syllables, that is, what it accounts for that the other approaches do not. There are two important concepts that come from this model of syllable representation, namely the notion of *syllable weight* and the analogue to spoken language *sonority*. In addition, aspectual reduplication can be seen to operate prosodically on verb *roots*, whether one or two syllables.

7.1 *Syllable weight*

Consider the difference in speech between syllables of different structures CV, CVC, CVCC, CCVC, CCVCC. It is easy to identify an increase in syllable weight as more consonants are added to these syllables, even without knowing what those consonants are or the type of vowel in the nucleus or whether there is a distinction between short and long vowels or open syllable vowel lengthening (CHAPTER 57: QUANTITY-SENSITIVITY).

If sign syllables do not have the same internal structure as spoken language syllables, then is there a syllable weight distinction in sign languages, and, if so, how does it manifest itself? Brentari (1998) argues that there is a weight distinction in ASL, based on the number of simultaneous movements specified for the syllable. Syllables with one movement are light and those with two are heavy; more technically, a weight unit is constructed for every prosodic foot. With this analysis, she can explain the pattern of verbs that can and cannot take reduplication to form nouns, i.e. respectively light and heavy verbs. For example, the sign FLY in Figure 56.10a above is able to form a repeated nominalization for AIRPLANE because it is a light (one-movement) syllable. In contrast, syllables with complex movements (for example, a path movement combined with a handshape change) cannot undergo reduplicated nominalization, even if the verb qualifies semantically. Similarly, activity verbs that form activity nouns with the addition of the feature [trilled movement] must have light syllable structure to start with (Brentari 1998: 242–243). Brentari also shows the correlation between verb heaviness and preference for sentence-final position, that is, the word order is sensitive to the weight of the verb, which is determined by the number of movements in the syllable (and if reduplicated, the number of syllables).

In Brentari's analysis, the maximum number of weight units per syllable is two. Using data from Finnish Sign Language (FinSL), Jantunen (2005; see also Jantunen and Takkinen 2010) argues that an extended system is necessary for FinSL because more than two weight units per syllable are possible if one takes the non-manual movements into account – in fact, three or four may be possible. For Jantunen, a movement is complex (not simple) if more than one articulator is involved. He is then able to make a weight distinction between two monosyllabic signs, MUSTA 'black' which has only path movement, hence one weight unit, and UJO 'shy', which has local movement accompanied by a head movement, and hence has two weight units. Thus, even though both are monosyllabic, the difference in weight results from the non-manual head specification. Another benefit of this line of reasoning is that in FinSL there are lexical items which are made entirely with non-manual articulation (there are a few in ASL also) – these would be assigned a single weight unit, and that is the desired result. One additional generalization can be stated: both FinSL and ASL prefer syllables with simple movements over complex movements (Brentari 1998; Jantunen and Takkinen 2010). This generalization would be lost from a segmental perspective on syllable structure.

7.2 *Sonority*

Sonority is not built on syllable weight, as a mora-based generalization might suggest (Perlmutter 1992). Brentari suggests that sonority be approached as multidimensional salience. She suggests that sonority is correlated articulatorily with closeness of the articulator to the body's midline, and that articulation closer to the midline has greater visual salience than articulation further away. Thus, strengthening of visual salience (Enhancement Theory; Stevens and Keyser 1989) by choice of articulator higher up on the hierarchy (and likewise, reduction by choice lower down) is captured directly by the Prosodic Model in a way that segmental models cannot (Brentari 1998: 135). She suggests the following hierarchy, repeated here from (1) above:

(6) *Brentari hierarchy*

shoulder joint	>	elbow joint	>	wrist joint	>	base finger joints	>	non-base finger joints

This distinction can be observed in cases where a movement can be articulated by different articulators, that is, if a wrist movement and an elbow movement can convey the same phonology. In such cases, if the movement is made by an articulator that is up the hierarchy, say an elbow joint replacing a wrist movement, then "proximalization" is said to occur, whereas if the articulator used is down the hierarchy, from elbow to wrist, then "distalization" is said to happen (see Mirus *et al.* 2001 for an empirical test of the factors involved). Crasborn (2001) likewise provided evidence in NGT for some of the factors relating to proximalization/distalization. An important methodological aspect of Crasborn's study is that it looked at fluent L1 signers, whereas Mirus *et al.* looked at L2 acquisition, for which the presence of sonority effects might be obscured by developmental performance factors.

Looking at data from British Sign Language and the echoing of manual movements by the signer's mouth, Woll (2001) suggests than non-manual articulations can also provide insight into the sonority hierarchy. Based on detailed investigation of these aspects of Finnish Sign Language, Jantunen (2005, 2006, 2007) suggests that non-manual movement should be included in the hierarchy, as in (7), from Jantunen (2005: 56):

(7) upper body and head > hands (including Brentari's hierarchy) > mouth

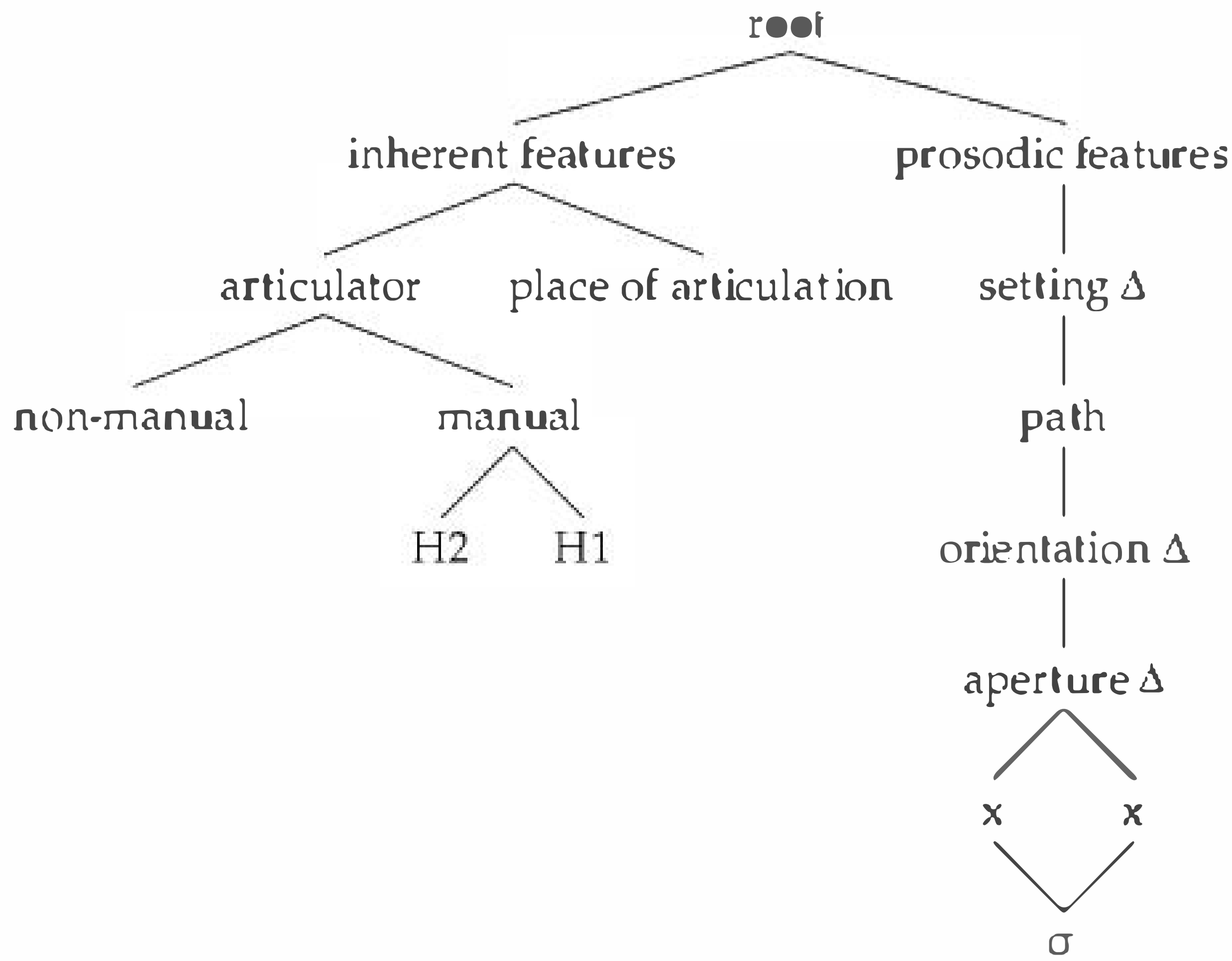
7.3 *Another perspective on reduplication: Templatic vs. prosodic*

Klima and Bellugi (1979) treat reduplication as part of a templatic approach to aspectual modification (CHAPTER 100: REDUPLICATION).[1] They refer to formational terms such as Planar locus (horizontal, vertical), Cyclicity (repetition), Direction (e.g. upward, downward), Geometric array (line, arc, circle, other arrangement), Quality (small, large), and Manner (continuous, hold, restrained). Thus, each morphological function (e.g. iterative, durative) involves a template composed of some of these formational features. But the choices of feature combinations in each template are not explained. Similarly, Sandler (CHAPTER 24: THE PHONOLOGY OF MOVEMENT IN SIGN LANGUAGE) argues for a templatic approach to reduplication, using additional M (movement) or L (location) segments to account for differences in movement type or final holding (what Klima and Bellugi refer to as "end marking").

Further discussion of reduplication with Klima and Bellugi led us to an interesting separation of function for spatial and temporal formational properties, with the spatial properties providing information about the arguments of the verb and the temporal/rhythmic properties providing information about aspect on the verb (Wilbur *et al.* 1983). We speculated that reduplication could be analyzed the same way as in spoken languages. It took over 20 years to work it out, but a standard Base-Copy reduplication approach can be applied to sign languages using the Prosodic Model (Wilbur 2005, 2009). In Brentari's model (8), the node dominating syllables and associated features is the root (CHAPTER 24: THE PHONOLOGY OF MOVEMENT IN SIGN LANGUAGE).

[1] It is important to distinguish repetition from reduplication. Here, repetition is viewed as prosodically driven, to fill the needs of a prosodic foot. Lexicalized repetition creates nouns from verbs, with only two formations of the lexically meaningful movement required. Reduplication is aspectually driven.

(8) *Brentari's Prosodic Model of syllable structure*



Whether one syllable or two in a root, the Base for reduplication is the root, and the entire Base is copied. A simple example with Base and Copies is illustrated in Figure 56.11; planar difference indicates argument differences.

There are, however, two modality differences between standard Base–Copy and what occurs in sign languages. First, multiple copies are common (Figure 56.11); indeed, a single copy implies "dual," so aspectual reduplication typically has two or more copies of the Base. A second difference is that for many aspectual reduplications the hand must return to its initial position in the Base before it can articulate the Copy, thus the sequence is Base–Return–Copy. Aspectual reduplication is a combinatorial system of Base event (verb root), followed by Return to initial position, which reflects the time between the end of the Base event and the onset of the repeated event in aspects involving iteration (Wilbur 2005, 2009). Different aspects determine the size of the Return (smaller than, equal to or greater than the Base) (Table 56.1). Whether the shape includes a stopping point or is smooth (circular, elliptical) is dependent entirely on whether the verb is telic (contains a stop) or atelic (cannot stop) (Wilbur 2008).

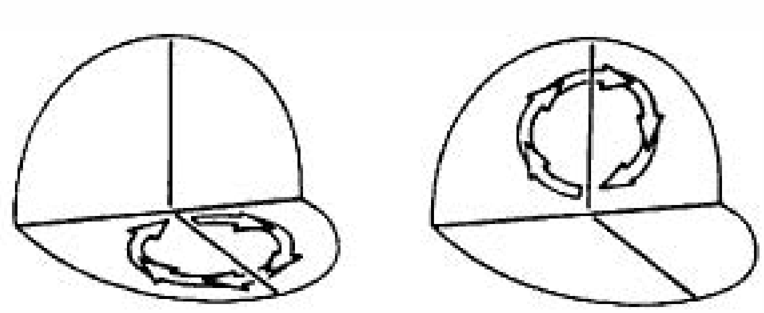


Figure 56.11 Base–Copy reduplication schemata for apportionative external and internal. By permission of Ursula Bellugi, The Salk Institute for Biological Studies

Table 56.1 Combinations of Return options and Base event type yield aspectual inflections

time between events	*telic event root*	*atelic event root*
[return = root]	habitual	durative
[return < root]	incessant	n/a
[return > root]	iterative	continuative

For telic events (Figure 56.12), when Return and Base are equal in size, there is the appearance of equal prominence on both (habitual); when the Return is smaller, there is a tendency for the Base to reduce as well (incessant). When the Return is larger than the Base, an arc is added (the morpheme EXTRA; Wilbur 2008). Thus, featurally, incessant aspect has [repeat] [return] [less than], habitual [repeat] [return] [equal] and iterative [repeat] [return] [greater than].

For atelic events (Figure 56.13), only two of the three options are possible, and the Base must be curved. Both of these requirements result from the absence of stops in the formation of atelic roots. Durative has [repeat] [return] [equal] and continuative has [repeat] [return] [greater than]. The atelic equivalent of the incessant, [repeat] [return] [less than], is not possible, because shortened movements would be perceptually equivalent either to stops, creating confusion with telics, or to trilled movement, which has a different interpretation (stative, not repeated). The difference between the modifications shown in Figures 56.12 and 56.13 is the Base root, which reflects the event structure in the semantics of the verb.

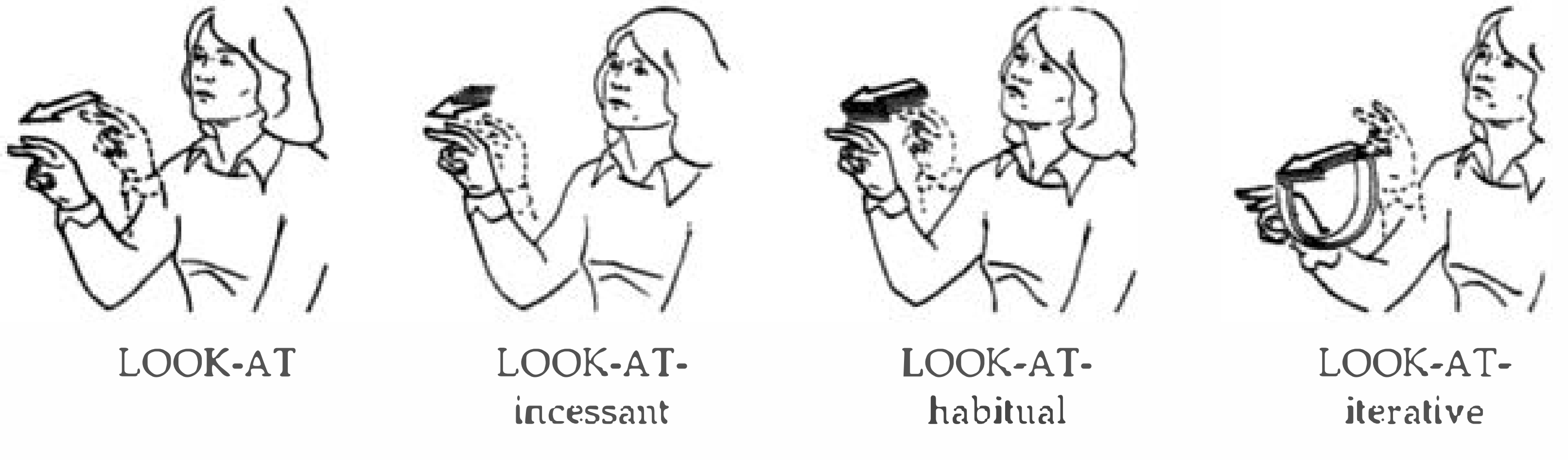


Figure 56.12 LOOK-AT and three inflections. By permission of Ursula Bellugi, The Salk Institute for Biological Studies

Figure 56.13 LOOK-stative and durative and continuative aspects. By permission of Ursula Bellugi, The Salk Institute for Biological Studies

These observations are not captured by a templatic approach to reduplication. Sandler's (1989, 2008) evidence for templatic analysis (for signs like BLOW-TOP, FAINT, and SNOOP) are all compatible with the generalization that aspectual reduplication copies the verb root. She argues that for the sign BLOW-TOP, which is a two-syllable compound created from HEAD and EXPLODE-OFF, EXPLODE-OFF is copied but not HEAD. Her explanation is that the rightmost M is copied in reduplication, whereas in lexicalized monosyllabic (single M) signs like FAINT (originally from MIND+DROP), the whole form is repeated. But in the alternative analysis presented here it is expected that EXPLODE-OFF will be copied. Similarly, Sandle cites SNOOP (from NOSE+STICK-IN) as an exception, with the syllable associated with STICK-IN reduplicating, even though the entire form is monosyllabic (single M). An explanation for this is that the initial movement *to the nose* for NOSE is purely epenthetic (as for MOTHER (Figure 56.3), and one of the two versions of PARENTS in §4.3). That is, SNOOP starts its STICK-IN movement at the nose but does not return there for subsequent repetitions. If so, it might be appropriate to consider the initial location at the nose to be something akin to a prefixal location adjoined/cliticized onto the beginning location of STICK-IN, resulting from the compound-to-lexical-item reduction process, leaving STICK-IN to be copied by reduplication with its original location, as shows up in subsequent repetitions.

This discussion highlights the kind of phonological level analysis that can be conducted with respect to syllables and their contents. Thus, factors other than phonology, such as verbal telicity and type of aspectual morphology, affect the final form of reduplicated signs (Wilbur 2005, 2009).

8 Conclusion

In this chapter, we started with two basic criteria for discussion of syllables in sign languages: linguistic meaningfulness and reliable measurement. To address the requirements of these criteria, we introduced the reader to relevant historical discussions of the status of syllables in sign languages. The key is that movement is the defining feature of a phonological syllable in sign languages. Then, we provided evidence that syllables in sign languages can be reliably measured. As for the criteria of linguistic meaningfulness, we reviewed several phenomena for which one has to make reference to the syllable for a reasonable account. Among those discussed were the phonologization of fingerspelled loan words, contact metathesis, handshape changes within and between signs, and some prosodic constraints.

We then turned to the debate concerning the formal representation of sign syllables. We reviewed the sequential and simultaneous models of syllable representation. Data presented – tapping, slips of the hand, and backwards signing – strongly favor the Prosodic Model proposed by Brentari (1998), resulting in a syllable model that is internally different from those that exist in speech. We then considered the implications of this conclusion, especially since it goes against expectations of similarity between signed and spoken languages. Issues of sonority, syllable weight, and Base–Copy reduplication indicate that the syllable performs similarly in sign language and spoken language phonologies despite the internal differences in organization.

To get to the point where we can use syllables for explaining other phenomena, we have needed a consistent and well-developed syllable model that makes empirically testable claims. This model has been tested in a variety of ways reviewed in this chapter, and there is more that has been omitted for reasons of space (variability of non-manual marking; judgments of well-formed syllables; Brentari and Wilbur 2008). The critical feature of the model is that it is a *prosodic* model, and the lowest level of the prosodic hierarchy is the syllable, for sign languages as well as spoken languages.

ACKNOWLEDGMENTS

The preparation of this chapter was funded in part by NIH-NIDCD: "Prosodic features of American Sign Language" (1993–98); NIH-NIDCD: "Modeling the nonmanuals of American Sign Language" (2004–10); NSF: "A basic grammar of Croatian Sign Language" (2004–10); NSF: "Syllables in American Sign Language" (1984–87). A special acknowledgment to Diane Brentari for discussion in the process of developing this chapter, and to Kadir Gökgöz for suggestions and revisions.

REFERENCES

Adams, Corinne. 1979. *English speech rhythm and the foreign learner*. The Hague: Mouton.

Akamatsu, Carol. 1982. The acquisition of fingerspelling in pre-school children. Ph.D. dissertation, University of Rochester.

Akamatsu, Carol. 1985. Fingerspelling formulae: A word is more or less than the sum of its letters. In Stokoe & Volterra (1985), 126–132.

Allen, George. 1972. The location of rhythmic stress beats in English speech, Parts I & II. *Language and Speech* 15. 72–100, 179–195.

Allen, George, Ronnie Wilbur & Brenda Schick. 1991. Aspects of rhythm in American Sign Language. *Sign Language Studies* 72. 297–320.

Bagemihl, Bruce. 1989. The crossing constraint and "backwards languages." *Natural Language and Linguistic Theory* 7. 481–549.

Battison, Robbin. 1978. *Lexical borrowing in American Sign Language*. Silver Spring, MD: Linstok Press.

Bellugi, Ursula & Susan Fischer. 1972. A comparison of sign language and spoken language: Rate and grammatical mechanisms. *Cognition* 1. 173–200.

Blevins, Juliette. 1993. The nature of constraints on the nondominant hand in ASL. In Coulter (1993), 43–61.

Brentari, Diane. 1990a. Licensing in ASL handshape change. In Lucas (1990), 57–68.

Brentari, Diane. 1990b. Underspecification in American Sign Language phonology. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 16. 46–56.

Brentari, Diane. 1993. Establishing a sonority hierarchy in American Sign Language: The use of simultaneous structure in phonology. *Phonology* 10. 281–306.

Brentari, Diane. 1994. Prosodic constraints in American Sign Language. In Helen Bos & Trude Schermer (eds.) *Sign Language Research 1994*, 39–51. Hamburg: Signum Press.

Brentari, Diane. 1996. Trilled movement: Formal representation and phonetic representation. *Lingua* 98. 43–71.

Brentari, Diane. 1998. *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

Brentari, Diane & Laurinda Crossley. 2002. Prosody on the hands and face: Evidence from American Sign Language. *Sign Language and Linguistics* 5. 105–130.

Brentari, Diane & Howard Poizner. 1994. A phonological analysis of a Deaf Parkinsonian signer. *Language and Cognitive Processes* 9. 69–100.

Brentari, Diane & Ronnie Wilbur. 2008. A cross-linguistic study of word segmentation in three sign languages. In Ronice Müller de Quadros (ed.) *Sign languages: Spinning and unraveling the past, present and future – Theoretical Issues in Sign Language Research* 9, 48–63. Petrópolis, Brazil: Editora Arara Azul.

Chinchor, Nancy. 1978. The syllable in ASL. Paper presented at the MIT Sign Language Symposium.

Chinchor, Nancy. 1981. Numeral incorporation in American Sign Language. Ph.D. dissertation, Brown University.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable.* Cambridge, MA: MIT Press.

Corina, David. 1990. Handshape assimilation in hierarchical phonological representation. In Lucas (1990), 27–49.

Coulter, Geoffrey. 1982. On the nature of ASL as a monosyllabic language. Paper presented at the 56th Annual Meeting of the Linguistic Society of America, San Diego.

Coulter, Geoffrey. 1990. Emphatic stress in ASL. In Fischer & Siple (1990), 109–126.

Coulter, Geoffrey (ed.) 1993. *Current issues in ASL phonology.* New York: Academic Press.

Covington, Virginia. 1973. Features of stress in American Sign Language. *Sign Language Studies* 2. 39–58.

Cowan, Nelson. 1989. Acquisition of Pig Latin: A case study. *Journal of Child Language* 16. 365–386.

Cowan, Nelson & Lewis Leavitt. 1981. Talking backward: Exceptional speech play in late childhood. *Journal of Child Language* 9. 481–491.

Cowan, Nelson & Lewis Leavitt. 1990. Speakers' access to the phonological structure of the syllable in word games. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26. 45–59.

Cowan, Nelson, Lewis Leavitt, Dominic Massaro & Raymond Kent. 1982. A fluent backward talker. *Journal of Speech and Hearing Research* 25. 48–53.

Cowan, Nelson, Martin Braine & Lewis Leavitt. 1985. The phonological and metaphonological representation of speech: Evidence from fluent backward talkers. *Journal of Memory and Language* 24. 679–698.

Cowan, Nelson, Cristy Cartwright, Carrie Winterowd & Molly Sherk. 1987. An adult model of preschool children's speech memory. *Memory and Cognition* 15. 511–517.

Crasborn, Onno. 2001. Phonetic implementation of phonological categories in sign language of the Netherlands. Ph.D. dissertation, University of Leiden.

Fischer, Susan D. & Patricia Siple (eds.) 1990. *Theoretical issues in sign language research, vol. 1: Linguistics.* Chicago: University of Chicago Press.

Friedman, Lynn A. 1974. On the physical manifestation of stress in the American Sign Language Unpublished ms., University of California, Berkeley.

Friedman, Lynn A. 1976. Phonology of a soundless language: Phonological structure of American Sign Language. Ph.D. dissertation, University of California, Berkeley.

Friedman, Lynn A. (ed.) 1977. *On the other hand: New perspectives on American Sign Language.* New York: Academic Press.

Frishberg, Nancy. 1978. The case of the missing length. *Communication and Cognition* 11. 57–67.

Fromkin, Victoria A. 1971. The non-anomalous nature of anomalous utterances. *Language* 47. 27–52.

Fromkin, Victoria A. 1973. Slips of the tongue. *Scientific American* 229. 109–117.

Geraci, Carlo. 2009. Epenthesis in Italian Sign Language. *Sign Language and Linguistics* 12. 3–51.

Green, Kerry. 1984. Sign boundaries in American Sign Language. *Sign Language Studies* 42. 65–91.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress.* Cambridge, MA: MIT Press.
Hoequist, Charles, Jr. 1983. Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica* 40. 203–237.
Hulst, Harry van der. 1993. Units in the analysis of signs. *Phonology* 10. 209–241.
Jantunen, Tommi. 2005. Mistä on pienet tavut tehty? Analyysi suomalaisen viittomakielen tavusta prosodisen mallin viitekehyksessä [What are the little syllables made of? A Prosodic Model account of the Finnish Sign Language syllable]. Licentiate thesis, University of Jyväskylä, Finland.
Jantunen, Tommi. 2006. The complexity of lexical movements in FinSL. *SKY Journal of Linguistics* 19. 335–344.
Jantunen, Tommi. 2007. Tavu suomalaisessa viittomakielessä [The syllable in Finnish Sign Language] (with English abstract). *Puhe ja kieli* 27. 109–126.
Jantunen, Tommi & Ritva Takkinen. 2010. Syllable structure in sign language phonology. In Diane Brentari (ed.) *Sign languages*, 312–331. Cambridge: Cambridge University Press.
Johnson, Robert. 1986. Metathesis in American Sign Language. Paper presented at the Conference on Theoretical Issues in Sign Language Research, Rochester, NY.
Kegl, Judy & Ronnie Wilbur. 1976. When does structure stop and style begin? Syntax, morphology and phonology vs. stylistic variation in American Sign Language. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 12. 376–396.
Klima, Edward S. & Ursula Bellugi. 1979. *The signs of language.* Cambridge, MA: Harvard University Press.
Kooij, Els van der. 2002. Phonological categories in Sign Language of the Netherlands: The role of phonetic implementation and iconicity. Ph.D. dissertation, University of Leiden.
Kooij, Els van der & Onno Crasborn. 2008. Syllables and the word-prosodic system in Sign Language of the Netherlands. *Lingua* 118. 1307–1327.
Liddell, Scott K. 1978. Non-manual signals and relative clauses in ASL. In Siple (1978), 59–90.
Liddell, Scott K. 1980. *American sign language syntax.* The Hague: Mouton.
Liddell, Scott K. 1984. THINK and BELIEVE: Sequentiality in American Sign Language. *Language* 60. 372–399.
Liddell, Scott K. & Robert E. Johnson. 1989. American Sign Language: The phonological base. *Sign Language Studies* 64. 197–277.
Lucas, Ceil (ed.) 1990. *Sign language research: Theoretical issues.* Washington, DC: Gallaudet University Press.
Mauk, Claude & Martha Tyrone. 2008. Sign lowering as phonetic reduction in American Sign Language. *Proceedings of the 8th International Speech Production Seminar*, 185–188. Available (September 2010) at http://issp2008.loria.fr/proceedings.html. Strasbourg.
Meier, Richard. 1993. A psycholinguistic perspective on phonological segmentation in sign and speech. In Coulter (1993), 169–188.
Miller, Christopher. 1997. Phonologie de la langue des signes québecoise: Structure simultanée et axe temporel. Ph.D. dissertation, Université de Québec à Montréal.
Mirus, Gene, Christian Rathmann & Richard Meier. 2001. Proximalization and distalization of sign movement in adult learners. In Valery L. Dively, Melanie Metzger, Sarah Taub & Anne Marie Baer (eds.) *Signed languages: Discoveries from international research*, 103–119. Washington, DC: Gallaudet University.
Newkirk, Donald. 1979. The form of the continuative aspect on ASL verbs. Unpublished ms., Salk Institute for Biological Studies, La Jolla, CA. Reprinted 1998 in *Sign Language and Linguistics* 1. 75–80.
Newkirk, Donald. 1980. Rhythmic features of inflection in American Sign Language. Unpublished ms., Salk Institute for Biological Studies, La Jolla, CA. Reprinted 1998 in *Sign Language and Linguistics* 1. 81–100.

Newkirk, Donald. 1981. On the temporal segmentation of movement in American Sign Language. Unpublished ms., Salk Institute for Biological Studies, La Jolla, CA. Reprinted 1998 in *Sign Language and Linguistics* 1. 59–97.

Padden, Carol. 1993. Response to Sandler's "Linearization of phonological tiers in ASL." In Coulter (1993), 131–134.

Perlmutter, David. 1992. Sonority and syllable structure in American Sign Language. *Linguistic Inquiry* 23. 407–442.

Pierrehumbert, Janet B. & Rami Nair. 1995. Word games and syllable structure. *Language and Speech* 38. 77–114.

Sandler, Wendy. 1986. The spreading hand autosegment of American Sign Language. *Sign Language Studies* 50. 1–28.

Sandler, Wendy. 1989. *Phonological representation of the sign: Linearity and non-linearity in American Sign Language.* Dordrecht: Foris.

Sandler, Wendy. 1993. A sonority cycle in American Sign Language. *Phonology* 10. 243–279.

Sandler, Wendy. 2008. The syllable in sign language: Considering the other natural modality. In Barbara Davis & Kristine Zajdo (eds.) *The syllable in speech production*, 379–408. New York: Taylor Francis.

Sandler, Wendy & Diane Lillo-Martin. 2006. *Sign language and linguistic universals.* Cambridge: Cambridge University Press.

Schmidt, Richard & Tim Lee. 2005. *Motor control and learning: A behavioral emphasis.* Urbana, IL: Human Kinetics.

Sherzer, Joel. 1970. Talking backwards in Cuna: The sociological reality of phonological descriptions. *Southwestern Journal of Anthropology* 26. 343–353.

Siple, Patricia (ed.) 1978. *Understanding language through sign language research.* New York: Academic Press.

Stevens, Kenneth N. & Samuel J. Keyser. 1989. Primary features and their enhancement in consonants. *Language* 65. 81–106.

Stokoe, William C. 1960. *Sign language structure: An outline of the visual communication systems of the American deaf.* Silver Spring, MD: Linstok Press.

Stokoe, William C. & Virginia Volterra (eds.) 1985. *SLR '83: Proceedings of the 3rd International Symposium on Sign Language Research.* Rome: Consiglio Nazionale delle Ricerche.

Supalla, Ted & Elissa Newport. 1978. How many seats in a chair? The derivation of nouns and verbs in American Sign Language. In Siple (1978), 91–132.

Treiman, Rebecca. 1983. The structure of spoken syllables: Evidence from novel word games. *Cognition* 15. 49–74.

Tyrone, Martha & Claude Mauk. 2008. Sign lowering in ASL: The phonetics of wonder. Paper presented at the SignTyp Conference, Storrs, CT.

Tyrone, Martha, Louis Goldstein & Gaurav Mathur. 2007. Movement kinematics and prosody in American Sign Language. Poster presented at the 20th Annual CUNY Conference on Human Sentence Processing, La Jolla, CA.

Valli, Clayton & Ceil Lucas. 1992. *Linguistics of American Sign Language.* Washington, DC: Gallaudet University.

Weast, Traci. 2008. Questions in American Sign Language: A quantitative analysis of raised and lowered eyebrows. Ph.D. dissertation, University of Texas, Arlington.

Wilbur, Ronnie. 1979. *American Sign Language and sign systems: Research and applications.* Baltimore: University Park.

Wilbur, Ronnie. 1985. Towards a theory of "syllable" in signed languages: Evidence from the numbers of Italian Sign Language. In Stokoe & Volterra (1985), 160–174.

Wilbur, Ronnie. 1987. *American Sign Language: Linguistic and applied dimensions.* Boston: College Hill Press.

Wilbur, Ronnie. 1990a. An experimental investigation of stressed sign production. *International Journal of Sign Language* 1. 41–60.

Wilbur, Ronnie. 1990b. Why syllables? What the notion means for ASL research. In Fischer & Siple (1990), 81–108.

Wilbur, Ronnie. 1993. Segments and syllables in ASL phonology. In Coulter (1993), 135–168.

Wilbur, Ronnie. 1994. Eyeblinks and ASL phrase structure. *Sign Language Studies* 84. 221–240.

Wilbur, Ronnie. 1999a. Metrical structure, morphological gaps, and possible grammaticalization in ASL. *Sign Language and Linguistics* 2. 217–244.

Wilbur, Ronnie. 1999b. Stress in ASL: Empirical evidence and linguistic issues. *Language and Speech* 42. 229–250.

Wilbur, Ronnie. 2005. A reanalysis of reduplication in American Sign Language. In Bernhard Hurch (ed.) *Studies in reduplication*, 593–620. Berlin & New York: Mouton de Gruyter.

Wilbur, Ronnie. 2008. Complex predicates involving events, time and aspect: Is this why sign languages look so similar? In Josep Quer (ed.) *Signs of the time: Selected papers from TISLR 2004*, 217–250. Hamburg: Signum.

Wilbur, Ronnie. 2009. Effects of varying rate of signing on ASL manual signs and nonmanual markers. *Language and Speech* 52. 245–285.

Wilbur, Ronnie & George Allen. 1991. Perceptual evidence against internal structure in ASL syllables. *Language and Speech* 34. 27–46.

Wilbur, Ronnie & Susan Nolen. 1986. Duration of syllables in American Sign Language. *Language and Speech* 29. 263–280.

Wilbur, Ronnie & Lesa Petersen. 1997. Backwards signing and ASL syllable structure. *Language and Speech* 40. 63–90.

Wilbur, Ronnie & Brenda Schick. 1987. The effects of linguistic stress on ASL signs. *Language and Speech* 30. 301–323.

Wilbur, Ronnie, Edward Klima & Ursula Bellugi. 1983. Roots: On the search for the origins of signs in ASL. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 19. 314–336.

Wilcox, Sherman. 1992. *The phonetics of fingerspelling*. Amsterdam & Philadelphia: John Benjamins.

Woll, Bencie. 2001. The sign that dares to speak its name: Echo phonology in British Sign Language. In Penny Boyes Braem & Rachel Sutton-Spence (eds.) *The hands are the head of the mouth: The mouth as articulator in sign languages*, 87–98. Hamburg: Signum.

57 Quantity-sensitivity

DRAGA ZEC

1 Introduction

Quantity-sensitivity is an important property of prosodic constituents, which are subclassified along this dimension as either light or heavy. In a typical hierarchical organization of prosodic units, as in (1) (Selkirk 1978, 1980; Nespor and Vogel 1986), each of the prosodic levels may be instantiated by constituents that vary in length, segment quality, or structural complexity (see CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE; CHAPTER 40: THE FOOT; CHAPTER 51: THE PHONOLOGICAL WORD; CHAPTER 84: CLITICS; CHAPTER 50: TONAL ALIGNMENT).

(1) *Prosodic hierarchy*

Prosodic phrase
Prosodic word
Foot
Syllable

This variation, in at least some of its aspects, introduces distinctions in quantity among constituents at the same level of the hierarchy, evidenced by distinctions in phonological behavior. While quantity-sensitivity is most clearly manifested at the level of the syllable, other prosodic levels exhibit this property as well. Quantity-sensitivity characterizes a wide range of phonological phenomena, including stress, tone, poetic meter, and various prosodic effects on morphosyntax. Moreover, quantity-sensitivity can be manifested either as a binary or as a scalar property. For these and other reasons to be addressed in this chapter, prosodic quantity needs to have its place in the formal representation of prosody, and is a central issue in any discussion of phonological representations.

The chapter is organized as follows: §2 addresses crucial aspects of weight-sensitivity in the syllable, providing a typology of weight patterns supported by a wide range of attested cases. §3 focuses on formal representations of the syllable and its weight, while §4 addresses the relevance of vowel length for quantity-sensitivity. §5 shows that weight distinctions could be binary in some languages, and multivalued in others. §6 documents inconsistencies in weight patterns, both with respect to phonological processes and phonological contexts. §7 addresses quantity-sensitivity in feet, focusing on binary patterns, and §8 focuses on scalar patterns of quantity. §9 touches upon quantity-sensitivity at the higher levels of the prosodic hierarchy, in prosodic words and prosodic phrases. §10 offers some remarks on markedness, and §11 concludes.

2 The syllable and quantity-sensitivity

Phonological quantity is primarily associated with the syllable. One of the traditional classifications of syllables is into those that are light and those that are heavy. This distinction is motivated on empirical grounds and is brought to relief by a number of quantity-sensitive phonological phenomena, including stress, tone, and poetic meter. We focus here only on those languages that do exhibit quantity-sensitivity at the syllable level, as this is not a universal prosodic property. In §2.1 we present a paradigm case of syllable weight, and then, in §2.2, turn to the typology of syllable-based quantity-sensitivity, supported by a wide range of phonological phenomena.

2.1 A paradigm case of quantity-sensitivity

Quantity-sensitivity has figured prominently in studies of classical languages and their prosody (Allen 1973). Latin provides a paradigm case of quantity-sensitivity, already known to the early grammarians such as Quintilian.

Latin stress (CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE) is quantity-sensitive, as illustrated in (2). If the penultimate syllable is heavy it is stressed, as in (2a), but if it is light the antepenultimate syllable is stressed, as in (2b).

(2) *Latin stress*

a. If the penultimate syllable is heavy, it is stressed.

for'tūna	'fortune (NOM SG)'
'gaudēns	'rejoicing (NOM SG)'
gau'dentem	'rejoicing (ACC SG)'

b. If the penultimate syllable is light, the antepenultimate is stressed.

'anima	'soul (NOM SG)'

Syllables that function as light are of the CV type, containing a short vowel, as the penultimate syllable in (2b). Syllables that function as heavy are more diverse, as shown in (2a). They are either of the CVV type, containing a long vowel or a diphthong, or of the CVC type, with a short vowel followed by a consonant. Significantly, syllables that are functionally equivalent may differ in their segmental content (CHAPTER 54: THE SKELETON). Within the set of heavy syllables, differences in segmental content are found not only across CVV and CVC syllables but also within the class of CVV syllables, which may contain either a long vowel or a diphthong. In this case, as in many others, onset consonants are excluded from the computation of weight (CHAPTER 55: ONSETS).

This same pattern of syllable weight also figures in Latin poetic meter. In one of the meters of Horace's *Odes*, known as the *First Asclepiad* (borrowed from

Greek), a line of verse contains a sequence of metrical positions that admit either heavy or light syllables (marked as – and ∪, respectively, with || marking the caesura), as in (3). In (3a), the first three metrical positions, all heavy, are filled with CVV syllables that contain either a diphthong or a long vowel, while in (3b) these same metrical positions are filled with CVC syllables. The light metrical positions, such as the next two, are filled in both lines with CV syllables. (The remaining metrical positions are filled in the same fashion, with the exception of the last vowel of *tollere* in (3b) which is elided, and therefore not scanned.)

(3) *Latin poetic meter: Horace,* First Asclepiad

a. – – – ∪∪ – || – ∪∪ – ∪ x
Maecēnās atavīs ēdite rēgibus
'O Maecenas, born from kingly ancestors!' (*Odes* 1.1.1)

b. – – – ∪ ∪ – || – ∪ ∪ – ∪ x
Certat tergeminīs tollere honoribus
'vies to lift [him] with triple magistracies.' (*Odes* 1.1.8)

Thus, in Latin, both stress and poetic meter are sensitive to distinctions in quantity, with weight characterized identically in the two phonological subsystems.

2.2 *Patterns of quantity-sensitivity*

The system of quantity in Latin exemplified in (2) and (3) was taken in much relevant work to be the standard mode of computing quantity, with broad empirical support. This is how syllable weight was characterized in Kuryłowicz (1948) and later in Newman (1972), among others. Newman (1972), in particular, identifies quantity-sensitivity in a number of languages, all exemplifying the pattern of quantity with light CV and heavy CVV and CVC syllables. In addition to Latin, the list includes Classical Greek, Finnish, Estonian, Classical Arabic, and Gothic, as well as three Chadic languages, Bolanci, Kanakuru, and Hausa. In fact, a number of researchers stated important generalizations about quantity-sensitivity solely in terms of the Latin pattern of weight (e.g. Kiparsky 1979, 1981; Halle and Vergnaud 1980; Clements and Keyser 1983).

The Latin pattern, however, is not the only empirically attested mode of computing quantity, as shown in Hyman's (1977) broad survey of stress systems and in much later work. In what follows we present the range of quantity patterns that have been empirically attested, and a typology of weight distinctions.

2.2.1 *Weight patterns: A typology*

McCarthy (1979) made the crucial theoretical statement that quantity-sensitivity can be instantiated in more than one way. In addition to the Latin weight pattern, with light CV and heavy CVV and CVC syllables as in (4a), henceforth type 1, there is a further weight pattern, one in which only CVV syllables are heavy and both CV and CVC syllables are light, as in (4b), henceforth type 2. A number of languages were identified to belong to this weight type: for example, Huasteco Mayan in Hyman (1977) and Yidin^y and Tiberian Hebrew in McCarthy (1979). Many more such cases figure in Hayes (1980, 1995) and Gordon (2006).

(4) *Possible weight patterns* (first approximation)

a. Type 1: heavy CVV, CVC *vs.* light CV
b. Type 2: heavy CVV *vs.* light CV, CVC

Thus the weight of CVC syllable is "parameterized": while in (4a) such syllables form a natural class with CVV, in (4b) they form a natural class with CV. This is crucially due to the status of the final consonant in a CVC syllable, which contributes to weight in (4a), but not in (4b) types of languages. McCarthy (1979) further identifies an important implicational relation: a language with heavy CVC syllables also has heavy CVV syllables. This supports the prediction about the following impossible weight pattern: no language can have heavy CVC but light CVV syllables.

But computation of quantity can be even more fine-grained than in (4), and in order to show this we invoke the sonority of segments. In particular, vowels are more sonorous than consonants, and within the class of consonants, sonorants (CHAPTER 8: SONORANTS) are more sonorous than obstruents. (For a general discussion of sonority, see CHAPTER 49: SONORITY.) In addition to the two weight systems in (4), one in which all consonants contribute to weight, and one in which no consonants contribute to weight, there is also a type 3 system, in which only some consonants contribute to weight (see Prince 1983; Zec 1988, 1995). In such split systems, the subset of consonants contributing to weight is generally sonorants. Such a case is exemplified by Kwakw'ala, to be discussed in §2.2.2, in which heavy syllables are CVV and CVR (sonorant), while light syllables are CV and CVO (obstruent). This yields the implicational relation that if CVO syllables are heavy, so are CVR syllables; and excludes the impossible system, with CVV and CVO syllables being heavy, and CV and CVR syllables being light (Zec 1995). Furthermore, while other splits in the hierarchy should in principle be possible, say, with liquids being weight-bearing to the exclusion of obstruents and nasals, such systems have not been attested. Only major splits within the sonority hierarchy appear to be exploited for distinctions in quantity, those in particular that correspond to splits imposed by the major class features (Chomsky and Halle 1968).

To summarize, a basic typology of weight patterns is given in (5). In type 1 languages, all segments contribute to weight, so that both CVV and CVC syllables are heavy; in type 2 languages only vowels are weight-bearing, which makes CVC syllables light; and in type 3 languages vowels and sonorant consonants are weight-bearing, to the exclusion of obstruents. The set of weight-bearing segments follows the sonority hierarchy: if a less sonorous segment contributes to weight, so does a more sonorous segment.

(5) *Typology of weight patterns*

	Heavy	*Light*
Type 1	CVV,CVC	CV
Type 2	CVV	CV, CVC
Type 3	CVV,CVR	CV, CVO

A special case of type 2 languages is those that lack CVC syllables. In a syllable inventory including only CV and CVV syllables, the former are light and the latter are heavy, as in Fijian (Hayes 1995; among others). Likewise, a special case of type 3 languages consists of those that lack CVO syllables, with an inventory that includes light CV and heavy CVV and CVR syllables, as in Tiv (Zec 1995); or Manam, in which the set of heavy syllables includes CVV and CVN (nasal), but excludes syllables closed with liquids (Lichtenberk 1983; Buckley 1998).

Crucially, the onset is excluded from the computation of weight: the number of segments in the onset does not affect the weight status of a syllable. This empirically grounded property of onsets will need to be captured in the representation of the syllable, an issue to be addressed in §3. But although broadly attested, this is not a universal property of onsets; see CHAPTER 55: ONSETS for cases of weight-sensitive onsets, which constitute counterexamples to this claim.

To conclude, it has been shown that there is a measure of language-specificity, with different modes of quantity computation employed in different languages. It has also been shown that there is an implicational relation across occurring weight patterns, or, more specifically, across the sets of heavy syllables in different languages, as in (6):

(6) *Implicational relations among heavy syllables*

a. If a language has heavy CVC syllables, it also has heavy CVV syllables.
b. If a language has heavy CVO syllables, it also has heavy CVR syllables.

2.2.2 *Weight patterns: Case studies*

The three patterns of quantity in (5) are documented below with two types of quantity-sensitive phonological phenomena: stress and tone. We begin with stress, which provides the most striking cases of quantity-sensitivity. It should be noted though that only some stress systems are quantity-sensitive. According to Gordon's (2006: 20–21) extensive survey, based on 408 languages, 310 languages have culminative accent systems. Out of those, 136 (43.9 percent) exhibit quantity-sensitivity, and 86 belong to one of the three weight systems we exemplify here.

Languages with quantity-sensitive stress show a clear preference for placing stress on heavy syllables (cf. Hyman 1977; Hayes 1980, 1995; Halle and Vergnaud 1987; Halle and Idsardi 1995). Simply stated, heavy syllables attract stress (Prince 1990). Moreover, languages with quantity-sensitive stress systems are of either type 1 or type 2, and rarely of type 3. The Latin stress pattern illustrated in §2.1, which belongs to type 1, is found in a number of languages. Out of 86 languages with quantity-sensitive stress in Gordon's survey, 42 languages are of type 1. It is found, for example, in Modern Classical Arabic (as described in Ryding 2005), where stress falls on the penultimate heavy syllable, CVV or CVC, otherwise on the antepenultimate syllable. Note, however, the pattern in Classical Arabic, where stress falls on the rightmost (non-final) heavy syllable, as in (7a), otherwise on the first syllable, as in (7b) (McCarthy 1979, and the references therein).[1]

[1] Final syllables have a special status, in at least two respects. Stress does not fall on final CVC syllables, but CVVC and CVCC syllables, which are only found word-finally, do bear stress. The special behavior of final elements is a more general issue, to be addressed in §6.2.

(7) *Type 1: Classical Arabic*

a.	ki'taabun	'book (NOM SG)'
	manaa'diilu	'kerchiefs (NOM)'
	ju'ʃaariku	'he participates'
	'mamlakatun	'kingdom (NOM SG)'
b.	'kataba	'he wrote'
	'balaħatun	'date (NOM SG)'

The type 1 weight pattern is also noted in English, although the overall stress system is rife with idiosyncrasies. A small portion of the English lexicon, the set of underived nouns, has a relatively regular stress pattern: stress falls on the heavy penult, either CVV as in *e'litist*, *ma'rina*, and *Ari'zona*, or CVC as in *a'genda*, *a'malgam*, and *co'nundrum*; otherwise on the antepenult, as in *'discipline*, *'labyrinth*, and *A'merica* (Hayes 1982). This stress pattern is again reminiscent of Latin. Many more type 1 stress systems are documented in Hayes (1995) and Gordon (2006).

Quantity-sensitive stress systems of type 2 are evidenced in a wide range of languages, just like type 1 (Hayes 1995; Gordon 2006; among others): 40 out of 86 quantity-sensitive stress systems in Gordon's survey. It is found, for example, in the Mongolian language Buriat (Poppe 1960; Walker 1996), illustrated in (8): stress falls on the initial syllable in words with no long vowels, as in (8a), and on the rightmost *non-final* heavy syllable in words with more than one long vowel or diphthong, as in (8b). If a word has only one CVV syllable, stress falls on that syllable even if it is final, as in (8c). Note that CVC syllables figure in the language, yet do not attract stress, for example, the third syllable in /ta'ruulagdaxa/, in (8b).

(8) *Type 2: Buriat*

a.	'xada	'mountain'
b.	mo'rʲooroo	'by means of his own horse'
	dalai'gaaraa	'by one's own sea'
	ta'ruulagdaxa	'to be adapted to'
c.	xa'daar	'through the mountain'

Another type 2 system is Huasteco Mayan (Larsen and Pike 1949; Hyman 1977; Hayes 1995: 296): stress falls on the rightmost CVV syllable, otherwise on the initial CV syllable. Again, CVC syllables pattern with CV rather than CVV syllables. And, in Aguacatec Mayan (McArthur and McArthur 1956; Hayes 1980, 1995), stress falls on a CVV syllable regardless of its position within a word, as in (9a); stress is final in words with no long vowels, as in (9b).[2]

(9) *Type 2: Aguacatec Mayan stress*

a. *Forms with* CVV *syllables*

ʔin'ta:	'my father'
'ʧi:bah	'meat'
'ʔe:q'um	'carrier'

[2] All examples are from McArthur and McArthur (1956), who list no cases of final CV syllables. Also, they claim that stress falls on the rightmost CVV syllable, yet no words with more than one CVV syllable are found in this source.

b. *Forms with no* CVV *syllables*
wu'qan 'my foot'
ʔal'k'om 'thief'
tpil'taʔ 'courthouse'

Quantity-sensitive stress systems are very rarely of type 3. In Gordon's (2006) survey, only four out of 86 languages are of type 3. Here we illustrate the distribution of stress in Kwakw'ala, in which CVV and CVR syllables pattern as heavy, while CV and CVO pattern as light (Boas 1947; also Zec 1988 and references therein). Stress falls on the leftmost heavy syllable, either CVV, in (10a), or CVR, in (10b). In words that contain only light syllables, CV or CVO, stress is final, as in (10c) and (d).

(10) *Type 3 language: Kwakw'ala*

a.	'qa:sa	'to walk'
	'n'a:la	'day'
	'ts'e:kwa	'bird'
	t'ə'li:dᶻu	'large board on which fish are cut'
b.	'm'ənsa	'to measure'
	'dəlxa	'damp'
	'dzəmbətəls	'to bury in hole in ground'
	mə'xənxənd	'to strike edge'
c.	nə'pa	'to throw a round thing'
	bə'ha	'to cut'
	m'əkʷə'la	'moon'
	ts'əxə'laə	'to be sick'
d.	ts'ət'xa	'to squirt'
	təʎ'ts'a	'to warm oneself'
	kʷ'əs'xa	'to splash'

The three weight patterns in (5) can be further exemplified with tonal phenomena, those provided by languages with lexical, i.e. contrastive tone, in simpler systems commonly High, or High and Low (CHAPTER 45: THE REPRESENTATION OF TONE). Quantity-sensitive tonal phenomena differ substantially from quantity-sensitive stress.[3] Crucial evidence for quantity-sensitivity comes from the so-called contour tones. If no more than one tone is sponsored by a light syllable and no more than two by a heavy one, we can say that multiple tones, standardly referred to as contour tones, may occur on heavy, but not on light, syllables. In other words, we focus on those languages in which a light syllable has one tone-bearing unit and a heavy syllable has two (see Zhang 2002 for different characterizations of

[3] It is not typical for tone to be attracted to a heavy syllable, although some cases have been interpreted in this light. Thus Hopi, as described in Jeanne (1982), has been interpreted as a quantity-sensitive stress system (Hayes 1995): stress occurs on initial heavy syllables, either CVV or CVC, otherwise on non-final peninitial syllables; stress is initial in all disyllables. However, because stress is realized as tonal prominence, this system has also been interpreted as a tonal system in which High tone is attracted to the initial heavy syllable, otherwise to the second syllable, if non-final (Yip 2002: 245). This stress-like behavior of tone, if indeed correctly interpreted, is truly atypical.

contour tones).[4] We further focus on those languages in which the mapping between tones and tone-bearing units is fairly straightforward: a tone-bearing unit may be associated with at most one tone. With this background, we turn to the evidence for the three weight patterns coming from the tonal domain.

We again rely on Gordon's (2006: 32–33) survey: out of 408 languages in his survey, 111 use contrastive tone and, of those, 61 use tone in a quantity-sensitive mode. Type 2 and type 3 weight patterns are widely exploited by weight-sensitive tonal phenomena, while type 1 is rarely associated with quantity-sensitive tone. Type 2 pattern is found in 28 languages (four without CVC syllables), and type 3 is found in 30 languages. In type 2 languages, contour tones occur on CVV syllables, but are absent from both CV and CVC syllables. In Navajo, contour tones occur only on CVV syllables as in (11a), while simple tones occur on all syllable types; (11b) exemplifies the absence of contour tones on CV and CVC syllables (Zhang 2002, based on Young and Morgan 1987).

(11) *Type 2: Navajo contour tone*

a.	sáànìì	'old woman'
	hákòónèèʔ	'let's go'
	tèíIʔá	'they extend'
b.	hááʔált'èʔ	'exhumation'
	pìkʰìn	'his house'

Another type 2 language is Ju|'hoansi, in which, as reported in Miller-Ockhuizen (1998; also Zhang 2002), contour tones are found only on long vowels and diphthongs, but not on CV syllables or syllables closed with nasals (the only type of closed syllable in the language).

Type 3 weight pattern is exemplified by a number of languages, including Nama (Khoisan), Lithuanian (Indo-European), and Tiv (Niger-Congo). Lithuanian has a pitch accent system, in which contour tones appear on heavy, but not on light, syllables. In particular, a Low High tonal contour, the so called circumflex accent, occurs on heavy syllables: CVV, as in /víinas/ 'wine', /zúikas/ 'rabbit', and CVR, as in /gársas/ 'sound', /bálsas/ 'voice', and /lánkas/ 'rainbow'. Syllables that pattern as light are CV and CVC, and those that pattern as heavy are CVV and CVR (Zec 1995 and references therein).

We now turn to the tonal evidence for the type 1 weight pattern. Contour tones are rare, and phonetically difficult to realize, on syllables closed with an obstruent. In his broad survey of quantity-sensitive tone, Gordon (2006) documents only three such cases: Hausa, Luganda, and Musey. Zhang (2002: 51) also lists Ngizim, and Yip (2002: 141–142) mentions the Nilo-Saharan language Kunama (Eritrea). Here we present evidence from Hausa, based on Gordon's (2006) experimental data. Hausa has three tones, two level tones, High and Low, and a contour High Low tone. As shown in (12), on the targeted initial syllables, the two level tones occur on all syllable types, while the contour tone occurs on CVV,

[4] Note, however, that tone languages vary as to what constitutes a tone-bearing unit. What we described here is one of several modes of selecting a tone-bearing unit. On tone and tone-bearing units, see CHAPTER 45: THE REPRESENTATION OF TONE.

CVR and CVO, but not on CV syllables. That is, the contour tone occurs on heavy, but not on light syllables.

(12) *Type 1: Hausa contour tone*

	L	H	HL
CV	fàsá:	sáfú:	–
CVV	mà:má:	rá:ná:	lâ:lá:
CVR	ràndá:	mándá:	mântá:
CVO	fàskí:	máskó:	râssá:

Of interest here is the fact that while sonorants, both vowels and consonants, are capable of phonetically realizing pitch, obstruents are not. As shown by Gordon (2006: 92), although the phonological weight of the CVO syllable provides the two tone-bearing units required for the realization of contour tone, the contour is phonetically realized on the vowel, which in this case has greater duration. No comparable increase in duration is evidenced in CVV and CVR syllables with contour tones. Thus, Hausa presents an interesting case of a mismatch between phonology and phonetics.

Other phonological phenomena that provide evidence for quantity-sensitivity include vowel shortening in closed syllables, to be addressed in §6, as well as compensatory lengthening (see CHAPTER 64: COMPENSATORY LENGTHENING) and poetic meter, which in §2.1 served as evidence for Latin. Onsets may on occasion exhibit quantity-sensitivity; for such cases, see CHAPTER 55: ONSETS and CHAPTER 47: INITIAL GEMINATES.

3 Representation of syllable quantity

The relevance of quantity-sensitivity, as well as its representation, was clearly recognized in early theoretical approaches to phonology. Both Jakobson (1931) and Trubetzkoy (1939) document weight distinctions among syllables, and cast them in terms of the unit of weight traditionally referred to as the mora: a light syllable contains one mora, and a heavy syllable contains two moras. Quantity-sensitivity was also recognized by Kuryłowicz (1948), who pursued the characterization of quantity in configurational terms, that is, in terms of a subconstituent of the syllable, the rhyme, whose structure is branching for heavy, and non-branching for light syllables. These two theoretical approaches to quantity-sensitivity, one in terms of constituency and the other in terms of arboreal configuration, emerged again in the 1970s and 80s, as competing representations of syllable weight, as well as the weight of other constituents in the prosodic hierarchy.

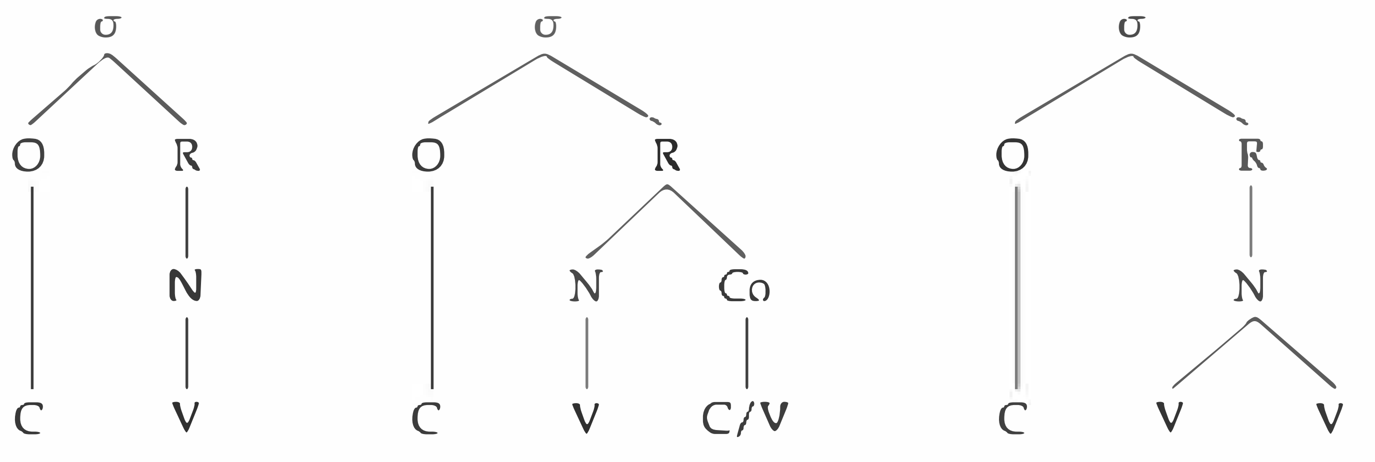
These two approaches both express an important intuition: that quantity formally corresponds to a binary structure. This will emerge as highly relevant in the representation of the syllable and its internal structure. This is also relevant for the representation of feet, as will be shown in §7.

The questions to be addressed in this section are: (i) how is weight computed from the representation of the syllable?; and (ii) how are different weight patterns represented? (For a general discussion of syllable structure and its representation, see CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE.)

3.1 Quantity represented in configurational terms

We begin with the configurational approach to syllable weight. In the representation in (13), the syllable branches into an onset and a rhyme, with the latter obligatorily dominating the nucleus and, optionally, the coda. The sub-syllabic constituent which is taken to be the domain of weight is the rhyme: if the rhyme branches, the syllable is heavy (13b); otherwise it is light (13a). An alternative assumption has been that a branching nucleus, as in (13c), has its role in the computation of quantity.

(13) a. *Light* b. *Heavy: Branching rhyme* c. *Heavy: Branching nucleus*



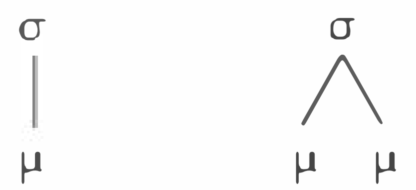
While this constituency was motivated on other grounds as well, capturing syllable quantity has been one of its important rationales. It was generally assumed that encoding weight distinctions is a crucial role of syllable structure. This representation was advocated, in this or somewhat modified form, by Kiparsky (1979, 1981), McCarthy (1979), Halle and Vergnaud (1980), Hayes (1980), Steriade (1982, 1988), and Levin (1985), among others. In all these approaches, the weight domain, provided by the rhyme subconstituent, crucially excludes the onset consonants, which do not participate in any of the weight patterns characterized in §2.2.

How does this representation capture the three weight patterns presented in (5)? In some proposals that primarily focus on type 2 languages (e.g. Halle and Vergnaud 1980), both CVV and CVC syllables are represented in terms of a branching rhyme, that is, as (13b). Capturing both type 1 and type 2 languages called for modifications. In one modification, CVV syllables are represented in terms of a branching nucleus, as in (13c), and CVC syllables in terms of a branching rhyme, as in (13b) (e.g. Hayes 1980). In another modification, different configurations are posited for type 1 and type 2 languages (e.g. McCarthy 1979). Type 3 language posed a special challenge: heavy CVR syllables in this language type were represented in terms of a branching nucleus (13c), with the weight-bearing sonorants residing in the nucleus together with vowels (Steriade 1990).

3.2 Quantity represented by constituency

Another way of capturing syllable weight is in terms of constituency. By positing the mora as a sub-syllabic constituent, syllable weight is represented in terms of the number of moras that the syllable dominates. A syllable with one mora is light, and a syllable with more than one mora is heavy.

(14) a. *Light* b. *Heavy*



While the mora as a unit of syllable weight goes back at least as far as the study of classical languages, it was introduced to theoretical phonology by Jakobson (1931) and Trubetzkoy (1939). Arguments for representing the mora as a sub-syllabic entity are primarily due to Hyman (1985), McCarthy and Prince (1986), Hayes (1989), and Zec (1988). Crucially, moras do not uniquely map to the level of segments. Moraic representations in (14) are thus sufficiently flexible to capture all three systems of syllable quantity. What needs to be stated is the set of segments that can be dominated by the second mora of the syllable: all segments, as in the type 1 weight pattern, only vowels, as in type 2, and vowels and sonorant consonants, as in type 3. How this is to be implemented varies with specific phonological models, which may rely either on rules or on constraints. Thus Hayes (1989) posits a weight-by-position rule, Zec (1988, 1995) posits language-specific sets of moraic segments that act as constraints on the second mora of a syllable, and Morén (1999) proposes optimality-theoretic constraints on moraic segments that parallel Prince and Smolensky's (1993) constraints on syllable nuclei.

4 Quantity-sensitivity and vowel length

Quantity-sensitivity is not a necessary property of the syllable. A number of languages, some listed in Hayes (1995), do not exhibit quantitative distinctions at the level of the syllable, for example, Bulgarian (Indo-European), Piro (Arawakan), Garawa (Karawic), and Modern Greek (Indo-European). Significantly, all these languages also lack vowel length (CHAPTER 20: THE REPRESENTATION OF VOWEL LENGTH). This strongly suggests that the basic weight contrast is in fact that between short and long vowels, and raises the question of possible implicational relations between syllable weight and vowel length, either phonemic or non-phonemic.

A strong claim on the relation of CVV and CVC syllables, proposed by Kuryłowicz (1948) and Newman (1972), among others, is that a language with heavy CVC syllables also has phonemic vowel length. While true in a number of specific cases, including Latin, Classical Arabic, and Fijian, this claim is too strong. A weaker claim is that the CVV syllable type is available in languages with heavy CVC syllables even if a language does not have phonemic vowel length (cf. Hayes 1989; Zec 1988, 1995). In such languages, vowel length could arise due to phonological processes such as compensatory lengthening, as in Ilokano (Hayes 1989), or iambic lengthening, as in Hixkaryana (Hayes 1995: 205 and the references therein). This claim rests crucially on a representation already available in a language (see §3), rather than on its phonemic distinctions.

5 Are weight distinctions binary or multivalued?

Cases of quantity-sensitivity presented thus far are characterized by two degrees of weight: a syllable is either light or heavy. The representations of syllable weight in §3 characterize quantity-sensitivity as a binary opposition, with two degrees of weight. However, a further question to be explored is whether there are cases of more than two degrees of weight, that is, whether quantity distinctions can be construed as scalar in nature.

Weight patterns with weight-bearing consonants, types 1 and 3, present an obvious point of departure. In a language with light CV and heavy CVV and CVC syllables, what is the status of CVVC and CVCC syllables? Are such syllable shapes allowed? And, if allowed, are they superheavy? That is, do they call for syllable structures that are either ternary branching or trimoraic? Likewise, what is the status of CVVR (and the less likely CVRR) syllables in type 3 languages?

Starting with type 1 languages, we find the following two cases. First, a language may have a syllable inventory that includes CVVC and CVCC syllables. In Hindi, such syllables give rise to three degrees of weight, as in (15a). Evidence for this ternary weight pattern comes from quantity-sensitivity in the stress system. Stress falls on a superheavy syllable if there is one, otherwise on a heavy syllable, otherwise on a light syllable (glossing over the complexities of this system, for details and examples, see §8). By contrast, Latin also has CVVC and CVCC syllables in addition to the standard type 1 inventory, yet exhibits only two degrees of weight, as in (15b). In this case, CVVC and CVCC syllables are functionally non-distinct from heavy syllables, CVV and CVC. This functional identity is supported by both stress and poetic meter.

(15) a. *Hindi*

light	CV
heavy	CVV,CVC
superheavy	CVVC, CVCC

b. *Latin*

light	CV
heavy	CVV, CVC, CVVC,CVCC

Newman (1972) claims that all weight distinctions are binary, pointing to languages like Latin. However, languages like Hindi clearly show that ternary weight distinctions are an attested reality.

Second, a language may have a restricted syllable inventory, with only CV, CVV, and CVC, excluding both CVVC and CVCC syllable shapes. Such languages impose binarity as an upper limit to syllable complexity both in terms of weight, or mora count, and in terms of the number of consonants that may occur at the right margin of the syllable. This situation is clearly illustrated by Turkish (Clements and Keyser 1983). The syllable inventory of Turkish, a type 1 language, includes light CV and heavy CVV and CVC syllables, and systematically lacks CVVC and CVCC syllables. If the prohibited syllable types arise by virtue of morpheme concatenation, they are eliminated by phonological processes. In (16a), the underlying long vowel is shortened in a closed syllable, (nominative and ablative), but not in the open syllable (accusative). And in (16b), the two post-vocalic consonants in the underlying form are split by an epenthetic vowel (CHAPTER 67: VOWEL EPENTHESIS), in order to avoid a CVCC syllable (nominative and ablative).

(16) *Turkish*

			accusative	*nominative*	*ablative*
a.	CVVC → CVC				
	/zamaːn-/	'time'	zamaːnɯ	zaman	zamandan
	/ispaːt-/	'proof'	ispaːtɯ	ispat	ispattan
b.	CVCC → CVCVC				
	/karn-/	'abdomen'	karnɯ	karɯn	karɯndan

Type 3 languages, or at least the known cases, do not provide evidence for ternary weight distinctions. Lithuanian, for example, has the following syllable shapes in its inventory, classified in terms of weight:

(17) *Lithuanian*

Light	CV, CVO, CVOO
Heavy	CVV, CVR, CVVO, CVRO

This weight pattern, as we saw in §2, is supported by the system of Lithuanian pitch accents (CHAPTER 42: PITCH ACCENT SYSTEMS): only heavy syllables, that is, CVV and CVR, can have contour tones. Lithuanian also provides evidence for strict binarity. This is evidenced by the process known as ablaut which applies in verbal morphology, with the effect of lengthening the root vowels in the preterite and infinitive, but not in the present form (Zec 1995). Vowel lengthening due to ablaut takes effect in all preterite forms: the root vowel occurs in an open syllable, due to the vowel-initial ending *-ee*, and is free to lengthen. In the infinitive forms, the root vowel is in a closed syllable, due to the consonant-initial ending *-ti*. Lengthening takes place in (18a), i.e. in roots that end in an obstruent, but not in roots that end in a sonorant (18b).

(18) *Lithuanian: Ablaut in verbal forms*

		root	*present*	*preterite*	*infinitive*	
a.	CVO	tup-	tupia	tuupee	tuupti	'perch'
		dreb-	drebia	dreebee	dreebti	'splash'
b.	CVR	vir-	viria	viiree	virti	'boil'
		mir-	miria	miiree	mirti	'die'

That is, ablaut may not create a superheavy CVVR syllable, and is therefore prevented from taking effect in the infinitives of the roots in (18b).

While type 2 languages may tolerate CVCC and CVVC syllables in their syllable inventories, such syllables do not form a natural class: the former has the weight of CV, and the latter has the weight of CVV syllables.

The extended syllable inventories we document in this section call for representations richer than those discussed in §3. This was directly addressed in moraic representations of the syllable and its weight: a constraint restricts the number of moras per syllable to no more than two; and this constraint can be violated in some languages, giving rise to trimoraic syllables, as in Hindi. The syllable inventory in Latin is accommodated by allowing some non-moraic consonants at the syllable's right margin (for a detailed discussion, see Sherer 1994).

6 Inconsistencies in weight patterns

The representations in §3, despite some conceptual differences, make the strong prediction that quantity distinctions in a language will be of the same type across the board, that is, in all relevant phonological processes, and in all contexts. However, a challenge to this strong position comes from many known cases of weight inconsistencies.

6.1 *Weight inconsistencies with respect to phonological process*

In §2.1 we saw that Latin belongs to the type 1 weight pattern both in its stress system and in the system of poetic meter. The phenomenon of compensatory lengthening (CHAPTER 64: COMPENSATORY LENGTHENING) conforms to this same weight pattern, as in /kasnus/ → [kaːnus]. While not uncommon, weight consistency across different phonological processes, as evidenced in Latin, is not the general case. Weight inconsistencies are encountered in a number of languages, as noted by Steriade (1990), as well as Hayes (1995) and Gordon (2006). One such case is Kiowa (Watkins 1984). As shown in (19), vowels are shortened in syllables closed by sonorants (19a), as well as those closed by obstruents (19b) and (19c), suggesting a type 1 weight system that obeys strict binarity.

(19) *Kiowa short vowels in closed syllables: Type 1*

a.	gúːlêː	'write-IMPERF-FUT'
	gûl	'write-IMP'
	gúltɔ̀ː	'write-FUT'
b.	cáːdɔ̀ː	'from the doorway'
	cát	'entrance, doorway'
	cátpé	'against the doorway'
c.	tʰɔ́ː	'beyond'
	tʰɔ́ːdèkʰìː	'next day'
	tʰɔ́p	'away beyond'

However, the distribution of contour tones, shown in (20), clearly points to a weight system of type 3. Contour tones occur on CVV syllables and syllables closed by a sonorant, as in (20a), but not on either CV syllables or syllables closed with an obstruent, as exemplified in (20b).

(20) *Kiowa contour tones: Type 3*

a.	páːlêː	'weak'
	sân	'child'
	kʰûl	'pull off'
b.	sàːné	'snake'
	sép	'rain'

Another case is Lhasa Tibetan (Gordon 2006, based on Dawson 1980), in which the stress system treats only CVV syllables as heavy, while the system of tone treats as heavy both CVV and CVR syllables (Gordon 2006 and references therein). In other words, Lhasa Tibetan is a type 2 language in its stress system, and a type 3 language in its tonal system. According to Steriade (1990), variability in weight is also found in Classical Greek, in which CVV syllables are heavy for the purposes of tone, yet all syllables are heavy for the purposes of stress distribution. Thus, stress falls on the penultimate syllable if the final syllable is heavy, either CVC(C) or CVV(C), otherwise it is antepenultimate. However, only CVV syllables can sustain tonal contours, either HL or LH.

Cases of weight variability in different phonological subsystems within a single language present an important challenge to formal representations, and call for fresh perspectives on the syllable and its quantity.

6.2 *Weight inconsistencies with respect to phonological context*

It has been noted in much work on stress that the weight of a syllable may be computed differently in word-internal and word-final positions. Thus in Classical Arabic, as shown in §2, stress falls on the rightmost CVV or CVC syllable, yet never on the final CVC. That is, CVC syllables are computed as heavy word-internally and as light word-finally. A further fact, not mentioned in §2, is that final CVCC syllables are always stressed, i.e. they are computed as heavy (CVCC do not occur word-internally). In other words, word-final consonants do not contribute to weight. Such cases of variable weight were subsumed in Hayes (1980) under the more general rubric of extrametricality (CHAPTER 43: EXTRAMETRICALITY AND NON-FINALITY), according to which certain phonological entities, segments as well as higher constituents, are "invisible" to phonological processes at word edges. There have been proposals, however, to treat contextual differences in weight as representational differences (Davis 1987; Kager 1989; Rice 1995; Rosenthall and van der Hulst 1999; see also CHAPTER 36: FINAL CONSONANTS). Under this view, the CVC sequence in Classical Arabic would be parsed as a heavy syllable word-internally, and as a light syllable word-finally.

It has been shown, however, that contextual weight differences are not restricted to word edges. Several cases of this type have been reported in Hayes (1994, 1995), among them Cahuilla and Eastern Ojibwa, as well as Central Alaskan and Pacific Yupik. In the Pacific Yupik dialect of Chugach, CVV syllables are heavy in all positions, while CVC syllables are heavy only initially, and light elsewhere. The distribution of stress in Chugach is fairly complex, and there can be more than one stress per word (for details, see Leer 1985; Kager 1993; Hayes 1995). We focus here on the evidence for the variable weight of CVC syllables. While initial CVV and CVC syllables are stressed, as in /'taːta'qa/ 'my father' and /'anciku'kut/ 'we'll go out', initial CV syllables are not, as in /mu'lu'kuːt/ 'if you take a long time'. But in medial position, CVC syllables pattern with CV rather than CVV syllables. Note that the second syllable in /'kal'maːnuq/ 'pocket', a CVV syllable, is stressed. Neither CV nor CVC syllables are stressed in this same environment, as in the forms /'anku'taχtu'a/ 'I'm going to go out' and /'atmaχ'ʧiqu'a/ 'I will backpack'.

Another relevant case is Goroa (Hayes 1980; Rosenthall and van der Hulst 1999, and references therein), in which stress falls on the leftmost CVV syllable, as in

(21a), or on the final CVC syllable, as in (21b); or on the penultimate syllable, as in (21c). Crucially, CVC syllables in positions other than final are not heavy: the second CVC syllable in /giram'bo:da/ does not win over the following CVV syllable, nor do the CVC syllables in /axe'mis/ and /idir'dana/ attract stress.

(21) *Gorоa stress: Variable weight of CVC syllables*

a. *Leftmost* CVV *stressed*

du:gnuno:	'thumb'
giram'bo:da	'snuff'
heni'nau	'young'

b. *Final* CVC *stressed*

a'dux	'heavy'
axe'mis	'hear'

c. *Penultimate syllable stressed*

oro'mila	'because'
am'rami	'ivory arm ring'
idir'dana	'sweet'

Contextually conditioned variation in syllable quantity affects CVC syllables, those that cross-linguistically could be either light or heavy. Thus the variability of the weight of CVC syllables found across languages has also been evidenced within individual languages. The phenomenon of contextually conditioned weight inconsistency of CVC syllables has been addressed, with a fair amount of success, in the Optimality Theory framework, most notably in Rosenthall and van der Hulst (1999).

7 Quantity-sensitivity of the foot

Syllables are grouped into feet, which belong to the next higher level of the prosodic constituency in (1) (see CHAPTER 40: THE FOOT; CHAPTER 41: THE REPRESENTATION OF WORD STRESS). Quantity-sensitivity of the syllable is directly reflected at the level of the foot, as noted in Hayes (1980, 1995), McCarthy and Prince (1986), and Prince (1990), among others. Feet play an important role in the characterization of stress and in prosodic morphology, and our examples will come from both domains.

As shown in a vast body of literature, feet tend to be binary. That is, feet are prosodic constituents resulting from grouping at most two constituents at the next lower level (Hayes 1995; McCarthy and Prince 1986; Prince 1990; among others). How this proceeds depends crucially on whether a language has a quantity-sensitive or a quantity-insensitive foot system (Hayes 1980). In quantity-insensitive systems, pairs of syllables are incorporated into feet regardless of their weight. Relevant for our discussion is foot formation in quantity-sensitive systems, in which syllable weight plays a crucial role. An important property of such systems is the commensurability of a heavy syllable with two lights. There are two types of quantity-sensitive feet, trochaic and iambic (CHAPTER 44: THE IAMBIC–TROCHAIC LAW).

In quantity-sensitive trochaic systems a foot corresponds to either one heavy syllable, as in (22a), or two light syllables, as in (22b); feet are left-headed, that is, have initial prominence, shown in (22b) by underlining.

(22) *Trochaic foot inventory*

a. σ_H
b. $\underline{\sigma_L}\ \sigma_L$

This receives a straightforward interpretation in moraic theory of syllable structure: a foot contains two moras, a condition met either by one heavy syllable, as in (22a), or by two lights, as in (22b). A heavy syllable has a dual status: it counts not only as a syllable but also as a foot. This foot inventory is active in the stress system of Fijian, a type 2 language (Hayes 1995, and references therein). In words with only a light syllable, pairs of syllables are incorporated into feet, computing from right to left, and foot-initial syllables are assigned stress. As a result, stress falls on every second syllable, computed from the right edge, as shown in (23). Parsing of syllables into feet obeys strict binarity, but is not necessarily exhaustive. In words with an odd number of syllables, as in (23c) and (23e), a syllable at the left edge is not footed. (The rightmost stressed syllable bears primary stress; others bear secondary stress.)

(23) *Fijian stress: Light syllables only*

a.	('lako)	'go'
b.	('talo)	'pour'
c.	βi ('naka)	'good'
d.	(ˌⁿdiko) ('nesi)	'deaconess'
e.	pe (ˌresi) ('teⁿdi)	'president'

(24) *Fijian stress: Light and heavy syllables*

a.	ki ('la:)	'know'
e.	(ˌᵐbe:) ('leti)	'belt'
c.	(ˌᵐbele) (ˌᵐbo:) ('tomu)	'bellbottoms'
d.	pa(ˌro:) ka ('ramu)	'program'
e.	(ˌmi:) (ˌsini) ('ᵑgani)	'machine-gun'

In words with both light and heavy syllables, each heavy syllable corresponds to a foot, and is stressed. Right-to-left footing is thus disrupted by heavy syllables, and has to work around them. In the disyllabic form with a heavy final syllable, in (24a), the initial syllable is left unfooted. And the form in (24d), which has five syllables, two light syllables, the first and the third, are left unfooted. All syllables are footed in the remaining forms in (24).

The inventory of feet in (22) captures the distribution of stress in a number of trochaic quantity-sensitive systems, including some of the cases presented in §2. In particular, stress in Latin follows the same pattern as in Fijian, with one notable difference: The final syllable is ignored for the purposes of stress (another case of so-called extrametricality, see §6.2). As a result, trisyllabic forms with only light syllables have initial stress, as in ('*ani*)*ma* 'soul (NOM SG)'. Likewise, final heavy syllables are not stressed: in ('*gau*)*dēns* 'rejoicing (NOM SG)' the penultimate heavy, but not the final heavy, is footed, and stressed (for a detailed analysis, see Mester 1994; Hayes 1995).

We also present a case of prosodic morphology that employs the foot inventory in (22). In the system of Japanese hypocoristic formation, as characterized in

Poser's (1990) detailed study, hypocoristics are formed by adding the suffix *-tjan* to proper names, either to their full or modified form. As shown by Poser, what is considered to be modification is really a case of template satisfaction. Crucially, the template corresponds to a trochaic foot: either to two light syllables or one heavy syllable. Japanese, a type 1 language, has light CV and heavy CVV and CVC syllables. As shown by the truncated versions of the proper name *Hanako*, the suffix is added to two light syllables, as in (25a) or one heavy, as in (25b) and (25c). The truncated form cannot be smaller than a foot, corresponding to a single light syllable, as in the ill-formed (25d). Nor can the truncated form be greater than a foot, corresponding to three light syllables, as shown by the ill-formed (26b). Proper names corresponding to a light syllable are converted to a heavy syllable, that is, to a foot; in (27a) this is accomplished by vowel lengthening. Note that (27b) is also available, as *-tjan* can be added to any proper name in its full form regardless of its size.

(25) *Hypocoristic forms for* Hanako

a. hanatjan
b. haatjan
c. hattjan
d. *hatjan

(26) *Hypocoristic forms for* Takatugu

a. takatjan
b. *takatutjan

(27) *Hypocoristic forms for* Ti

a. tiitjan
b. titjan

Thus, in trochaic prosodic morphology, just as in trochaic stress systems, a heavy syllable is functionally equivalent with two light syllables.

Quantity-sensitive iambic feet differ somewhat in shape from the trochaic set, as shown by the inventory in (28). Iambic feet are right-headed, indicated by the underlining.

(28) *Iambic foot inventory*

a. σ_H
b. $\sigma_L\ \underline{\sigma_L}$
c. $\sigma_L\ \underline{\sigma_H}$

In this case, as well, syllable quantity plays a central role: for a foot to be well-formed, it needs to contain syllables of the correct weight. The iambic system of quantitative feet captures the distribution of stress in Asheninca (Hayes 1995; Payne 1990). Asheninca has a type 2 weight system, with only CVV heavy syllables. The forms in (29a) contain only light syllables: binary right-headed feet are computed from right to left. The final syllable is regularly left unfooted, which yields initial stress in disyllables, as in /'haka/ 'here'. Crucial are the forms in (29b), which contain both light and heavy syllables, and can therefore exemplify all members of the foot inventory.

(29) *Asheninca stress system*

a.	(pa.'me).(na.'ko).(wen.'ta).ke.ro	'take care of her'
	(ha.'ma).(nan.'ta).(ke.'ne).ro	'he bought it for her'
	(no.'ko).(wa.'we).ta.ka	'I wanted (it) in vain'
	(no.'ton).(ka.'men).to	'my gun'
	(ka.'man).ta.ke	'he/she said'
b.	(no.'ma).(ko.'rjaa).('wai).(ta.'paa).ke	'I rested a while'
	(pi.'ɲaa).('paa).ke	'you saw on arrival'
	(i.'kjaa).('piin).ti	'he always enters'
	('poo).(ka.'na).ke.ro	'you threw it out'
	('paa).(ti.'ka).ke.ri	'you stepped on him'

Quantity-sensitive iambic feet also figure in prosodic morphology. In Ulwa, which has a type 1 weight system, the suffix /-ka/ is attached to the leftmost iambic foot, as in (30). It occurs at the right edge of a stem only when the entire stem corresponds exactly to a foot, as in (30a). In (30b), the only way for /-ka/ to be attached to an iambic foot is to occur stem-internally.

(30) *Ulwa construct state* (from McCarthy and Prince 1990: 228)

	base	*possessed*	
a.	al	al-ka	'man'
	bas	bas-ka	'hair'
	kii	kii-ka	'stone'
	sana	sana-ka	'deer'
	amak	amak-ka	'bee'
	sapaa	sapaa-ka	'forehead'
b.	suulu	suu-ka-lu	'dog'
	kuhbil	kuh-ka-bil	'knife'
	baskarna	bas-ka-karna	'comb'
	siwanak	siwa-ka-nak	'root'
	anaalaaka	anaa-ka-laaka	'chin'
	karasmak	karas-ka-mak	'knee'

Trochaic and iambic systems differ with regard to the role of quantity, as noted in Hayes (1985) and Prince (1990) as well as CHAPTER 44: THE IAMBIC–TROCHAIC LAW. The preferred type of trochaic disyllabic feet includes two light syllables, while iambic feet optimally correspond to a sequence of a light and heavy syllable. Thus, disyllabic trochaic feet are preferably even, while iambic feet are preferably of uneven quantity. Evidence for even trochaic quantity comes from the so-called trochaic shortening, which makes an uneven trochee even by vowel shortening, as exemplified by Fijian. The form in (31a), with an underlying long vowel, undergoes shortening when integrated into a disyllabic foot, as in (31b).

(31) *Fijian: Trochaic shortening*

a.	'ta:	'chop'
b.	'ta-ja	'chop-TRANS-3SG OBJ'

By contrast, uneven quantity is an important feature of iambic systems. A number of iambic stress systems are characterized by iambic lengthening, including

Menomini, Hixkaryana, and Kashaya (Hayes 1995). In Hixkaryana (Caribian), prominent CV syllables undergo vowel lengthening, as in (32a) and (32b), and thus become heavy. Note that prominent CVC syllables, which are already heavy, are not subject to lengthening, as in the second foot in (32c), and the initial syllables in (32a) and (32b).

(32) *Hixkaryana: Iambic lengthening*

a.	owtohona	('ow)(to'ho:)na	'to the village'
b.	tohkurʲehonahaʃaka	('toh)(ku'rʲe:)(ho'na:) (ha'ʃa:)ka	'finally in Tohkurye'
c.	nɨhananɨhno	(mɨ'ha:)(na'nɨh)no	'you taught him'

Generalizations about the quantity of trochaic and iambic groupings are stated in Hayes (1995) as the Iambic–Trochaic Law (see CHAPTER 44: THE IAMBIC–TROCHAIC LAW):

(33) *The Iambic–Trochaic Law*

a. Elements contrasting in intensity naturally form groupings with initial prominence.
b. Elements contrasting in duration naturally form groupings with final prominence.

8 Scalar quantity systems

While binary quantity systems are based primarily on grouping syllables into feet, scalar quantity systems are based on prominence, defined along some dimension (Prince and Smolensky 1993; Hayes 1995). A central prominence dimension is syllable weight, although other dimensions, such as tone and vowel height, have been evidenced as well.

We present two cases with syllable weight as the prominence dimension. One is Kashmiri, with examples given in (34) (Kenstowicz 1993; Rosenthall and van der Hulst 1999). In Kashmiri, CVV syllables are heavier than CVC, which in turn are heavier than CV. Thus, in words with only CV and CVV syllables, stress falls on the leftmost CVV, as in (34a). In words with only CV and CVC, stress falls on the leftmost CVC, as in (34b). In words with both CVC and CVV syllables, stress falls on the CVV syllable, as in (34c). Finally, with only CV syllables present, stress is initial, as in (34d). The final syllable is excluded from scansion. (None of the sources supply glosses for Kashmiri forms.)

(34) *Kashmiri stress:* CVV > CVC > CV

a.	mu'si:ba[illegible]	c.	am'ri:ka
	a'jo:gja ta:		mas'ra:wun
b.	ba'gandarladin	d.	'tsaripop
	juni'varsiti		'paharadari:

Languages in which stress is assigned on the basis of scalar syllable prominence may have several degrees of syllable weight. Thus Hindi (for the dialect described in Kelkar 1968) has three degrees of syllable weight: superheavy syllables CVVC and CVCC are more prominent than heavy syllables, CVV and CVC, which in turn are more prominent than CV syllables, as stated in (35).

(35) CVVC, CVCC > CVV, CVC > CV

Excluding the final syllable from scansion, stress is assigned to the heaviest available syllable, as in (36). In both forms stress falls on a CVVC syllable, which in (36b) wins over a CVV syllable, and in (36a) over both a CV and CVV syllable.

(36) a. 'ʃoːxʤabaːniː 'talkative'
b. 'reːzgaːriː 'small change'

If there is a tie, stress is assigned to the rightmost (non-final) syllable: to a CV syllable in (37a), and a CVV syllable in (37b) and (37c).

(37) a. sa'miti 'committee'
b. roː'zaːnaː 'daily'
c. kaː'riːgariː 'craftsmanship'

Interestingly, when the final syllable is the heaviest in the word, it is not excluded from scansion, as in (38):

(38) ki'dʰar 'which way'
ru'pia 'rupee'
as'baːb 'goods'

Quantity in Hindi is thus computed along a scale of syllable weight, with the superheavy syllable being most prominent, followed by the heavy syllable and then by the least prominent light syllable. This case is analyzed in precisely these terms in Hayes (1995) and Prince and Smolensky (1993), although in different frameworks: in rule-based metrical theory and in Optimality Theory, respectively.

An interesting mode of computing prominence is found in Pirahã, a Mura language of Brazil (Everett 1988; Hayes 1995). The Pirahã prominence scale combines syllable weight and onset quality (on onsets, see CHAPTER 55: ONSETS). While CVV syllables are more prominent than CV syllables, voiceless onsets are more prominent than voiced onsets, and presence of onset is more prominent than its absence, yielding the scale in (39).

(39) KVV > GVV > VV > KV > GV [K = voiceless, G = voiced]

Stress falls on one of the last three syllables of the word that is highest on this scale, as in (40a). In the event of ties, the rightmost syllable wins, as in (40b).

(40) *Pirahã prominence-based stress*

a.	'kaːgai	b.	ko'po
	ʔapa'baːsi		ʔaba'pa
	'ʔibogi		paohoa'hai

Further dimensions for computing prominence are in no obvious way related to patterns of syllable quantity we surveyed here. Yet, because of their scalar nature, they are highly reminiscent of quantity-based systems of prominence. One such dimension is vowel quality: given the sonority scale, stress falls on the most sonorous vowel. Prominence systems of this type have been analyzed in Kenstowicz (1997) and de Lacy (2004). In Mordwin, for example, non-high vowels are more prominent than high vowels (CHAPTER 21: VOWEL HEIGHT). In words with only non-high vowels, or with only high vowels, stress falls on the leftmost syllable. However, in words that contain both high and non-high vowels, stress falls on a non-high vowel, regardless of its position in the word. Another dimension is tonal prominence: syllables associated with High tones are more prominent than syllables associated with Low tones, and thus more likely to be associated with stress. Prominence systems of this type are described in Hayes (1995) and de Lacy (2002); for a somewhat different perspective, see Zec (2003). Of particular interest is the complex case of Nanti, a Kampa language of Peru: its stress system, which is of the iambic type, is also governed by several types of prominence, including syllable quantity and vowel quality (Crowhurst and Michael 2005).

9 Quantity-sensitivity at the higher levels of the prosodic hierarchy

When focusing on higher levels of the prosodic hierarchy, the prosodic word and the prosodic phrase, we are in fact dealing with morphosyntax/prosody interface. Quantity-sensitivity is a specifically prosodic phenomenon and is not known to play any role in other modules of the grammar. Any effects of quantity-sensitivity in either morphology or syntax are therefore to be attributed to prosody. We addressed the interfaces with morphology in §8, with two cases of affixes that select not only a morphological class, but also a prosodic type of the stem; both in this case select for the foot. Many more such cases are found in the literature (McCarthy and Prince 1986, 1990, 1995; among others).

The word level of the prosodic hierarchy is constituted by a grouping of feet (CHAPTER 41: THE REPRESENTATION OF WORD STRESS; CHAPTER 51: THE PHONOLOGICAL WORD). In practice, however, one foot is sufficient for a prosodic word to achieve the desired quantity, as documented by numerous cases of minimal word size. Moreover, in a number of languages, no minimal size beyond a single syllable is imposed on prosodic words. This is broadly documented in Hayes (1995) and Downing (2006), among others. In sum, the prosodic word provides no evidence for quantity-sensitivity of the sort found at the level of the syllable and the foot: its binary structure is not distinct from that of a foot. There are no known cases of a prosodic word minimally branching into two feet, yet this would be expected, based on the situation at the lower end of the prosodic hierarchy.

However, quantity-sensitivity has been evidenced at the higher end of the hierarchy, that is, at the level of the prosodic phrase. The distribution of a syntactic constituent should not be affected by its length or internal complexity. When such effects arise, they are generally attributed to prosody. We focus here on cases of branching in prosodic phrases, typical cases of apparent quantity-sensitivity of syntactic constituents. Cases of binary branching prosodic phrases were reported by Nespor and Vogel (1986), with evidence from Italian, French, and English. In Italian, for example, a prosodic phrase preferably contains more than one prosodic word, as shown by the following cases (Nespor and Vogel 1986):

(41) *Prosodic phrase formation in Italian*

a. *Av'rà trovato (il pescecane)*$_\varphi$
'He will have found the shark.'

b. *(I cari'bù nanni)*$_\varphi$ *sono estinti*
'Dwarf caribous are extinct.'

c. *Hanno dei (car'ibù)*$_\varphi$ *(molto piccoli)*$_\varphi$
'They have very small caribous.'

While complements that correspond to single prosodic words, as in (41a), form one-word prosodic phrases, multiple word complements, as in (41b) and (41c), correspond to branching prosodic phrases. The prosodic phrasing in (41c) further shows that complements with three prosodic words do not correspond to a single prosodic phrase, as branching prosodic phrases contain at most two prosodic words. By contrast, Serbo-Croatian sentence-initial topics have to include at least two prosodic words (Zec and Inkelas 1990), and thus exemplify obligatory branching in prosodic phrases. This line of research has been further continued by Ghini (1993), Selkirk (2000), and Sandalo and Truckenbrodt (2002), among others.

10 Remarks on markedness

It is important to note that the markedness (CHAPTER 4: MARKEDNESS) of light and heavy constituents is not identical across prosodic levels: heavy constituents are marked at the level of the syllable, while light constituents are marked at the level of the foot. This is directly encoded in Optimality Theory. Constraints listed in (42) assign marked status to heavy syllables: to CVV syllables, as in (42a), and to syllables with coda consonants, as in (42b) and (42c). While (42b) targets any coda consonant, (42c) targets any weight-bearing segment.

(42) a. NoLongVowel (Rosenthall 1994)
A vowel should not be long, i.e. linked to more than one mora.

b. NoCoda (Prince and Smolensky 1993)
Syllables must not have a coda.

c. *Mora[seg] (Morén 1999)
Do not associate a mora with a particular type of segment.

These constraints, which belong to the markedness family, penalize binary structures at the syllable level, thus favoring a simple CV syllable, which is light. Thus, light syllables emerge as the unmarked case: all languages have light syllables, and some may also have heavy syllables. Superheavy, i.e. trimoraic, syllables are, of course, also marked, and are penalized as such by a constraint against trimoraic syllables proposed by Sherer (1994).

By contrast, heavy feet are preferred over light ones. Binary constituents are highly desirable at the level of the foot, both in trochaic and iambic systems. Non-binary, or light, feet are permitted in some languages under very special conditions and banned in others. The unmarked condition for feet is thus to be binary, that is, heavy, either under moraic or syllabic analysis, and this is codified in Optimality Theory by a corresponding constraint:

(43) FootBinarity (McCarthy and Prince 1993)

Feet must be binary under a syllabic or moraic analysis.

At the higher prosodic levels, constituent size is largely determined by morphosyntax, as is the distribution of light and heavy constituents. However, where permitted by morphosyntax, heavy, i.e. branching, constituents are preferred over light ones (see §9).

11 Conclusion

Quantity-sensitivity is an important property of prosodic structure, evidenced at each of its levels. As we have seen, constituents at any level of the prosodic hierarchy can be classified into those that are light and those that are heavy. While quantity-sensitivity is typically associated with the syllable and the foot, all prosodic levels exhibit this property. Whether a syllable is light or heavy crucially depends on its segmental setup; quantity at the level of the foot relies on, and is largely characterized in terms of syllable quantity; quantity-sensitivity of the prosodic word is non-distinct from that of the foot; and quantity-sensitivity at the higher prosodic levels is heavily influenced by morphosyntax.

While characterization of quantity largely depends on level-specific criteria, a general property of heavy constituents is their greater size and complexity, and often their binary structure. It is interesting, however, that preference, or dispreference, for heavy constituents varies across prosodic levels. The unmarked condition for syllables is to be light while the unmarked condition for feet is to be heavy. The latter condition persists through the higher levels of the prosodic hierarchy. Thus, while light syllables are preferred over heavy ones, feet and prosodic words are preferably heavy. Heavy prosodic phrases are preferred as well, although in a very weak sense.

ACKNOWLEDGMENTS

I am most grateful to Adam Cooper, Michael Weiss, and Kimberly Will, to an anonymous referee, and to the editors, Marc van Oostendorp and Keren Rice, for their invaluable comments and suggestions, which have led to major improvements to this chapter.

REFERENCES

Allen, W. Sidney. 1973. *Accent and rhythm.* Cambridge: Cambridge University Press.

Boas, Franz. 1947. Kwakiutl grammar with a glossary of the suffixes. *Transactions of the American Philosophical Society: New Series* 37. 201–377.

Buckley, Eugene. 1998. Alignment in Manam stress. *Linguistic Inquiry* 29. 475–496.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable.* Cambridge, MA: MIT Press.

Crowhurst, Megan J. & Lev Michael. 2005. Iterative footing and prominence-driven stress in Nanti (Kampa). *Language* 81. 47–95.

Davis, Stuart. 1987. Coda extrametricality in English nouns. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23. 66–75.

Dawson, Willa. 1980. Tibetan phonology. Ph.D. dissertation, University of Washington.

de Lacy, Paul. 2002. The interaction of tone and stress in Optimality Theory. *Phonology* 19. 1–32.

de Lacy, Paul. 2004. Markedness conflation in Optimality Theory. *Phonology* 21. 145–199.

Downing, Laura J. 2006. *Canonical forms in prosodic morphology.* Oxford: Oxford University Press.

Everett, Daniel L. 1988. On metrical constituent structure in Pirahã phonology. *Natural Language and Linguistic Theory* 6. 207–246.

Ghini, Mirco. 1993. X-formation in Italian: A new proposal. *Toronto Working Papers in Linguistics* 12. 41–78.

Gordon, Matthew. 2006. *Syllable weight: Phonetics, phonology, typology.* London: Routledge.

Halle, Morris & William J. Idsardi. 1995. General properties of stress and metrical structure. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 403–443. Cambridge, MA & Oxford: Blackwell.

Halle, Morris & Jean-Roger Vergnaud. 1980. Three-dimensional phonology. *Journal of Linguistic Research* 1. 83–105.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress.* Cambridge, MA: MIT Press.

Hayes, Bruce. 1980. A metrical theory of stress rules. Ph.D. dissertation, MIT.

Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13. 227–276.

Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 11. 429–446.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce. 1994. Weight of CVC can be determined by context. In Jennifer Cole & Charles W. Kisseberth (eds.) *Perspectives in phonology*, 61–79. Stanford: CSLI.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Hyman, Larry M. 1977. On the nature of linguistic stress. In Larry M. Hyman (ed.) *Studies in stress and accent*, 37–82. Los Angeles: Department of Linguistics, University of Southern California.

Hyman, Larry M. 1985. *A theory of phonological weight.* Dordrecht: Foris.

Jakobson, Roman. 1931. Die Betonung und ihre Rolle in der Wort- und Syntagmaphonologie. *Travaux du Cercle Linguistique de Prague* 2. Reprinted (1962) in *Selected writings*, vol. 1: *Phonological studies*, 117–136. The Hague: Mouton.

Jeanne, LaVerne Masayesva. 1982. Some phonological rules of Hopi. *International Journal of American Linguistics* 48. 245–270.

Kager, René. 1989. *A metrical theory of stress and destressing in English and Dutch.* Dordrecht: Foris.

Kager, René. 1993. Alternatives to the iambic-trochaic law. *Natural Language and Linguistic Theory* 11. 381–432.

Kelkar, Ashok R. 1968. *Studies in Hindi-Urdu I: Introduction and word phonology*. Poona: Deccan College.

Kenstowicz, Michael. 1993. Peak prominence stress systems and Optimality Theory. *Proceedings of the 1st International Conference of Linguistics and Chosun University*, 7–22. Gwangju: Foreign Culture Research Institute, Chosun University, Korea.

Kenstowicz, Michael. 1997. Quality-sensitive stress. *Rivista di Linguistica* 9. 157–187.

Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10. 421–441.

Kiparsky, Paul. 1981. Remarks on the metrical structure of the syllable. In Wolfgang U. Dressler, Oskar E. Pfeiffer & John R. Rennison (eds.) *Phonologica 1980*, 245–256. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Kuryłowicz, Jerzy. 1948. Contribution à la théorie de la syllabe. *Bulletin de la Société Polonaise de Linguistique* 8. 80–113.

Larsen, Raymond S. & Eunice V. Pike. 1949. Huasteco intonations and phonemes. *Language* 25. 268–277.

Leer, Jeff. 1985. Toward a metrical interpretation of Yupik prosody. In Michael Krauss (ed.) *Yupik Eskimo prosody systems: Descriptive and comparative studies*, 159–172. Fairbanks: Alaska Native Language Center Research Papers.

Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.

Lichtenberk, Frantisek. 1983. *A grammar of Manam*. Honolulu: University of Hawaii Press.

McArthur, Henry & Lucille McArthur. 1956. Aguacatec (Mayan) phonemes within the stress group. *International Journal of American Linguistics* 22. 72–76.

McCarthy, John J. 1979. On stress and syllabification. *Linguistic Inquiry* 10. 443–465.

McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.

McCarthy, John J. & Alan Prince. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory* 8. 209–283.

McCarthy, John J. & Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology 1993*. 79–153.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.

Mester, Armin. 1994. The quantitative trochee in Latin. *Natural Language and Linguistic Theory* 12. 1–61.

Miller-Ockhuizen, Amanda. 1998. Towards a unified decompositional analysis of Khoisan lexical tone. In Mathias Schladt (ed.) *Language, identity, and conceptualization among the Khoisan*, 217–243. Cologne: Rüdiger Köppe Verlag.

Morén, Bruce. 1999. Distinctiveness, coercion and sonority: A unified theory of weight. Ph.D. dissertation, University of Maryland at College Park.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Newman, Paul. 1972. Syllable weight as a phonological variable. *Studies in African Linguistics* 3. 301–323.

Payne, Judith. 1990. Asheninca stress patterns. In Doris L. Payne (ed.) *Amazonian linguistics: Studies in lowland South American languages*, 185–209. Austin: University of Texas Press.

Poppe, Nicholas N. 1960. *Buriat grammar*. Bloomington: Indiana University Publications.

Poser, William J. 1990. Evidence for foot structure in Japanese. *Language* 66. 78–105.

Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14. 19–100.

Prince, Alan. 1990. Quantitative consequences of rhythmic organization. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26(2). 355–398.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder.

Rice, Curt. 1995. An Optimality Theoretic analysis of weight variability and variation in primary stress in English, Dutch, and Norwegian. Unpublished ms., University of Tromsø.

Rosenthall, Sam. 1994. Vowel/glide alternation in a theory of constraint interaction. Ph.D. dissertation, University of Massachusetts, Amherst.

Rosenthall, Sam & Harry van der Hulst. 1999. Weight-by-position by position. *Natural Language and Linguistic Theory* 17. 499–540.

Ryding, Karin C. 2005. *A reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.

Sandalo, Filomena & Hubert Truckenbrodt. 2002. Some notes on phonological phrasing in Brazilian Portuguese. *Della* 19. 1–30.

Selkirk, Elisabeth. 1978. On prosodic structure and its relation to syntactic structure. In Thorstein Fretheim (ed.) *Nordic prosody II*, 111–140. Trondheim: Tapir.

Selkirk, Elisabeth. 1980. Prosodic domains in phonology: Sanskrit revisited. In Mark Aronoff & Mary-Louise Kean (eds.) *Juncture*, 107–129. Saratoga: Anma Libri.

Selkirk, Elisabeth. 2000. The interaction of constraints on prosodic phrasing. In Merle Horne (ed.) *Prosody: Theory and experiment*, 231–261. Dordrecht: Kluwer.

Sherer, Timothy. 1994. Prosodic phonotactics. Ph.D. dissertation, University of Massachusetts, Amherst.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT.

Steriade, Donca. 1988. Review of Clements & Keyser (1983). *Language* 64. 118–129.

Steriade, Donca. 1990. Moras and other slots. *Proceedings of the Formal Linguistics Society of Midamerica* 1. 254–280.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht.

Walker, Rachel. 1996. A third parameter for unbounded stress. *Papers from the Annual Meeting of the North East Linguistic Society* 26. 441–455.

Watkins, Laurel J. 1984. *A grammar of Kiowa*. Lincoln & London: University of Nebraska Press.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Young, Robert & William Morgan. 1987. *The Navajo language: A grammar and dictionary*. Albuquerque: University of New Mexico Press.

Zec, Draga. 1988. Sonority constraints on prosodic structure. Ph.D. dissertation, Stanford University.

Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12. 85–129.

Zec, Draga. 2003. Prosodic weight. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*, 123–143. Cambridge: Cambridge University Press.

Zec, Draga & Sharon Inkelas. 1990. Prosodically constrained syntax. In Sharon Inkelas & Draga Zec (eds.) *The phonology–syntax connection*, 365–378. Chicago: University of Chicago Press.

Zhang, Jie. 2002. *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis*. New York & London: Routledge.

Volume I: General Issues and Segmental Phonology
Volume II: Suprasegmental and Prosodic Phonology
Volume III: Phonological Processes
Volume IV: Phonological Interfaces
Volume V: Phonology across Languages

The Blackwell Companion to Phonology is a major reference work drawing together 124 new contributions from leading scholars in the field. Led by a team of renowned international scholars, the *Companion* represents a diverse range of approaches and methodologies to key phenomena in phonological research. In contrast to other handbooks and reference works currently available for phonology, these volumes focus on phenomena and case studies that highlight past and ongoing debates in the field. Interdisciplinary connections, such as those with the social and computational sciences, are covered, as well as statistical and experimental techniques. The *Companion* will be a touchstone for future phonological theorists, giving an overview of data and insights which any theory of phonology should be able to cover.

Simultaneously offering broad coverage and a high level of detail, *The Blackwell Companion to Phonology* is a landmark work available in both print and online editions that will be indispensable to students and researchers in the field for years to come.

www.companiontophonology.com

Cover image: Theo van Doesburg, Color scheme for Café Aubette, Strasbourg, France 1926. New York, Museum of Modern Art (MoMA). Ink and gouache on paper, 27.3 x 62.9 cm. Gift of Lily Auchincloss, Celeste Bartos, and Marshall Cogan 390.1982 © 2010. Digital image, The Museum of Modern Art, New York / Scala, Florence.

Cover design by Nicki Averill

WILEY-BLACKWELL
www.wiley.com/wiley-blackwell

VOLUME III
Phonological Processes

Edited by
van Oostendorp, Ewen, Hume, and Rice

THE BLACKWELL COMPANION TO
Phonology

WILEY-BLACKWELL

VOLUME III
Phonological Processes

Edited by
Marc van Oostendorp
Colin J. Ewen
Elizabeth Hume
Keren Rice

THE BLACKWELL COMPANION TO Phonology

WILEY-BLACKWELL

86 Morpheme Structure Constraints

Geert Booij

1 Introduction

Morpheme structure constraints are constraints on the segmental make-up of the morphemes of a language. A textbook example of such a constraint is that *bnik* is an impossible morpheme of English, whereas *blik* is a possible English morpheme that happens not to exist. Hence, *bnik* is a systematic gap in the morpheme inventory of English, whereas *blik* is an accidental gap in this inventory. This can be taken to imply that there is a morpheme structure constraint that prevents English morphemes from beginning with a /b/ followed by a nasal consonant.

Halle (1959: 38) proposed to account for such distributional generalizations by means of morpheme structure rules, which define the class of possible morphemes of a language. Morpheme structure rules were conceived of as rules that fill in predictable specifications of the sound segments of a morpheme. For instance, in the case of English morphemes that begin with the consonant cluster *b*C, such as *brick*, it is predictable that the C must be a liquid, i.e. a non-nasal sonorant consonant. That is, the feature specifications [−nasal] and [+sonorant] of the second consonant of *brick* are predictable. They can therefore be omitted in the lexical phonological specification of the relevant morphemes. Morpheme structure rules fill in the blank cells of the lexical phonological matrix, and thus turn this underspecified matrix into a systematic phonological matrix, with all feature values of its segments specified. This is the underlying phonological form of a morpheme to which the phonological rules of a language apply. In sum, morpheme structure rules function as redundancy rules that specify predictable information, and at the same time they define the set of possible morphemes of a language.

Stanley (1967) proposed to replace Halle's notion "morpheme structure rule" by the notion "morpheme structure condition" (MSC). All morpheme structure conditions function as redundancy statements with respect to fully specified lexical phonological matrices, which form the input for the phonological rules (P-rules).

The notion "morpheme structure condition" as discussed above forms part of the theoretical machinery of classical generative phonology, but has been subject to debate. In this chapter this debate will be summarized. Before doing that, I will provide some examples of phonotactic properties of morphemes in §2. The problems raised by the concept of "morpheme structure condition" will then be discussed in §3–§5. These problems are the following:

(i) The redundancy problem: is there any need for a specific set of morpheme structure conditions, or can they be made to follow from other types of phonological rules or constraints? (§3 and §4)
(ii) The duplication problem: how can we avoid the same distributional generalization (for example the homorganicity of a nasal consonant and a following obstruent in consonant clusters) being expressed by both an MSC and a phonological rule (P-rule), and thus making the grammar unnecessarily complex? (§4)
(iii) The status of MSCs: are they absolute constraints, or statistical tendencies only? (§5)

The chapter will conclude with some observations on the expressive functions of morphemes with specific phonotactic properties (§6) and a summary of our findings on the status of MSCs (§7).

2 Morpheme structure conditions

The unequal distribution of phonemes across words and morphemes was an important topic of research in structuralist phonology, because distributional facts were interpreted as signaling the presence or absence of grammatical boundaries, as in the work of Trubetzkoy (Wiese 2001). For instance, in German the phoneme /j/ only occurs at the beginning of lexical morphemes (as in *jagen* 'to hunt' and its derivative *verjagen* 'to chase away'). Hence, the presence of the /j/ is a positive signal of a left-edge morpheme boundary (van Wijk 1939: 125). Such facts show that the phonological and grammatical dimensions of linguistic structure are not completely autonomous, but are related in a systematic fashion (Jakobson 1949). The relation between the distribution of phonemes and grammatical units such as morphemes and words is therefore an aspect of the interface between phonology and morphology. Jakobson (1949) drew attention to the fact that different grammatical units may have different phonotactic properties. For instance, he observed that of the 23 Czech consonants, only eight are found in inflectional suffixes. Jakobson also mentioned that only the following consonants appear in the inflectional suffixes of English: /z d n ŋ/.

Dutch exhibits a number of such asymmetries between lexical morphemes on the one hand and derivational and inflectional suffixes on the other (Booij 1995). For Dutch, the following generalizations hold:

(1) a. Suffixes may consist of consonants only (/s/, /t/ or a combination thereof).
b. Suffixes may begin with the vowel /ə/.
c. Suffixes may have /ə/ as their only vowel.

Lexical morphemes of Dutch, on the other hand, do not have the phonotactic possibilities listed in (1) for suffixes, and require the presence of at least one full vowel (that is, a vowel that is not /ə/; see CHAPTER 26: SCHWA), and cannot be schwa-initial. Dutch prefixes cannot begin with a schwa either, but can have the schwa as their only vowel, as is the case for the Dutch prefix *be-* /bə/. Thus, we can sometimes tell from the phonological make-up of a morpheme whether it is a lexical morpheme or an affix.

A famous type of morpheme structure constraint is the restricted distribution of consonants in Semitic roots (see CHAPTER 108: SEMITIC TEMPLATES). Most Semitic roots are triliteral, that is, they contain three consonants, the consonantal skeleton. These skeletons are intercalated with vowels, and these vowel patterns are the exponents of grammatical information. Greenberg (1950) observed that the first two consonants of a Semitic CCC skeleton cannot be identical, whereas the last two can. Furthermore, homorganic consonants, i.e. consonants with the same place of articulation, are excluded, unless they are identical, even if they are the last two consonants. This is exemplified by the following distributional patterns in Arabic:

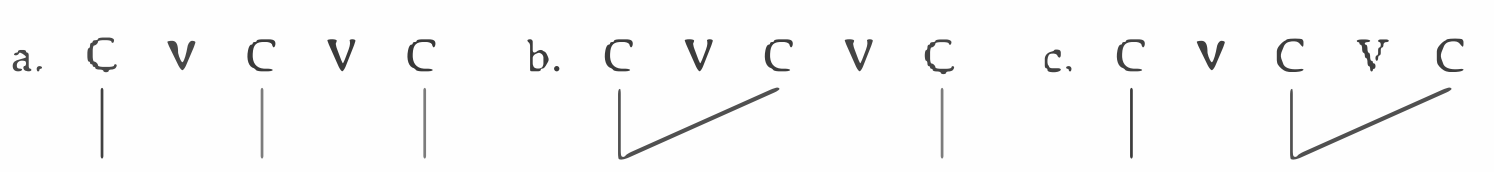
(2)
(2)	*m-m-d	
	m-d-d	'to stretch'
	f-r-r	'to flee'
	*b-m-C	
	*C-b-m	
	*g-k-C	
	ʃ-k-k	'to split'
	*ʃ-k-g	

Similar facts are reported for Modern Hebrew in Bar-Lev (1978: 321): "well-formed roots contain only consonants from different places of articulation." For instance, the following patterns can be observed in existing Modern Hebrew roots (cf. also Berent *et al.* 2002):

(3)	labial – velar – dental	*bagad*	'to betray'
	velar – labial – dental	*gibor*	'hero'
	dental – velar – labial	*dégem*	'model'
	labial – dental – velar	*mélex*	'king'

Crucially, these constraints apply to morphemes only; hence they are tautomorphemic constraints. As McCarthy (1986: 209) observes, there is no Arabic root *tatak*, with two identical consonants, but there are inflected forms of verbs like *ta-takallam* 'you converse' in which the first *t* belongs to a prefix, and hence does not belong to the same morpheme as the second *t*. Thus the prohibition on identical consonants is not violated.

McCarthy (1986) proposes to analyze this constraint as following from the Obligatory Contour Principle (OCP), which states that identical elements on the melodic tier of a morpheme are not permitted (see CHAPTER 14: AUTOSEGMENTS). If the OCP applies to the lexical representations of Arabic roots, and assuming that all autosegmental spreading in Arabic is rightwards, it is predicted that, of the following three structures, only the third one is well-formed (McCarthy 1986: 209):

(4)


In (4a), the OCP is violated on the melodic tier, whereas in (4b) leftward spreading has taken place. Hence, if the V is /a/, the root **sasam* is excluded, and only the root *samam* is well-formed. In order to exclude sequences of homorganic but not identical consonants as well, the OCP must be interpreted here as OCP-Place. That is, assuming that there is a separate tier for Place specifications of sounds, OCP-Place forbids adjacent identical specifications on the Place tier, and this excludes adjacent homorganic consonants (including identical ones, which by definition have the same specification on the Place tier).

Phonotactic properties of morphemes may also reveal that they belong to a particular stratum of the lexicon of that language, and may differentiate between native and borrowed morphemes (see CHAPTER 95: LOANWORD PHONOLOGY). Itô and Mester (1995: 819) mention examples of constraints from several languages that are specific for native morphemes of those languages. Japanese is interesting in that its morphemes can be divided into four subclasses: Yamato (native), Sino-Japanese, Foreign and Mimetic. Each subclass is characterized by a set of constraints, some of which are valid for more than one subclass. For instance, Sino-Japanese roots consist of one syllable only, and Lyman's Law (morphemes contain at most one voiced obstruent) holds for Yamato morphemes only (Itô and Mester 1995).

Dutch words that begin with *pn-* in the spelling, such as *pneumatisch* 'pneumatic' and *pneumonie* 'pneumonia', betray their non-native origin: *pn-* is a well-formed word-initial consonant cluster in Greek, but not in Dutch. This suggests that morphemes with *pn-* do not belong to the set of possible native morphemes of Dutch, and the constraint *[pn- partially characterizes the set of native Dutch morphemes. In English this constraint applies to all morphemes, and hence the combination *pn-* is realized as /n/.

The range of phonotactic patterns found in morphemes may be smaller than those in words. English morphemes, for instance, never end in a cluster of voiced obstruents (there are no morphemes like **lovd* or **dubd*), whereas such clusters do occur in complex words like past tense forms of verbs (as in *loved* /vd/, and *dubbed* /bd/). Dutch morphemes are subject to the constraint that voiced obstruent clusters only occur in complex words: morpheme-internally we only find clusters like /pt/ and /st/, but in complex words we find clusters like /bd/ and /zd/, as in the past tense forms *eb-de* /ɛbdə/ 'receded' and *raas-de* /raːzdə/ 'raged'. The only exceptions to this Dutch MSC are a few loanwords like *labda* /labdaː/ 'lambda' and *budget* /[bœdʒɛt/ 'budget'. Hence, the occurrence of voiced obstruent clusters morpheme-internally makes the relevant Dutch morphemes recognizable as loans (Zonneveld 1983).

As observed by Shibatani (1973), MSCs may have a different status from phonological rules or constraints, in that loanwords are not necessarily adapted to the MSCs of a borrowing language, whereas the application of the phonological constraints cannot be suppressed. Hence, the Dutch loan *labda* keeps its voiced obstruent cluster, and is not pronounced as *la*[pt]*a*. Dressler (1985: 219–245) also provides examples from various languages of distributional patterns that are characteristic of morphemes.

In sum, there are distributional constraints that are characteristic of morphemes, but not of words in general. The question is whether and how they have to be accounted for by a specific type of constraint, the MSC.

3 The redundancy problem

As we saw above, in classical generative phonology constraints on the segmental composition of (lexical) morphemes are interpreted as lexical redundancy rules or morpheme structure constraints (MSCs) (Halle 1959, 1964; Stanley 1967; Chomsky and Halle 1968). For instance, in many languages, nasal consonants in morpheme-internal clusters share their place of articulation with a following consonant. This generalization can be expressed by omitting the place of articulation of the nasals in the lexical representation of the relevant morphemes. A lexical redundancy rule will then fill in the proper value for the feature [place], and thus derive fully specified underlying phonological representations to which the phonological rules of a language apply. Thus, in the phonological component of the grammar the set of rules that express static phonotactic generalizations is ordered to apply before the set of phonological rules that account for alternations. The two sets of rules (MSCs and phonological rules) together are considered to express all the phonotactic regularities of a language (Postal 1968: 214; see also CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION).

The role and importance of MSCs have been questioned for a number of reasons. In the first place, as pointed out by Hooper (1972), the role of the syllable as a domain of phonotactic generalizations cannot be ignored. The notion "syllable" does not play any formal role in the type of generative phonology codified in Chomsky and Halle (1968), but since then a wealth of evidence for the crucial role of the syllable (and larger prosodic units such as the foot and the prosodic word) in phonological analysis has been amassed (Nespor and Vogel 1986). Constraints on syllable structure are by definition constraints on how phonemes can combine into larger units. Hence, a lot of constraints on phoneme sequences are in fact syllable structure constraints (Hooper 1972). For example, the constraint that an English word cannot begin with a consonant cluster of the type nasal + obstruent follows from the universal principle of syllable structure that the sonority of consonants must decrease towards the edges of the syllable (see CHAPTER 49: SONORITY). Thus, sequences like **mpat* and **ntak* are impossible English morphemes. This means that morphemes must have a phonological composition that will lead to well-formed prosodic structures.

A second argument for the syllable as a phonotactic unit is that we cannot determine whether a particular segmental string is ill-formed without taking syllabification into account. For instance, the consonant sequence /bkm/ is always phonotactically ill-formed in English, because there is no possible division across two syllables that leads to a sequence of well-formed syllables. On the other hand, the consonant cluster /kn/, which does not appear word-initially in English words, can appear word-internally, as in *acne*, because this word can be syllabified as *ak.ne*, with two well-formed syllables (dots indicate syllable boundaries). That is, the following generalization holds:

(5) A (grammatical) word is phonotactically well-formed *iff* it can be parsed exhaustively into one or more well-formed prosodic constituents.

The class of well-formed syllables and higher prosodic constituents (foot, prosodic word) can be defined by the prosodification algorithm of each language, which is partially universal and partially language-specific. This algorithm groups the sounds of words into syllables, feet, and prosodic words (Rubach and Booij 1990). Alternatively, the grammar contains a set of ranked prosodic constraints that determine the optimal prosodification of a string of sounds, as in Optimality Theory (Kager 1999).

We might therefore claim that MSCs are superfluous because phonotactic restrictions on morphemes can be seen as the effects of phonological constraints on the output forms of words. For instance, English does not have a morpheme *abkmer*, since this morpheme cannot be prosodified exhaustively: the /k/ cannot be made part of the first or the second syllable. Similarly, the reason that a Dutch lexical morpheme requires the presence of at least one full vowel is that, otherwise, such a morpheme cannot yield a well-formed prosodic word. Dutch suffixes, on the other hand, can be vowelless or contain the vowel /ə/ only, as mentioned in (1), because they always combine with a lexical morpheme. That is, the phonotactic shape of Dutch suffixes has to do with their being dependent on a host morpheme to which they attach. We might therefore consider these suffix-specific phonotactic properties as something that need not be expressed separately, since it follows from the mapping of morphological structure onto prosodic structure.

Hooper (1972) offers a second argument against MSCs, that formulating phonotactic constraints with the morpheme as domain may also lead to spurious generalizations. For instance, Dutch lexical morphemes of Romance origin may end in obstruent clusters that are unpronounceable in isolation:

(6) *castr-eer* 'castrate'
celebr-eer 'celebrate'
emigr-eer 'emigrate'
penetr-eer 'penetrate'

One might conclude that Dutch morphemes can end in clusters of the type /Cr/, but this generalization does not reveal what is really at stake: those morphemes are only acceptable because they are bound roots, obligatorily followed by a vowel-initial suffix. Hence, these consonant clusters will form proper syllable onsets, as in *pe.ne.treer*. Similar observations can be found in Kenstowicz and Kisseberth (1977: 145) for Tunica, and they therefore concluded that in such cases it is the word rather than the morpheme that is the domain of phonotactic constraints. Nevertheless, the occurrence of such root-final consonant clusters is revealing in the sense that they betray the Romance origin of those roots: Germanic roots of Dutch never have this form because they can be used as words without further suffixation.

A third example of the role of prosody in the phonotactics of morphemes is that in many languages lexical morphemes are subject to prosodic minimality conditions. For instance, Dutch lexical morphemes are subject to the constraint that they consist of at least one heavy syllable (with either a long vowel or a short vowel followed by a consonant). That is, a lexical morpheme cannot consist of a light syllable only; bimoraicity is required. It is only in exclamations like *hè* /hɛ/ that the use of such light syllables with a short vowel is possible.

Prosodic conditions on morphemes create a problem for the classical MSCs: the syllable structure of a morpheme is not part of its lexical representations, but a derived property. Therefore, MSCs cannot refer to derived prosodic properties such as bimoraicity (McCarthy 1998). The only way to circumvent this problem is to phrase the constraint in terms of segment sequences: a lexical morpheme must contain either a long vowel, or a short vowel followed by at least once consonant. However, we then miss the generalization that it is a prosodic syllable weight condition that is involved. Once more, this suggests that the segmental composition of morphemes is governed by phonological output conditions. A similar problem occurs when we want to express the following generalization for Dutch: "In mono-morphemic forms we do not find sequences of schwa-headed syllables" (van Oostendorp 1995: 141). Again, this MSC refers to the derived property of syllable structure (cf. Downing 2006 for a cross-linguistic survey of prosodic minimality conditions).

In sum, we have to find a way in which prosodic constraints can account for at least part of the phonotactic constraints on morphemes.

3.1 Non-syllabic sequential constraints

Not all constraints on segmental sequencing can be reduced to syllable structure or prosodic minimality requirements. There are sequential constraints on consonant clusters that hold independently from the tautosyllabic or heterosyllabic status of these clusters. For instance, Yip (1991) proposes the following generalization for English (see also CHAPTER 12: CORONALS):

(7) *Consonant Cluster Condition*

In consonant clusters, consonants may have at most one other articulator feature than Coronal.

Thus, we find English clusters like /pt/ and /kt/ (*apt, act*), but not (tauto- or heterosyllabic) clusters like /kp/, /pk/, /km/, /mk/, /xm/, and /gm/ (loanwords like *drachma* and *stigma* are exceptions to this generalization). Note that the ill-formedness of such clusters does not follow from syllable structure constraints since they could be heterosyllabic. Yet they do not occur. If we come across such clusters in words (as in *zipcode* and *backpack*), we can conclude that these words must be compounds, consisting of more than one lexical morpheme.

An example of a sequential constraint that holds both for tautosyllabic and heterosyllabic sound sequences, observed for English by Davis (1991), and for Dutch by Booij (1995: 46), is that in the sequence sCVC the two Cs should not be identical, unless they are coronal. Here are some Dutch examples with labial and coronal consonants (such sequences of velar consonants do not occur for independent reasons):

(8)

CVC			sCVC		
poep	/pup/	'shit'	**spoep*	/spup/	
mam	/mam/	'mother'	**smam*	/smam/	
toet	/tut/	'face'	*stoet*	/stut/	'procession'

This constraint is also valid for heterosyllabic sequences: they are not acceptable when followed by a vowel, as shown by forms like **spupo* and **smama*.

The point that not all phonotactic constraints can be reduced to syllable structure constraints is particularly clear for word-edge constraints, which are discussed in the next subsection.

3.2 *Word edges*

The difference between syllable structure constraints and sequential constraints is stressed by Kristoffersen (2000: 46–48), in relation to the distribution of consonants at word edges. In Norwegian, the cluster *tl-* is a proper syllable onset. It occurs word-internally in words like *Be.tlem* 'Bethlehem' and *A.tle* (proper name). Yet Norwegian words never begin with this cluster. Kristoffersen also observed that, although Norwegian words never begin with *pn-*, a cluster that does not violate the Sonority Hierarchy Constraint on syllable structure, Norwegians have no difficulty in pronouncing loan words like *pneumatisk* 'pneumatic'. These observations imply that /tl-/ and /pn-/ are proper syllable onsets in Norwegian, and that the non-occurrence of initial /tl-/ and /pn-/ is not due to a syllable structure constraint, but to a constraint that holds for the left edge of Norwegian root morphemes or prosodic words. A similar example from Dutch is that lexical morphemes do not begin with *pj-*, *tj-*, or *kj-*; however, the diminutive suffix allomorphs *-pje*, *-tje*, *-kje* begin with these clusters, and hence these clusters do appear in word-internal syllable onsets, as in *riem-pje* 'belt-DIM' with the prosodic structure $((\text{rim})_\sigma(\text{pjə})_\sigma)_\omega$ (ω = prosodic word, σ = syllable). Therefore, the non-occurrence of these clusters cannot be attributed to a syllable structure constraint. The word-initial sequences /pj- tj- kj-/ do occur in borrowed proper names for male persons, such as *Pjotr*, *Tjeerd*, *Kjeld*, and they do not cause pronunciation problems for speakers of Dutch.

The edges of words may have special phonotactic properties, since they may either impose more restrictions than what syllable well-formedness requires, or allow for extra consonants compared to what is possible in syllables in general. The Norwegian examples above (no *tl-* or *pn-* at the beginning of a word) are a case in point. Other examples of more restricted phonotactics at word edges can be found in Booij (1983): in Huichol, for example, words cannot end in a consonant but syllables can (source: Bell 1976).

In Polish, extra consonants may be added in word-initial position that violate the universal Sonority Sequencing constraint (Rubach and Booij 1990: 434; see also CHAPTER 109: POLISH SYLLABLE STRUCTURE):

(9)
rwać 'tear'
rdza 'rust'
łgnąć 'stick'
mdły 'tasteless'
mnich 'monk'

In these words, a sonorant consonant is followed by a consonant of the same or higher degree of sonority, in violation of the Sonority Sequencing requirement that the sonority of consonants must increase towards the direction of the nucleus. The account that Rubach and Booij (1990) propose is that Polish prosodic words have an extra optional word-initial slot for an extrasyllabic consonant preceding the regular syllables, which is exempt from the requirements of the Sonority

Sequencing condition. This analysis implies that allowing for these marked consonant clusters is not to be seen as a property of lexical morphemes, but of the prosodic words that corresponds with such morphemes.

The special phonotactics of word edges is dealt with in Optimality Theory in the form of alignment constraints (McCarthy and Prince 1993). The basic idea of this approach, which makes crucial use of ranked output constraints in computing the phonetic form of words, is that there are alignment constraints that require the alignment of prosodic and grammatical boundaries. According to McCarthy and Prince (1993), the language Axininca Campa has word-initial onsetless syllables, whereas word-internally a vowel hiatus must always be filled by an epenthetic consonant. The relevant alignment constraint blocks the insertion of an epenthetic consonant in word-initial position. If epenthesis took place, there would be no alignment of the left edge of the prosodic word with the left edge of the (vowel-initial) morpheme. That is, the alignment constraint is ranked higher than the constraint that penalizes empty onsets. Note, however, that this analysis does not directly express that the left edges of Axininca Campa morphemes can begin with a vowel, even though syllables in this language normally begin with a consonant. The alignment mechanism allows for a difference in make-up between the edges of morphemes and syllables, but does not express it.

3.3 *Phonotactic differences between simplex and complex words*

As briefly mentioned at the end of §2, the range of phonotactic patterns in morphemes may be smaller than in complex words. Harris (1994) presents a number of observations on the phonotactic differences between simplex words and complex words in English. For instance, one will not find a heterosyllabic sequence /pt/, as in *laptop*, within a morpheme (except for loans like *helicopter*), even though a heterosyllabic cluster /pt/ would not violate the syllable structure constraints of English: a syllable can end in a /p/, and begin with a /t/. The same applies to the cluster /pw/: a proper name like *Sopwith*, which is historically a compound, is exceptional in this respect, and thus betrays its historical origin as a compound. Such opaque compounds tend to be adapted to the phonotactic patterns. The proper name *Greenwich* with the sequence /nw/ is now pronounced without the /w/, thus adapting to the phonotactic constraints for monomorphemic words (Harris 1994: 51).

The observation that certain consonant clusters only occur at morpheme boundaries is often used in linguistic analyses for assigning multi-morphemic status to words (see CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS). For instance, many words in the Amerindian language Athapaskan are considered to be compounds, even though the constituents do not occur as words by themselves, because they contain consonant clusters that are characteristic of morpheme boundaries (Rice 2009: 546).

Phonotactic differences between root morphemes and complex words have also been observed for vowel harmony (cf. §4 for a more detailed discussion of such facts for Turkish). The necessity of a separate morpheme structure condition on vowel combinations in roots is explicitly defended in the analysis of Hungarian vowel harmony in Vago (1976); see also CHAPTER 123: HUNGARIAN VOWEL HARMONY. Harvey and Baker (2005: 1459) observed that in the Australian language

Warlpiri, a language with vowel harmony with respect to the feature [round], the sequence [−round][+round] is not permitted for two consecutive vowels (with intermediate consonants) within roots, whereas the disharmonic sequence [+round] [−round] is. They account for this difference not by assuming an agreement constraint, but by proposing separate constraints for each type of disharmonic sequence. In addition, there is a constraint of root identity that requires the feature specifications for [round] to be preserved in the output. Thus they do not need to assume two rules of vowel harmony, a morpheme structure constraint and a phonological constraint that applies to complex words, and the duplication problem is avoided. Note, however, that this analysis requires reference to the root, a type of morpheme, as the domain of an identity constraint. That is, reference to morphemes in phonological constraints is still required.

Different phonotactics may also play a role in recognizing the lexical category of a word. In Dutch, there is a marked difference in phonological make-up between simplex nouns and simplex verbs. Verbs tend to consist of at most two syllables; if there is a second syllable, it will end in a schwa followed by a liquid. Nouns, on the other hand, allow for a larger variety of phonological structures, such as those consisting of three or more syllables, or ending in a full vowel. It appears that speakers of Dutch are able to categorize words as nouns or verbs on the basis of such phonotactic knowledge (Don and Erkelens 2006).

In sum, the distributional properties of segments within morphemes relate to the phonological rules or constraints of the relevant language, but not all morpheme-internal phonotactics can be reduced to these more general phonological regularities. In the words of Stanley (1967: 397): "The constraints holding within single morphemes are more restrictive than the constraints which characterize larger units."

4 The duplication problem

The problem that the assumption of both MSCs and P-rules seems to lead to unnecessary complications of the grammar was noted by Stanley (1967). For instance, Turkish has two general P-rules of vowel harmony that also predict the distribution of vowels within morphemes: all vowels agree in backness, and high vowels agree in roundness (see CHAPTER 118: TURKISH VOWEL HARMONY). As Zimmer (1969: 310) points out:

> The restrictions on vowel co-occurrence within almost all bases of Turkic origin are nearly the same as those just described for suffix vowels; thus for the "harmonic" part of the lexicon, there are two MSC's which replicate, to a great extent, the vowel-harmony rules that determine the selection of vowels in suffixes. There is, however, a large number of loanwords to which these vowel harmony MSC's do not apply, – e.g. /günah/ 'sin', /kalem/ 'pen', /sosis/ 'sausage', /viraž/ 'curve'.

In addition, there is an MSC that does not double as a P-rule, the Labial Consonant MSC (Zimmer 1969: 312):

(10) After /a/, a [+high] vowel agrees in labiality with a preceding [+labial] consonant.

An example of a morpheme that obeys MSC (10) is *karpuz* 'watermelon', in which the second vowel is round, even though the first vowel is non-round. This is an interesting MSC for the debate on the redundancy of MSCs, since it has no P-rule counterpart.

This duplication problem, already noted by Stanley (1967), is discussed by Anderson (1974: ch. 16). Anderson observes that many Turkish morphemes such as *kitap* 'book' are disharmonic, but do not block the application of vowel harmony rules once they have been suffixed. Anderson therefore concludes that we need both MSCs and P-rules for vowel harmony in Turkish, since they may be subject to different idiosyncrasies. In Anderson's view, the relation between MSCs and P-rules dealing with vowel harmony is a functional one, which need not and cannot be expressed formally by unifying them into one rule.

Shibatani (1973) proposes that such constraints should be considered to be both MSCs and Surface Phonetic Constraints (SPCs). A constraint can be marked as an MSC, a SPC, or both an MSC and a SPC. Clayton (1976) argues that constraints that hold for underlying forms of morphemes only are unmotivated, and do not reflect the speaker's knowledge of his/her language. Therefore, Clayton claims that Surface Phonetic constraints suffice.

The duplication problem is also considered by Kiparsky (1982: 167–170), in the framework of Lexical Phonology (see also Kaisse and Shaw 1985: 25; CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME). In this framework, phonological rules apply cyclically. Rules apply either in a structure-adding fashion or in a structure-changing one. Rules only apply in a structure-changing fashion in derived environments, i.e. in environments created by the previous application of a morphological or a phonological rule (see Booij 2000 for a survey of this theory). Kiparsky (1982) proposed that there are no Morpheme Structure Rules. The lexical representations of morphemes are underspecified; that is, predictable properties are omitted. On the first cycle, phonological rules specify these features; i.e. they fill in the blanks. If a word is complex, the same rule can apply in a structure-changing fashion on the next cycle, since the complex word is a derived environment. For instance, the Dutch rule that requires obstruents to have the same specification for the feature [voice] as an adjacent obstruent can be applied as a blank-filling rule to the underspecified feature matrix for /x/ in a word like *achter* /ɑxtər/ 'back', and it can be applied in a structure-changing fashion in a complex word like *as-bak* 'ash-tray' (underlying /ɑs-bɑk/; phonetic form [ɑzbɑk]). In English, the place of articulation of the nasal consonant in *damp* can be left unspecified, and filled in by the rule of Nasal Place assimilation, whereas the same rule will change the underlying coronal nasal /n/ into [m] in the derived word *compress* (underlyingly *con-press*).

Such an analysis can also deal with exceptions to MSCs. For instance, the Dutch word *imker* /ɪmkər/ 'bee-keeper', which is synchronically a simplex word, violates the constraint on homorganicity of nasal–obstruent clusters. In Kiparsky's proposal, this is no problem: the nasal consonant will be fully specified as being labial, and the rule that predicts the feature [velar] for a nasal followed by /k/ is blocked from applying because feature-changing applications of this rule are allowed in derived environments only.

In the case of Turkish vowel harmony, the same solution would apply. Disharmonic roots are fully specified, and therefore the P-rules of vowel harmony are blocked from applying to these morphemes, whereas they will apply in derived environments, to the vowels of the suffixes.

In short, in Kiparsky's proposal, the duplication problem is solved by abolishing the class of morpheme structure rules, and having P-rules apply in two different fashions. However, not all types of generalizations over the phonological shape of morphemes mentioned above can be expressed this way. This applies in particular to prosodic conditions on the shape of morphemes.

4.2 MSCs in Optimality Theory: Lexicon optimization and output–output faithfulness

Optimality Theory (OT) does not allow for constraints on the inputs of phonological evaluation. Output constraints are the only mechanisms for expressing phonotactic patterns. This idea of OT is referred to as the Richness of the Base hypothesis. For instance, there is no input constraint that forbids the morpheme **bnik* as a morpheme of English. The output constraints will penalize such a form, and evaluate this form in such a way that the optimal output form is not faithful to this form, but different, e.g. *blik*. Since forms such as *bnik* will never surface in English, it does not make sense to store an underlying form *bnik* for *blik*. This is the effect of lexicon optimization. Thus, the phonological output constraints of a language will be reflected by the input forms. This point of view is foreshadowed in Sommerstein (1974: 73), who argued that judgments about whether a sound sequence is a possible morpheme must be made on the basis of surface representations.

This idea is discussed in more detail in McCarthy (1998, 2002, 2005), and can be illustrated as follows. Suppose there is a language with the constraint that obstruents are voiceless at the end of a syllable, and with the suffix /-ən/ as plural ending for nouns, as in [hut] – [hutən] 'hat(s)'. Furthermore, this language has no alternations of the type [hut] – [hudən]. That is, morphemes that end in an obstruent will always end in a voiceless obstruent. Given the word [hut], the Richness of the Base hypothesis implies that we might assume the underlying form /hud/ for the singular form. The correct phonetic form [hut] will be computed anyway. However, in an optimal lexicon the underlying form to be chosen will be /hut/ because of lexicon optimization. This means that of the possible tableaux that select the right form, the most harmonic one will be selected, i.e. the one with the minimal number of violations of constraints. The assumption of the underlying form /hud/ will imply violation of the input–output (IO) faithfulness condition, unlike the underlying form /hut/. IO faithfulness requires the underlying form to be selected as the surface form, unless it is overruled by higher-ranked constraints. The optimal underlying form can be selected by comparison of tableaux, and the selection of the most harmonic one. Thus, lexicon optimization makes restrictions on input forms superfluous.

If there are no MSCs, the question arises of how to account for constraints that hold for morphemes only. One example of a distributional difference between morphemes and words concerns the distribution of nasals in Dutch. Within morphemes, nasal consonants are always homorganic with a following obstruent (with the exception of *imker*; cf. §4.1). Hence, we find *damp* /dɑmp/ 'damp', *tand* /tɑnt/ 'tooth' and *dank* /dɑŋk/ 'thanks', but no morphemes ending in */-mt -mk -ŋp -ŋt/. On the other hand, complex words such as the 3rd singular present forms of verbs always end in /-t/, preceded by all three types of nasals: *klim-t* 'climbs', *zon-t* 'sunbathes', *zing-t* /zɪŋt/ 'sings'. If we assume a markedness constraint NC (nasals are homorganic with a following consonant), this constraint must be blocked from changing a verb form like *klimt* into *klint*. McCarthy (1998) argued that this can be achieved by making use of output–output (OO) correspondence constraints. If we rank OO faithfulness conditions for the relation between a base word and its derivatives higher than the markedness constraint NC, a verb form like *klimt* cannot be changed to *klint*, because this would violate the requirement of correspondence of the stem of this inflected form with the verbal stem *klim*.

Nasal assimilation should not be blocked in all derived environments, however. In a prefixed word like *compress*, the prefix-final /n/ of *con-* does assimilate to the next /p/. This can be accounted for if we assume that the NC constraint is ranked higher than faithfulness to the underlying form of the prefix, /kɔn/. The more general observation is that affixes tend to adapt to roots rather than the other way around. Hence, in OT analyses it is often assumed that faithfulness constraints for affixes rank lower than those for roots (Alderete 2003). This implies that constraints have to be indexed for particular morphological categories such as root and affix. Therefore, we have to allow for reference to morphological domains in a system of phonological output constraints.

4.3 *Domains and strata*

Our conclusion so far is that, even when we do not allow for constraints on underlying forms of morphemes, it should be possible to index a phonological output constraint for a particular morphological domain such as the lexical morpheme. This will make it possible to specify distributional constraints that hold for lexical morphemes only. For instance, the Dutch constraint that lexical morphemes and prefixes cannot begin with /Cj/, whereas suffixes can, is expressible by indexing this constraint for the relevant morphological domains.

As observed in §3, particular phonotactic properties may only hold for certain strata of the lexicon. This is discussed in detail for Japanese by Itô and Mester (1995, 1999, 2001), who argue that "phonological generalizations can be covert by being lexically partial: they hold within a subdomain of the lexical space, but are violated in peripheral areas occupied e.g. by loanwords or onomatopeia" (Itô and Mester 2001: 274). For instance, in Japanese the palatalization constraint that changes /t/ into /ts/ before /i/ (this constraint also excludes tautomorphemic /ti/ sequences) does not affect loanwords like English *tea* and *party*. Therefore, Itô and Mester defend the idea of stratum-specific (ranking of) faithfulness constraints.

A Dutch example of a stratum-specific constraint was mentioned in §2: in native Dutch morphemes morpheme-internal obstruent clusters are always voiceless, but this constraint does not hold for non-native morphemes such as *labda* /lɑbdaː/ 'lambda', or the brand name *Mazda* /mɑzdaː/. Such non-native morphemes preserve their foreign pronunciation. The word *labda*, for instance, will not be changed to [lɑptaː], and Dutch speakers will recognize it as a loan due to this phonological property. Another example is the word-initial cluster *sk-*, which does not occur in Dutch native words, but only in loans from English, e.g. *scan* and *Skype*.

A similar distinction between native and non-native morphemes can be observed for languages with vowel harmony: non-native lexical morphemes may be disharmonic, and this is not changed through application of the vowel harmony constraints. For instance, the Hungarian noun *soför* 'chauffeur' is disharmonic (the first vowel is back, the second one is front), and remains so, even though it selects its suffix vowels in accordance with the frontness/backness vowel harmony constraint.

In sum, phonological constraints may have to be indexed for particular morphological categories or for lexical strata.

5 Absolute constraints or tendencies?

A final important point of debate concerning MSCs is whether they are absolute constraints or just statistical tendencies. Zimmer (1969) investigated the psychological reality of the Labial Consonant MSC of Turkish (10) mentioned above. Recall that this constraint holds for morphemes only, and is not supported by the two P-rules of vowel harmony. Zimmer made up lists of pairs of nonsense words, and asked subjects, Turkish students in California, to determine which word of such a pair sounds more like a word that might actually occur in Turkish. In the case of word pairs where the P-rules of vowel harmony played a role, the results were as expected, with a strong preference for the word in accordance with the vowel harmony constraints. For the pairs that involved the Labial Consonant MSC, on the other hand, there was hardly any difference in number between expected responses (the words in accordance with the MSC) and unexpected responses. Zimmer (1969: 320) therefore concluded that an MSC that is not supported by a P-rule might not be internalized by native speakers of Turkish.

5.1 OCP-Place

The psychological reality of OCP-Place, discussed in §2, which excludes identical adjacent place specifications, has been investigated for speakers of Jordanian Arabic (Frisch and Zawaydeh 2001; Frisch *et al.* 2004). It appears that "Jordanian Arabic speakers do recognize systematic gaps that are violations of OCP-Place as different from accidental gaps involving unrelated consonant pairs" (Frisch and Zawaydeh 2001: 99), even though there are violations of OCP-Place. Frisch *et al.* (2004) argue that OCP-Place is not an absolute, universal constraint. They consider the constraint as reflecting the generalizations that Arabic speakers make on the basis of their lexicon. OCP-Place is claimed to be a gradient constraint, since there are quite a number of words that violate it, but in different degrees. "Forms that violate the constraint to a lesser degree are more frequent than forms that violate the constraint to a greater degree" (Frisch *et al.* 2004: 182). Frisch *et al.* also point out that the co-occurrence of homorganic consonants that are non-adjacent (occurring in the first and third positions) is less restricted than the co-occurrence of adjacent homorganic consonants. In other words, the OCP-Place constraint is a gradient, but psychologically real constraint: "the native speaker knows an abstract but gradient OCP Place constraint ('Roots with repeated homorganic consonants are unusual') based on generalizations over the statistical patterns found in the lexicon" (Frisch *et al.* 2004: 216).

Frisch *et al.* also looked at the effect of OCP-Place on the borrowing of Italian verbs in Maltese, a variety of Arabic with many loans from Italian. The number of Italian verbs whose consonant pattern conforms to OCP-Place is significantly higher than that of the Italian verbs that violate OCP-Place (but these latter verbs may also be borrowed, and adapted to Maltese). This again supports the psychological reality of such a constraint, without it being categorical.

These findings suggest that OCP-Place is a gradient constraint that aims at avoidance of similarity: the more similar adjacent consonants are, the more they are avoided. Speakers are able to make phonotactic generalizations about lexical morphemes, but the corresponding constraints need not be categorical.

Statistical tendencies for the composition of various morphological categories such as the root and the stem have been observed by Wiese (2001): 94% of all German roots begin with a consonant, and 96% of all German roots end in a consonant. In OT, this can be expressed by alignment constraints that require the left and right edges of a root to coincide with the feature [+cons]. For those roots that violate the constraint, IO faithfulness will preserve the vowel at the edges. Note in particular that the tendency to have consonants at the end of roots does not follow from a syllable constraint, since the universally most unmarked syllable is the open syllable. Thus, this type of distribution may function as a boundary signal.

5.2 *Constraints on underlying forms*

Dutch exhibits intriguing constraints on sequences of vowels followed by fricatives. The basic generalization is that a vowel is short before /f s/, whereas it is long before /v z/. Let us call this the VZ constraint. Due to the effect of Final Devoicing, the constraint that obstruents are voiceless at the end of a syllable, this constraint can only be observed directly if the fricative is not morpheme-final. The following morphemes illustrate this constraint:

(11)	*short vowel*			*long vowel*			*excluded*
	effen	/ɛfən/	'even'	*even*	/e:vən/	'even'	/ɛvən, e:fən/
	dissel	/dɪsəl/	'pole'	*vezel*	/ve:zəl/	'fiber'	/ɛzəl, e:səl/

This constraint is violated by a few loanwords like *mazzel* /mɑzəl/ 'good luck' and *puzzel* /pœzəl/ 'puzzle', and by the native morpheme *oefen-* /u:fən/ 'to exercise'. This shows that this constraint is not an absolute condition on pronounceability, but a statistical generalization about morphemes.

This VZ constraint seems to apply to intervocalic sequences only, since we do find long vowels followed by /f s/ at the end of morphemes, as in the singular forms of the following nouns:

(12)	*graaf*	[ɣra:f]	'earl (SG)'	*grav-en*	[ɣra:vən]	(PL)
	kaas	[ka:s]	'cheese (SG)'	*kaz-en*	[ka:zən]	(PL)

However, we can interpret this constraint as also applying to morpheme-final sequences if we assume the constraint to hold for the underlying forms of morphemes. Morphemes like *graaf* and *kaas* end in a voiced fricative underlyingly, as shown by their plural forms, and hence the underlying forms of these morphemes are /ɣra:v/ and /ka:z/, respectively. There are a few exceptions, such as the non-native word *graaf* 'graph' with the plural form *graf-en* [ɣraːfən]. In the case of /s/ *vs.* /z/, the number of exceptions is much higher, since there are a number of verbs like *eis-en* [ɛisən] 'require (INF)' and *ruis-en* [rœysən] 'rustle (INF)', in which the diphthong, which counts as a long vowel, is followed by [s]. Speakers of Dutch may thus recognize a plural form as *grafen* as being non-native, whereas the phonetic form of its singular form *graaf* [ɣraːf] does not betray this stratal property.

If we allow for phonotactic constraints on underlying forms that cannot be observed from their corresponding surface forms, this enables us to make generalizations about the kind of alternations we may find in a language (Booij 1999). This topic is also broached in Ernestus and Baayen (2003). They raise the question to what extent the occurrence of alternations between morpheme-final voiceless and morpheme-final voiced obstruents in Dutch morphemes is predictable. There appear to be clear regularities. For instance, if a Dutch morpheme ends in a long vowel plus a labial stop, that stop is always an underlying /p/; that is, we do not find the alternation [p – b] for such morphemes. In the case of the fricatives discussed above, we saw that the length of the preceding vowel is a strong predictor of whether the final obstruent is underlyingly voiced or voiceless, although stronger for /f v/ than for /s z/. The question is whether language users possess this kind of knowledge. Ernestus and Baayen (2003) tested this by asking subjects to make past tense forms for nonsense words. If the underlying form of the verbal root morpheme ends in a voiceless obstruent, the past tense suffix *-te* /tə/ will be chosen, and *-de* /də/ otherwise. It appeared that language users do make use of the phonotactic tendencies involved: there is a strong correlation between the proportion of *-te*/*-de* choices for nonsense morphemes and the proportion for existing morphemes with a similar phonotactic make-up. Ernestus and Baayen (2003) therefore concluded that the speaker chooses an underlying representation for a nonsense morpheme that makes it resemble similar morphemes in the lexicon. As was the case for the Arabic roots discussed in §5.1, such phonotactic generalizations concerning morphemes may be statistical rather than absolute in nature. Moreover, they may pertain to underlying forms, where properties are present that may not be accessible in surface form.

This kind of knowledge about the type of alternations that occur in a language may also be formalized without restricting such MSCs to underlying forms. Consider the singular/plural pairs of nouns in Dutch with a stem-final obstruent such as *hoed* [hut] – *hoed-en* [hudən] 'hat (SG, PL)'. The voice specification of the stem-final obstruent of the morpheme *hoed* can only be determined on the basis of the plural form. The plural form is the most informative form of the paradigm (Albright 2005, 2008), and we may assume that it is stored in lexical memory. The relation between the two forms can be specified by a schema of the following type:

(13) $[x]_{SG} \leftrightarrow [x - \text{ən}]_{PL}$

(The symbol ↔ indicates the correlation between the two forms; *x* is a variable for a string of segments.) The plural form is the only reliable form for the computation of the underlying form, that is, the form on the basis of which new derived words and inflected forms can be computed. For instance, if we were to coin the adjective *hoed-ig* [hudəx] 'hat-like', the stem has to end in a /d/, since the phonetic form [hutəx] is wrong. That is, an underlying form is not necessarily a lexically stored representation, but may be computed when necessary for a morphological operation. In the case of Dutch verbs, we need to compute the underlying form of the verbal stem for the choice of the proper form of the past tense suffix (*-te* or *-de*).

Recall now the generalization for Dutch verbal stems that after a long vowel (VV) there is never a *p/b* alternation: if a singular form ends in [VVp], its plural form will never end in [VVbən]. This generalization also holds for nouns, and this can be expressed by the following subschema of (13) (*y* is a variable for segmental strings):

(14) $[y\ \text{VVp}]_{\text{SG}} \leftrightarrow [y\ \text{VVp-ən}]_{\text{PL}}$

In the case of the Dutch s/z alternations discussed above, we might assume subschemas like the following for nouns:

(15) a. $[y\ \text{VVs}]_{\text{SG}} \leftrightarrow [y\ \text{VVz-ən}]_{\text{PL}}$ (as in *kaas* – *kazen* 'cheese (SG, PL)')
b. $[y\ \text{Vs}]_{\text{SG}} \leftrightarrow [y\ \text{Vs-ən}]_{\text{PL}}$ (as in *kas* – *kassen* 'greenhouse (SG, PL)')

If such schemas do not apply to all words, that is, if they are statistical generalizations only, they can be given a weight that indicates their probability.

The generalizations expressed in (15) apply almost exceptionlessly to nouns, and are confirmed by irregular pairs of singular and plural nouns with vowel length alternation. Vowel Lengthening is no longer a regular rule of Dutch, but an idiosyncratic alternation, a relic of Prokosch's Law that applied to Early Germanic, illustrated here for the noun *glas*:

(16) glas [ɣlɑs] 'glass (SG)' glaz-en [ɣlaːzən] 'glass (PL)'

The correspondence between the length of the vowel and the [voice] specification given in (15) is maintained in these irregular pairs by the combination of vowel length alternation and choice of obstruent: the forms *[ɣlɑzən] and *[ɣlaːsən] are both ill-formed.

In sum: whether there is an alternation between a voiced and a voiceless stem-final obstruent for Dutch lexical morphemes can only be determined with 100 percent certainty on the basis of inflected forms such as plurals. Yet, segmental composition of the lexical morpheme may give a clue, in some cases with almost 100 percent reliability. This type of knowledge may be modeled by constraints on the underlying forms of lexical morphemes, or by alternation schemas of the type proposed in (14) and (15).

6 The expressive value of phonotactics

A final phenomenon to be discussed is that particular sound sequences may have a specific semantic or pragmatic value. Sound symbolism is the usual term for such phenomena (see Hinton *et al.* 1994 for a number of detailed studies). In particular, there are phonaesthemes, recurring sounds or sound sequences, with a particular value. For instance, Marchand (1969: 397) argued that "/ɪ/ is suggestive of the subjective, emotionally small and is therefore frequent with diminutive and pet suffixes," and Bauer (1996) also found a cross-linguistic tendency for the use of this vowel in diminutive suffixes (see also Nichols 1971 for consonant sound symbolism in diminutives). According to Marchand (1969), the word-initial sequence *fl-* in words like *flick, flip, flap, flash* is expressive of brisk, quick movement, and Marchand provides many examples of such phonaesthemes. Such sound combinations are not to be considered as morphemes by themselves; yet they have a particular value. Hence, one may claim that phonotactic properties of morphemes can have an expressive role.

Japanese has a class of mimetic morphemes, which are sound-imitating or manner-symbolic roots. These morphemes have to be minimally bimoraic, and usually they appear in reduplicated or some other bipodic form (Mester and Itô 1989: 268):

(17) *poko-poko* 'up and down movement'
noro-noro 'slow movement'
paku-paku 'munching'
pata-pata 'palpitating'

In her study of the expressive value of lexical patterns, Klamer (2002) observed that the violation of general phonotactic constraints in specific classes of lexical items may have an expressive value. That is, a marked semantic value correlates with a marked phonotactic structure. An example from Dutch is the class of monosyllabic words of the type /lVl/, that is, words with the same consonant /l/ in onset and coda. Such words violate the phonotactic constraint or tendency of Dutch that the liquid consonants /l r/ in a syllable should be different. Words with this kind of phonotactics may have marked interpretations, as the following examples from Klamer (2002: 273) illustrate:

(18) *lal* (V) 'jabber, babble, slur one's words'
lel (N) 'earlobe, clout, whopper'
lil (V) 'quiver'
lol (N) 'fun, lark, trick'
lul (N) 'prick, jerk'
lul (V) 'talk nonsense'

Morphemes in which the vowel of /lVl/ is long do not occur at all.

In sum, the expressive value of phonotactic patterns within morphemes may be considered from a different angle: the violation of a constraint may have expressive value.

7 Conclusions

There is no doubt that there are distributional generalizations concerning the phonological make-up of morphemes that need to be expressed somehow in a proper phonological theory. The main theoretical issues are to what extent they can be made to follow from phonological generalizations that also hold for larger units than morphemes, and whether they are absolute constraints, or gradient constraints that express statistical tendencies. MSCs may also reveal different layers of the lexicon. Thus this chapter provides a range of data, observations, and considerations that can be used as a testing ground for the adequacy of theoretical phonological models.

ACKNOWLEDGMENTS

I would like to thank Paulo Chagas de Souza, Moira Yip, two anonymous reviewers, and the editors for their constructive comments and advice on an earlier draft of this chapter. In writing the final version, I also profited from taking part in a phonology seminar on phonotactic learning taught by Adam Albright at MIT and Harvard in the Spring semester of 2010.

REFERENCES

Albright, Adam. 2005. The morphological basis of paradigm leveling. In Downing *et al.* (2005), 17–43. Oxford: Oxford University Press.

Albright, Adam. 2008. Inflectional paradigms have bases too: Evidence from Yiddish. In Asaf Bachrach & Andrew Nevins (eds.) *Inflectional identity*, 271–312. Oxford: Oxford University Press.

Alderete, John. 2003. Structural disparities in Navajo word domains: A case for LexCat-Faithfulness. *The Linguistic Review* 20. 111–157.

Anderson, Stephen R. 1974. *The organization of phonology*. New York: Academic Press.

Bar-Lev, Zev. 1978. The Hebrew morphemes. *Lingua* 45. 319–331.

Bauer, Laurie. 1996. No phonetic iconicity in evaluative morphology. *Studia Linguistica* 50. 189–206.

Bell, Alan. 1976. The distributional syllable. In Alphonse Juilland (ed.) *Linguistic studies offered to Joseph Greenberg*, vol. 2, 249–262. Saratoga, CA: Anma Libri.

Berent, Iris, Gary F. Marcus, Joseph Shimron & Adamantios I. Gafos. 2002. The scope of linguistic generalizations: Evidence from Hebrew word formation. *Cognition* 83. 113–139.

Booij, Geert. 1983. Principles and parameters in prosodic phonology. *Linguistics* 21. 249–280.

Booij, Geert. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.

Booij, Geert. 1999. Morpheme structure constraints and the phonotactics of Dutch. In Harry van der Hulst & Nancy Ritter (eds.) *The syllable: Views and facts*, 53–68. Berlin & New York: Mouton de Gruyter.

Booij, Geert. 2000. The phonology–morphology interface. In Lisa Cheng & Rint Sybesma (eds.) *The first Glot International state-of-the-art book: The latest in linguistics*, 287–306. Berlin & New York: Mouton de Gruyter.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clayton, Mary. 1976. The redundancy of underlying morpheme structure conditions. *Language* 52. 295–313.

Davis, Stuart. 1991. Coronals and the phonotactics of nonadjacent consonants in English. In Paradis & Prunet (1991), 49–60.

Don, Jan & Marian Erkelens. 2006. Vorm en categorie. *Taal en Tongval* 19. 40–53.

Downing, Laura J. 2006. *Canonical forms in prosodic morphology*. Oxford: Oxford University Press.

Downing, Laura J., T. A. Hall & Renate Raffelsiefen (eds.) 2005. *Paradigms in phonological theory*. Oxford: Oxford University Press.

Dressler, Wolfgang U. 1985. *Morphonology: The dynamics of derivation*. Ann Arbor, MI: Karoma.
Ernestus, Mirjam & R. Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79. 5–38.
Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22. 179–228.
Frisch, Stefan A. & Bushra A. Zawaydeh. 2001. The psychological reality of OCP Place in Arabic. *Language* 77. 91–106.
Greenberg, Joseph H. 1950. The patterning of root morphemes in Semitic. *Word* 6. 162–181.
Halle, Morris. 1959. *The sound pattern of Russian: A linguistic and acoustical investigation*. The Hague: Mouton.
Halle, Morris. 1964. On the basis of phonology. In Jerry A. Fodor & Jerrold J. Katz (eds.) *The structure of language: Readings in the philosophy of language*, 324–333. Englewood Cliffs, NJ: Prentice-Hall.
Harris, John. 1994. *English sound structure*. Oxford: Blackwell.
Harvey, Mark & Brett Baker. 2005. Vowel harmony, directionality and morpheme structure constraints in Warlpiri. *Lingua* 115. 1457–1474.
Hinton, Leanne, Johanna Nicols & John J. Ohala (eds.) 1994. *Sound symbolism*. Cambridge: Cambridge University Press.
Hooper, Joan B. 1972. The syllable in phonological theory. *Language* 48. 525–540.
Itô, Junko & Armin Mester. 1995. Japanese phonology. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 817–838. Cambridge, MA & Oxford: Blackwell.
Itô, Junko & Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.) *The handbook of Japanese linguistics*, 62–100. Malden, MA & Oxford: Blackwell.
Itô, Junko & Armin Mester. 2001. Covert generalizations in Optimality Theory: The role of stratal faithfulness constraints. *Studies in Phonetics, Phonology, and Morphology* 7. 273–299.
Jakobson, Roman. 1949. L'aspect phonologique et l'aspect grammatical du langage dans leurs interrelations. Reprinted 1963 in Roman Jakobson. *Essais de linguistique générale*, 161–175. Paris: Minuit.
Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.
Kaisse, Ellen M. & Patricia A. Shaw. 1985. On the theory of Lexical Phonology. *Phonology Yearbook* 2. 1–30.
Kenstowicz, Michael & Charles W. Kisseberth. 1977. *Topics in phonological theory*. New York: Academic Press.
Kiparsky, Paul. 1982. From Cyclic Phonology to Lexical Phonology. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part I, 131–175. Dordrecht: Foris.
Klamer, Marian. 2002. Semantically motivated lexical patterns: A study of Dutch and Kambera expressives. *Language* 78. 258–286.
Kristoffersen, Gjert. 2000. *The phonology of Norwegian*. Oxford: Oxford University Press.
Marchand, Hans. 1969. *The categories and types of present-day English word formation*. Munich: Beck.
McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17. 207–263.
McCarthy, John J. 1998. Morpheme structure constraints and paradigm occultation. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 32(2). 123–150.
McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Oxford: Oxford University Press.
McCarthy, John J. 2005. Optimal paradigms. In Downing *et al.* (2005), 170–210.
McCarthy, John J. & Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology 1993*. 79–153.
Mester, Armin & Junko Itô. 1989. Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language* 65. 258–293.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Nichols, Johanna. 1971. Diminutive consonant symbolism in western North America. *Language* 47. 826–848.

Oostendorp, Marc van. 1995. Vowel quality and phonological projection. Ph.D. dissertation, University of Tilburg.

Paradis, Carole & Jean-François Prunet (eds.) 1991. *The special status of coronals: Internal and external evidence*. San Diego: Academic Press.

Postal, Paul. 1968. *Aspects of phonological theory*. New York: Harper & Row.

Rice, Keren. 2009. Athapaskan: Slave. In Rochelle Lieber & Pavol Stekauer (eds.) *The Oxford handbook of compounding*, 542–573. Oxford: Oxford University Press.

Rubach, Jerzy & Geert Booij. 1990. Edge of constituent effects in Polish. *Natural Language and Linguistic Theory* 8. 427–463.

Shibatani, Masayoshi. 1973. The role of surface phonetic constraints in generative phonology. *Language* 49. 87–106.

Sommerstein, Alan H. 1974. On phonotactically motivated rules. *Journal of Linguistics* 10. 71–94.

Stanley, Richard S. 1967. Redundancy rules in phonology. *Language* 43. 393–436.

Vago, Robert M. 1976. Theoretical implications of Hungarian vowel harmony. *Linguistic Inquiry* 7. 243–263.

Wiese, Richard. 2001. The structure of the Germanic vocabulary: Edge marking of categories and functional considerations. *Linguistics* 39. 95–115.

Wijk, Nicolaas van. 1939. *Fonologie: Een hoofdstuk uit de structurele taalwetenschap*. The Hague: Martinus Nijhoff.

Yip, Moira. 1991. Coronals, consonant clusters, and the coda condition. In Paradis & Prunet (1991), 61–78.

Zimmer, Karl. 1969. Psychological correlates of some Turkish morpheme structure conditions. *Language* 45. 309–321.

Zonneveld, Wim. 1983. Lexical and phonological properties of Dutch voicing assimilation. In Marcel van den Broecke, Vincent van Heuven & Wim Zonneveld (eds.) *Sound structures: Studies for Antonie Cohen*, 297–312. Dordrecht: Foris.

MARC VAN OOSTENDORP is Senior Researcher at the Department of Variationist Linguistics at the Meertens Institute of the Royal Netherlands Academy of Arts and Sciences, and Professor of Phonological Microvariation at the University of Leiden. He holds an M.A. in Computational Linguistics and a Ph.D. from Tilburg University.

COLIN J. EWEN is Professor and Chair of English Language and Linguistics at the University of Leiden. He holds an M.A. in English Language and Linguistics and a Ph.D. from the University of Edinburgh. He is editor (with Ellen Kaisse) of the journal *Phonology*.

ELIZABETH HUME is Professor and Chair of the Department of Linguistics at The Ohio State University. She holds a Ph.D. and M.A. in Linguistics from Cornell University, an M.A. in French and Social Psychology of Language from McMaster University, and a B.A. in French from Université Laval.

KEREN RICE is University Professor and Canada Research Chair in Linguistics and Aboriginal Studies at the University of Toronto. She holds an M.A. and Ph.D. in Linguistics from the University of Toronto and a B.A. from Cornell University.

Blackwell Companions to Linguistics series

Series Editors:
Martin Everaert and Henk van Riemsdijk

The Blackwell Companion to Syntax
Edited by Martin Everaert and Henk van Riemsdijk

The Blackwell Companion to Phonology
Edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice

The Blackwell Companion to Phonology

Edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice

Volume III
Phonological Processes

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This edition first published 2011

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

The Blackwell companion to phonology / edited by Marc van Oostendorp . . . [et al.].
p. cm. — (Blackwell companions to linguistics series)
Includes bibliographical references and index.
ISBN 978-1-4051-8423-6 (hardcover : alk. paper)
1. Phonetics. 2. Grammar, Comparative and general—Phonology.
I. Oostendorp, Marc van, 1967–
P217.B53 2011
414—dc22 2010042206

A catalogue record for this book is available from the British Library.

Set in 10/12pt Palatino by Graphicraft Limited, Hong Kong

1 2011

Brief Contents

Volume I

Volume II

Volume III

Volume IV

Volume V

Full Table of Contents

Volume I

Volume II

Suprasegmental and Prosodic Phonology

Volume III

Phonological Processes

Volume IV

Phonological Interfaces

Volume V

Phonology across Languages

58 The Emergence of the Unmarked

Michael Becker
Kathryn Flack Potts

1 Introduction

The term "The Emergence of the Unmarked" (TETU), originally coined by McCarthy and Prince (1994), refers to situations where some marked structure is generally allowed in a language, but banned in particular contexts; the complementary unmarked structure thus "emerges." In Nuu-chah-nulth (Wakashan, referred to by McCarthy and Prince by its former name, Nootka), for example, syllables can generally have codas; reduplicants, however, are exceptional in that codas are banned. This results in words like [ʧ̓i-ʧ̓im.s'iːħ] 'hunting bear' and [waː-waːs.ʧ̓iɬ] 'naming where', in which unmarked (codaless) syllables emerge in reduplicants despite the presence of marked codas in bases.

TETU effects came to prominence in phonological theory with the advent of Optimality Theory (OT; Prince and Smolensky 1993). In OT terms, these effects typically follow from rankings like (1), where a markedness constraint M is dominated by a faithfulness constraint F1, which blocks M's activity in some, though crucially not all, contexts. M is free to become active in contexts where F1 isn't relevant; here, M can motivate violation of still lower-ranked faithfulness constraints (F2).

(1) F1 >> M >> F2

The Nuu-chah-nulth pattern described above results from a ranking of this type, as shown in (2) and (3). The markedness constraint NoCoda is dominated by the anti-deletion constraint IO-Max; this ranking protects underlying codas from deletion, eliminating the unmarked, codaless candidate (2b). Since reduplicants are assumed not to stand in correspondence with inputs, however (chapter 100: reduplication), high-ranking IO-Max is irrelevant in their evaluation.[1] Because NoCoda dominates BR-Max, the emergence of unmarked CV syllables is permitted

[1] Correspondence between input and output candidates is evaluated by input–output (IO) faithfulness constraints. Reduplicants stand in correspondence relationships with the output forms of their bases, and are evaluated by base-reduplicant (BR) faithfulness constraints (McCarthy and Prince 1999). Faithfulness constraints in this chapter assess IO correspondence, unless otherwise noted.

in reduplicants. Concretely, candidates (3a) and (3b) are identical, except that the reduplicant in (3b) contains a copy of the coda of the root-initial syllable, while the reduplicant in (3a) doesn't. Because NoCoda dominates BR-Max, the additional NoCoda violation in (3b) rules out this candidate in favor of the less marked (3a).

(2)

/ʧims-ʼiːh/	IO-Max	NoCoda	BR-Max
☞ a. ʧim.sʼiːħ		**	
b. ʧi.sʼiː	**!		

(3)

/red-ʧims-ʼiːħ/	IO-Max	NoCoda	BR-Max
☞ a. ʧi.ʧim.sʼiːħ		**	****
b. ʧim.ʧim.sʼiːħ		***!	***

Increasing attention to TETU effects was a natural result of inquiry into Optimality Theory. As McCarthy and Prince note, TETU is a direct result of two fundamental properties of OT. First, OT is a theory of ranked, violable constraints. Constraints are frequently active in a language even if they are not always satisfied; this is at the heart of TETU effects, which occur when a markedness constraint is dominated but still active. They observe that this "sharply differentiates OT from approaches to linguistic structure and interlinguistic variation based on parameters, rules, or other devices that see linguistic principles in globally all-or-nothing terms" (1994: 363–364).[2]

Second, distinctions between marked and unmarked structures are fundamental to OT, allowing the existence and emergence of unmarkedness to be formally defined. As McCarthy and Prince explain, "OT (Prince and Smolensky 1993) offers an approach to linguistic theory that aims to combine an empirically adequate theory of markedness with a precise formal sense of what it means to be 'unmarked'" (1994: 333). At the heart of OT are two basic constraint types: those demanding identity, typically between inputs and outputs (faithfulness), and those penalizing particular output structures (markedness) (chapter 63: markedness and faithfulness constraints; see also chapter 4: markedness). Marked structures are defined as exactly those structures which violate a markedness constraint.[3] "Emergence" can be defined with similar precision, again by reference to basic properties of OT: an unmarked structure can be said to emerge in a language if the markedness constraint violated by that structure is dominated by some (typically faithfulness) constraint which blocks its activity in some, but not all, contexts in that language.

§2 elaborates on this basic understanding of TETU as "activity despite domination," surveying three types of cases in which a dominating constraint is inactive in a particular evaluation, allowing a lower-ranked markedness constraint

[2] This view is elaborated in McCarthy (2002: 129–134), where it is noted that theories with ordered rules can mimic some TETU effects with the application of default rules.

[3] More precisely, structures which violate some markedness constraint M1 are marked with respect to M1; if these structures do not violate some other markedness constraint M2, no conflict arises in saying that they are also unmarked with respect to M2. In OT, markedness is multidimensional, assessed by each markedness constraint individually.

to emerge. §3 then describes gradient TETU effects found in languages where the emergent markedness constraint is never categorically active. Finally, §4 compares true TETU effects with situations where faithfulness, rather than markedness, constraints are active despite domination and thus emergent.

2 TETU typology

The typical TETU ranking is F1 >> M >> F2, with M emerging in evaluations where F1 is not decisive. This section will discuss three subclasses of TETU rankings, following from three different contexts in which high-ranking F1 may be rendered inactive. §2.1 looks at output segments and structures which have no input correspondents and so are invisible to IO-faithfulness constraints; these include reduplicants, epenthetic segments, and syllable boundaries. §2.2 considers evaluations in which multiple candidates tie on a particular high-ranking constraint, and §2.3 surveys faithfulness constraints which evaluate only some positions or aspects of outputs while ignoring others. In each of these situations, a high-ranking constraint is inactive and a dominated markedness constraint becomes active, choosing the winning output.

2.1 *Output segments and structures without input correspondents*

TETU is commonly observed in output structures which lack input correspondents and thus cannot be evaluated by IO-faithfulness. Recall the Nuu-chah-nulth ranking in (2) and (3), of the form IO-F >> M >> BR-F. Because reduplicants have no input correspondents in this theory, they cannot be evaluated by IO-faithfulness, allowing the effects of M (NoCoda in Nuu-chah-nulth) to emerge. This section describes similar TETU patterns found in two other structures which are present in outputs but not inputs: epenthetic segments and syllable boundaries.

2.1.1 *Epenthesis*

Kager (1999) observes that markedness constraints which are generally freely violated in a language often determine the quality of epenthetic segments (CHAPTER 67: VOWEL EPENTHESIS). These segments are typically featurally unmarked; epenthetic vowels like [i], [ɨ], and [ə], and consonants like [ʔ], [h], and glides, are cross-linguistically common, while marked segments like [f] and [æ] are rarely epenthesized.[4] This is due to TETU rankings like IO-Ident >> M, where M is a featural markedness constraint. When a constraint demanding identity between input and output features outranks markedness (here, IO-Ident >> M), the latter has little power to ban marked features in the language as a whole. While the presence of an epenthetic segment violates the anti-epenthesis constraint Dep, its lack of an input correspondent means that it is invisible to high-ranked IO-Ident; thus, epenthetic segments are subject to markedness constraints which require them to have unmarked feature values.

[4] See Vaux (2002, 2008) for a survey of epenthetic segments and a diachronic perspective. See also Steriade (2001, 2009) for the view that epenthetic segments are chosen by faithfulness constraints minimizing the perceptual distance between representations with and without the epenthetic segment.

2.1.2 *Syllable structure*

Not every aspect of linguistic outputs is evaluated by faithfulness constraints; some output properties, like prosodic structure above the mora level, are generally taken to be governed by markedness constraints only (CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE). In a language like Timugon Murut (Austronesian), where DEP >> ONSET as in (4) and (5) (McCarthy and Prince 1994), the dominated markedness constraint ONSET emerges to make decisions in cases where DEP cannot distinguish between candidates. DEP's high ranking results in a language where epenthesis never occurs in order to avoid onsetless syllables, thus allowing words like [ambi'luo] 'soul' in (4). DEP (and similarly MAX, IDENT, etc.) cannot, however, distinguish between the candidates in (5), which differ only in syllabification. Because faithfulness constraints cannot see these differences, the decision is handed down to the emergent markedness constraint ONSET.

(4)

/ambi'luo/	DEP	ONSET
☞ a. am.bi.'lu.o		**
b. ʔam.bi.'lu.ʔo	**!	

(5)

/ambi'luo/	DEP	ONSET
☞ a. am.bi.'lu.o		**
b. am.bil.'u.o		***!

Cross-linguistically, the markedness constraint ONSET commonly triggers epenthesis, deletion, and other changes to prevent onsetless syllables. But its effects can also emerge even in languages like Timugon Murut, where ONSET is crucially dominated and so cannot require all syllables to have onsets; here, ONSET nonetheless requires syllabification of available consonants as onsets rather than codas. This contrasts with a parameter-based view of phonology, where onsetless syllables are present only when the ONSET parameter is "off," and thus cannot affect syllabification in any way.

2.2 *Output candidates not distinguished by dominating constraints*

Unmarkedness can also emerge when multiple output candidates are evaluated identically by all constraints dominating the emergent markedness constraint. This section discusses allomorph selection, which has been traditionally analyzed as a TETU effect of this sort within OT, as well as a similar example from the syntax–phonology interface.

2.2.1 *Allomorphy*

Mascaró (2004) observes that when a morpheme has multiple underlying forms, Gen supplies candidates that vary in the forms they correspond to (CHAPTER 99: PHONOLOGICALLY CONDITIONED ALLOMORPH SELECTION). In cases like English *a/an*, where the indefinite article has two lexically listed allomorphs, some members of the candidate set stand in correspondence with underlying *a*, while others stand in correspondence with underlying *an*. For this reason, the two output candidates shown in (6), *a wug* and **an wug*, tie on all high-ranked IO-faithfulness constraints. While the ranking of faithfulness constraints (here, simply FAITH) above NOCODA generally permits codas throughout English, NOCODA nevertheless emerges as decisive here, ruling out **an wug* in this unique case where multiple possible outputs are equally faithful to their respective inputs.[5]

(6)

{*a, an*} /wʌg/	FAITH	NOCODA
☞ a. ə.wʌg		*
b. ən.wʌg		**!

A more complex TETU analysis of lexically specific allomorph selection is offered in Becker's (2009) discussion of the Turkish aorist (Lees 1961; Napikoğlu and Ketrez 2006). The aorist suffix has two allomorphs: /-Ir/, with a high vowel, is used after all polysyllabic roots; /-Er/, with a non-high vowel, is used after all monosyllabic obstruent-final roots (the backness and height of these vowels are determined by vowel harmony; CHAPTER 118: TURKISH VOWEL HARMONY). Monosyllabic sonorant-final roots allow lexical exceptions: some take /-Ir/, while others take /-Er/.

(7)

shape of stem	*affix*					
polysyllabic	-Ir	[gere'k-ir]	'need'	[ʧali'ʃ-ir]	'work'
obstruent-final monosyllabic	-Er	[sa't-ar]	'sell'	[œ'p-er]	'kiss'
{r l n}-final monosyllabic	-Ir	[ka'l-ir]	'stay'	[gœ'r-yr]	'see'
	-Er	[da'l-ar]	'dive'	[œ'r-er]	'knit'

Turkish vowels are typically faithful to their underlying height specification, both in roots and in affixes; for example, the affix /-E/ (DATIVE) (e.g. [je're] 'to the place') contrasts with the affix /-I/ (3SG POSS) (e.g. [je'ri] 'his/her place'). This indicates that IDENT[high] outranks both of the markedness constraints in (8). When two allomorphs are available to choose from, however, as in these aorist examples, IDENT[high] is satisfied regardless of the choice of allomorph; the markedness constraints can thus emerge as decisive. (9) illustrates how *'σ/high consistently selects the /-Er/ allomorph in monosyllabic obstruent-final roots.

(8) a. *'σ/high
No stressed high vowels.
b. *RER
No non-high vowels between sonorants.

(9)

/sat-{-Er, -Ir}/	IDENT[high]	*'σ/high	*RER
☞ a. sa't-ar			
b. sa't-ir		*!	

[5] The ranking FAITH >> ONSET similarly chooses [ən.ʌg] over *[ə.ʌg].

The situation is more complex in the monosyllabic sonorant-final roots, which don't behave uniformly. Some of these occur with /-Er/, violating *RER, as shown in (10), while others occur with /-Ir/, violating *'σ/high, as in (11). Becker argues that sonorant-final monosyllabic roots are linked to lexically specified constraint rankings: for /-Er/-selecting roots like /dal/, *'σ/high >> *RER, while the opposite ranking holds for /-Ir/-selecting roots like /kal/. The overall pattern is one where each markedness constraint is emergent for a particular class of roots. See Becker (2009) for further details of the analysis, including the treatment of polysyllables and mechanisms for learning both affix URs and lexically specific rankings.

(10)

/dal-{-Er, -Ir}/	Ident[high]	*'σ/high	*RER
☞ a. da'l-ar			*
b. da'l-ɨr		*!	

(11)

/kal-{-Er, -Ir}/	Ident[high]	*RER	*'σ/high
a. ka'l-ar		*!	
☞ b. ka'l-ɨr			*

This TETU analysis of the Turkish aorist accounts for the fact that the lexically specific distribution of this affix is limited to sonorant-final roots. Since *RER is ranked below Ident[high], its effect can only be observed when a root contributes one sonorant and one of two lexically listed allomorphs (here, of the aorist affix) contributes the other. This contrasts with a diacritic-based approach to exceptionality; since such an approach isn't based on markedness constraints, it runs the risk of missing phonological restrictions on the distribution of exceptions. See Gouskova (2010) for further arguments in favor of a grammar-based approach to exceptionality (also chapter 106: exceptionality).

Rankings in each of these allomorphy examples take the form F >> M, where F cannot distinguish between candidates containing different allomorphs. Because multiple candidates are equally faithful, satisfying M does not require violating a lower-ranked F2, as is required in the prototypical TETU cases discussed in §1. The Turkish example shows that satisfying an emergent markedness constraint can also require violating a lower-ranked markedness constraint, in a ranking like F >> M1 >> M2. This occurs because markedness constraints can conflict with each other, as well as with faithfulness constraints. The following discussion of phonological phrasing and the syntax–phonology interface carries this observation further, demonstrating that markedness constraints can also emerge in contexts where a higher-ranked, conflicting markedness constraint is inactive.

2.2.2 *Phonological phrasing*

Because faithfulness constraints do not evaluate prosodic structure, analyses of phonological phrasing are generally based on rankings of conflicting markedness constraints. While most familiar TETU rankings involve domination by a conflicting faithfulness constraint, the dominating constraint may also be a second markedness constraint. That is, dominated M2 may also emerge in a ranking like (12).

(12) M1 >> M2

Truckenbrodt (1999) proposes an analysis of phonological phrasing based only on markedness constraints; in contexts where high-ranking markedness constraints are rendered inactive, lower-ranking markedness constraints emerge. In Chewa (Bantu, referred to by Truckenbrodt as Chicheŵa), a complex VP like $[V\ NP\ NP]_{VP}$ is produced as single phonological phrase, $(V\ NP\ NP)_{PhP}$, rather than $^*(V\ NP)_{PhP}$ $(NP)_{PhP}$, as in (13). The large phrase satisfies WRAP-XP, a constraint that penalizes any syntactic phrase whose elements are parsed into smaller phonological phrases. Other dominated markedness constraints express conflicting preferences for smaller phonological phrases: ALIGN-XP demands alignment of the right edge of each syntactic phrase with the right edge of a corresponding phonological phrase. The winner in (13) incurs a violation due to the first NP, which has no phrase break at its right edge.

The ranking WRAP-XP >> ALIGN-XP generally thwarts ALIGN-XP's desire for additional phonological phrases. ALIGN-XP's effects emerge, however, under focus. ALIGN-FOC requires focused verbs to fall at the end of phonological phrases; the ranking ALIGN-FOC >> WRAP-XP rules out candidate (14a). Candidates (14b) and (14c) both satisfy ALIGN-FOC, and both violate WRAP-XP, rendering WRAP-XP inactive as well in selecting the optimal output. Because these candidates tie on high-ranking constraints, we again see a TETU effect: ALIGN-XP emerges, selecting the unmarked candidate, (14c).

(13)

$/[V\ NP\ NP]_{VP}/$	ALIGN-FOC	WRAP-XP	ALIGN-XP
☞ a. $(V\ NP\ NP)_{PhP}$			*
b. $(V\ NP)_{PhP}\ (NP)_{PhP}$		*!	

(14)

$/[V_{FOC}\ NP\ NP]_{VP}/$	ALIGN-FOC	WRAP-XP	ALIGN-XP
a. $(V_{FOC}\ NP\ NP)_{PhP}$	*!		*
b. $(V_{FOC})_{PhP}\ (NP\ NP)_{PhP}$		*	*!
☞ c. $(V_{FOC})_{PhP}\ (NP)_{PhP}\ (NP)_{PhP}$		*	

Truckenbrodt notes that this analysis of Chewa is particularly interesting, due to the non-local nature of the TETU effect: the appearance of an (unmarked) prosodic break after the focused verb causes another break to appear after a subsequent non-focused noun phrase.

2.3 *Output segments not evaluated by specific faithfulness*

This final subsection discusses situations where general IO-faithfulness is low-ranked, and the emerging markedness constraint is instead dominated by a different type of faithfulness constraint. In other words, these are TETU rankings of the type Special-F >> M >> General-F. We discuss three kinds of faithfulness that can outrank general IO-faithfulness: positional faithfulness, which protects strong positions inside a candidate; output-output faithfulness, which protects the base in a morphologically complex form; and USELISTED, which protects correspondents of existing forms in a speaker's lexicon.

2.3.1 *Positional faithfulness*

Beckman (1999) examines patterns where contrasts are licensed only in strong positions like initial syllables, stressed syllables, and onsets. She analyzes these using positional faithfulness constraints, which assess correspondence only for segments in particular output positions (here, onsets).

Catalan (Romance) allows contrastive voicing in onsets, but bans voiced obstruents in codas (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). Beckman accounts for this with the ranking shown in (15)–(17). Underlyingly voiced coda obstruents are devoiced in surface forms, due to the ranking *VoiObs >> Ident[voice], as in (15). In onsets, however, underlying voicing surfaces faithfully, due to the high-ranking positional faithfulness constraint Ident[voice]/Onset, as in (16)–(17). Here the markedness constraint *VoiObs is dominated, yet active in non-onset contexts, making this a TETU effect.

(15)

/griz/ 'gray (MASC)'	Ident[voice]/Onset	*VoiObs	Ident[voice]
a. 'griz		**!	
☞ b. 'gris		*	*

(16)

/gos-a/ 'dog (FEM)'	Ident[voice]/Onset	*VoiObs	Ident[voice]
a. 'go.zə	*!	**	*
☞ b. 'go.sə		*	

(17)

/griz-a/ 'gray (FEM)'	Ident[voice]/Onset	*VoiObs	Ident[voice]
☞ a. 'gri.zə		**	
b. 'gri.sə	*!	*	*

In addition to protecting phonologically strong positions (initial syllable, stressed syllables, onsets), positional faithfulness may also protect morphologically strong contexts such as roots (McCarthy and Prince 1995) and nouns (Smith 1999, 2001; see also CHAPTER 102: CATEGORY-SPECIFIC EFFECTS). Smith notes that in Spanish, stress is lexically marked in nouns but predictable in verbs. She analyzes a range of such patterns using noun-specific faithfulness constraints (F/Noun) in the ranking schema F/Noun >> M >> F. Here, nouns may be faithful to lexically specified stress thanks to high-ranking F/Noun; in verbs, however, stress is instead governed by emergent markedness constraints.

The activity of markedness constraints in all of these rankings, despite their domination by a conflicting (here position-specific) faithfulness constraint, identifies these as TETU effects. There is a significant difference, however, between many positional faithfulness patterns and most other TETU patterns. In the TETU ranking schemata discussed in previous sections, general IO-faithfulness constraints dominate emergent markedness constraints. Here, though, markedness dominates (general, though not position-specific) IO-faithfulness. This results in different surface distributions of the emergent unmarked structures.

Typically, when IO-F >> M, marked structures are licensed in most contexts throughout the language; unmarkedness emerges in specific, less frequent contexts like reduplicants, epenthetic segments, or allomorphs. When TETU results from high-ranking positional faithfulness, however, the reversed ranking M >> IO-F can result in a language which is largely *unmarked*; in these languages, marked structures are restricted to the specific set of contexts protected by the positional faithfulness constraint. When this set of markedness licensing contexts is small, as for FAITH/σ1 or FAITH/'σ (faithfulness to word-initial and stressed syllables, respectively), unmarked structures are required in the majority of contexts: the set of positions in which markedness may occur is atypically smaller than those where unmarkedness is required.[6] This distributional pattern will be discussed further in §4.

2.3.2 *Output–output faithfulness*

Another family of constraints which evaluates only some outputs and so gives rise to TETU effects is output-output (OO) faithfulness (Benua 1997). OO-faithfulness constraints evaluate correspondence between the base of a morphologically complex word and that base's stand-alone surface form. Harris (1990) discusses examples of Aitken's Law in dialects of the Central Scottish Lowlands. Here, stressed vowels in roots have predictable length: when followed by any consonant other than /r v ð z/, vowels are short; otherwise, they are long. /ɪ ʌ/ are exceptions, remaining short in all positions. (See also CHAPTER 20: THE REPRESENTATION OF VOWEL LENGTH.) For example, stop-final *feed* has a short vowel, while the open syllable *key* has a long vowel. The past tense *keyed*, however, keeps the long vowel which is present in its base *key*, despite its final stop coda. This can be attributed to protection from high-ranking OO-FAITH, as described below (Benua 1997; McCarthy 2002).

Because OO-faithfulness constraints target only a subset of a language's output forms – those which are morphologically complex – they can give rise to TETU effects. A ranking like OO-F >> M >> IO-F operates much like the positional faithfulness TETU ranking discussed above. The markedness constraint *V:C] ("no long vowels in syllables closed by any consonant other than /r v ð z/") dominates IO-faithfulness; tableau (18) shows that this results in a language which is typically unmarked: long vowels are absent from closed syllables. Long vowels appear in open syllables, as in (19); because OO-IDENT(length) >> *V:C], long vowels also appear in closed syllables in morphologically complex forms derived from roots with long vowels, as in *keyed* [ki:d] (cf. *key* [ki:]) in (20).

(18)

/fi:d/	OO-ID(length)	V:C]	IO-ID(length)
☞ a. fid			*
b. fi:d		*!	

(19)

/ki:/	OO-ID(length)	V:C]	IO-ID(length)
a. ki			*!
☞ b. ki:			

[6] Positional faithfulness TETU rankings can also result in languages where, as is more typical of TETU, unmarkedness is the less frequent pattern; this occurs when the positional faithfulness constraint targets a broad set of positions, e.g. FAITH/Root.

(20)

/kiː-d/	OO-ID(length)	V:C]	IO-ID(length)
a. kid	*!		*
☞ b. kiːd		*	

Here, again, a markedness constraint is active in the language despite its domination by (here, OO) faithfulness. Similar TETU effects are possible in other theories that use faithfulness relations between members of a paradigm, such as McCarthy's (2005) Optimal Paradigms.

2.3.3 UseListed

Zuraw (2000) proposes an additional novel kind of faithfulness constraint, UseListed, which protects items that are listed in a speaker's lexicon. Listed items include all roots and all morphologically complex forms that a particular speaker has heard, with more frequent items assumed to be more strongly listed.

In producing a previously heard, morphologically complex form, the speaker has two options: they could use either the lexically listed forms of the root and affixes as inputs to the grammar, or they could instead use the single lexically listed complex form (again, as input to the grammar). Zuraw proposes that these two possible input structures compete in a single evaluation, with UseListed penalizing outputs derived from productive combinations of morphemes.

For novel roots and novel complex forms (i.e. novel combinations of roots and affixes, even if a speaker is familiar with each morpheme in other contexts), however, no lexical listing is available. Thus outputs based on any of these forms will violate UseListed equally. Markedness constraints ranked below UseListed can therefore emerge in evaluations of unfamiliar items, as in Hayes and Londe's (2006) analysis of Hungarian vowel harmony.

The Hungarian dative appears with a back vowel when the root's final syllable has a back vowel ([glykoːz-nɔk] 'glucose-DAT'), and it appears with a front vowel when the root's final syllable has a front rounded vowel ([ʃoføːr-nɛk] 'chauffeur-DAT') (CHAPTER 123: HUNGARIAN VOWEL HARMONY). When the root's final syllable has a front unrounded vowel, some items take a front suffix ([tsiːm-nɛk] 'address-DAT') and others take a back suffix ([hiːd-nɔk] 'bridge-DAT'). Taking a back suffix is especially likely when the final front unrounded vowel is preceded by a back vowel ([aːtseːl-nɔk] 'steel-DAT'). Here, the relevant markedness constraints will be Local[eː], which penalizes back vowels in the syllable immediately following an [eː], and Distal[back], which penalizes front vowels in any syllable following a back vowel.

Hungarian speakers agree on the dative forms of familiar (lexically listed) items such as [aːtseːl-nɔk]. UseListed is decisive in these cases, preferring the listed form over productive combinations of the root and the suffix, and thus rendering lower-ranked markedness constraints on vowel harmony inactive. The two candidates (21a) and (21b) are generated from the listed form [aːtseːl-nɔk], and thus satisfy UseListed (despite the unfaithful surface form of this input in (21b)). The second two candidates are generated productively by combining the root /aːtseːl/ with the dative suffix, and are thus ruled out by UseListed.

(21)

/aːtseːl-{nɛk, nɔk}/, *listed*: [aːtseːl-nɔk]	USE LISTED	IDENT [back]	LOCAL [eː]	DISTAL [back]
☞ a. /aːtseːl-nɔk/ → aːtseːl-nɔk			*	*
b. /aːtseːl-nɔk/ → aːtseːl-nɛk		*!		**
c. /aːtseːl-{nɛk, nɔk}/ → aːtseːl-nɔk	*!		*	*
d. /aːtseːl-{nɛk, nɔk}/ → aːtseːl-nɛk	*!			**

But when Hungarian speakers hear a novel root containing a back vowel followed by a front unrounded vowel, e.g. [haːdeːl], the suffix vowel in the dative forms can agree with either root vowel: some speakers prefer [haːdeːl-nɔk], as in (22a), while others prefer [haːdeːl-nɛk], as in (22b). Both candidates in (22) violate UseListed, since no lexical listing exists for this novel item, and thus the dative form must be derived productively by combining the root /haːdeːl/ with the dative suffix.

(22)

/haːdeːl-{nɛk, nɔk}/, *listed*: [ ]	USE LISTED	IDENT [back]	LOCAL [eː]	DISTAL [back]
☞ a. /haːdeːl-{nɛk, nɔk}/ → haːdeːl-nɔk	*		*	*
☞ b. /haːdeːl-{nɛk, nɔk}/ → haːdeːl-nɛk	*			**

Hayes and Londe argue that a particular speaker's actual output depends on a stochastic ranking between the two competing markedness constraints on vowel harmony, Local[eː] and Distal[back]. Crucially, as in the Turkish example in §2.2.1, one of these two low-ranked markedness constraints emerges; here, this occurs when dominating UseListed cannot distinguish between candidate outputs for a novel input.

3 Gradient TETU

The previous sections have surveyed various ways in which high-ranking constraints can be rendered irrelevant in particular evaluations, allowing lower-ranked markedness constraints to emerge. Of course, not all markedness constraints which are dominated in a particular language emerge; many are ranked too low to ever be active in choosing a winning surface form. Recent work suggests, however, that subtle TETU effects can be identified even for markedness constraints which never distinguish between grammatical and ungrammatical forms.

Consonants in Arabic roots are subject to various co-occurrence restrictions (chapter 86: morpheme structure constraints). Among grammatical consonant combinations, preferences for particular combinations are found: some are much more frequent than others in the lexicon, and novel words conforming to the more frequent patterns are judged as more well-formed in rating tasks (chapter 90: frequency effects). Coetzee and Pater's (2008) analysis of these preferences among grammatical forms casts them as gradient TETU effects, where markedness constraints which are never categorically obeyed nevertheless exert subtle preferences for unmarked forms (chapter 89: gradience and categoricality in

PHONOLOGICAL THEORY). Rankings giving rise to these gradient effects are illustrated in (23) and (24). Roots including both coronal stops and fricatives (e.g. /dasar/ 'to push', represented here as TS) and those containing coronal stops and sonorants (e.g. /dalaq/ 'to spill', represented here as TL) both surface faithfully in Arabic, although these combinations are underrepresented, i.e. they are attested less often than expected, given the overall frequency of each type of consonant. In the lexicon, however, TS roots are more severely underrepresented than TL roots, suggesting that TL roots are in some sense more easily tolerated. Coetzee and Pater argue that both combinations violate a constraint against roots with two coronals (*TT), while only the dispreferred TS roots violate an additional constraint against roots with two coronals of similar sonority (*TT[son]).

(23)

/TS/	IDENT(place)	*TT[son]	*TT
☞ a. TS		*	*
b. PS	*!		

(24)

/TL/	IDENT(place)	*TT[son]	*TT
☞ a. TL			*
b. PL	*!		

Here, no markedness constraint is ranked highly enough to ban TS or TL outputs. These consonant combinations are protected by faithfulness, and are thus attested and grammatical, but they are not judged by speakers to be quite as well formed as roots that lack OCP violations. TS's additional violation of *TT[son] contributes to its decreased acceptability relative to TL, as observed in the results of word-likeness tasks and similar psycholinguistic experiments. In other words, the markedness constraint *TT[son] is active in Arabic even though it is crucially dominated by IDENT(place). This activity is evidenced by the underattestation of actual TS roots and the decreased acceptability of novel TS roots, even though it doesn't force unfaithful mappings.

The incorporation of gradient generalizations into the grammar can also be used to identify relative rankings of undominated markedness constraints, i.e. the opposite of gradient TETU. If neither of two markedness constraints is ever crucially dominated by a conflicting constraint in some language, the relative ranking of these constraints cannot be determined from either categorical phonotactics or paradigmatic information. This approach, however, allows evidence for their relative ranking to come from gradient phonotactics and psycholinguistic data. Coetzee (2009) compares the grammaticality of English homorganic stops after [s], noting that coronals are attested, as in *state*, but labials and dorsals are not, as in **skake* or **spape* (see also Davis 1984, 1991; Frisch 1996; Frisch *et al.* 2004).

Coetzee's psycholinguistic experiments show that speakers rate **spape* as less acceptable than **skake*, and both are less acceptable than *state*. He uses this result to propose that while *sPVP and *sKVK are both undominated in English, the constraint penalizing *sPVP is more highly ranked than the constraint penalizing *sKVK. This view is also supported by the existence of words that come close to violating *sᴋVᴋ, such as *skag*, *skulk*, or *squeak*, compared with the non-existence of **spab*, **spulp*, or **spweep*.

4 The emergence of the faithful

As mentioned in §2.3.1, there is an important distinction between the formal definition of TETU and the most intuitive surface-oriented descriptions of these patterns. McCarthy and Prince (1994: 334) define TETU as follows:

> Even in languages where [some markedness constraint] C is crucially dominated and therefore violated, the effects of C can still be observed under conditions where the dominating constraint is not relevant. . . . this [is] "emergence of the unmarked."

The same passage describes the typical surface pattern that results from constraint activity despite domination:

> in the language as a whole, C may be roundly violated, but in a particular domain it is obeyed exactly. In that particular domain, the structure unmarked with respect to C emerges.

Patterns where high-ranking positional faithfulness constraints allow unmarkedness to emerge (e.g. Iᴅᴇɴᴛ[voice]/Oɴsᴇᴛ >> *VᴏɪOʙs >> Iᴅᴇɴᴛ[voice]) demonstrate that activity-despite-domination rankings can also give rise to a converse pattern: a markedness constraint may in fact be obeyed in the language as a whole, but violated in a particular domain. In these cases, the structure unmarked with respect to the markedness constraint emerges in the language as a whole, despite its ungrammaticality in a particular domain.

The lack of a necessary connection between a markedness constraint's activity despite domination and the relative rarity of the resulting unmarkedness is also illustrated in patterns following from the activity of positional markedness constraints. Like positional faithfulness constraints, these are versions of regularly attested markedness constraints which evaluate only structures in particular output positions, e.g. Oɴsᴇᴛ/σ1, a constraint that penalizes onsetlessness in the initial syllable only (ᴄʜᴀᴘᴛᴇʀ 55: ᴏɴsᴇᴛs).

An example of this pattern comes from Arapaho (Algonquian, Smith 2002: 127, from Salzman 1956: 53–54). In this language, onsetless syllables occur in non-initial syllables (e.g. the onsetless third syllable in [wo.ˈʔo.uː.soː] 'kitten'), as shown in (25). Word-initial vowels are, however, banned (e.g. *[o.toʔ]), as shown in (26). These patterns follow from the ranking Oɴsᴇᴛ/σ1 >> Dᴇᴘ >> Oɴsᴇᴛ, and are identical in character to the typical TETU surface pattern: marked structures are licensed in most of the language, but a small pocket of enforced unmarkedness is found in initial syllables. Many other languages of this type are discussed by Smith (2002) and Flack (2009).

(25)

/woˈʔouːsoː/	Oɴsᴇᴛ/σ1	Dᴇᴘ	Oɴsᴇᴛ
☞ a. wo.ˈʔo.uː.soː			*
b. wo.ˈʔo.ʔuː.soː		*!	

(26)

/oto?/	Onset/σ1	Dep	Onset
☞ a. ho.to?		*	
b. o.to?	*!		*

Despite the surface similarities between positional markedness patterns and classic TETU patterns, the formal structure of these rankings distinguishes them from TETU rankings. These follow the schematic form M1 >> F >> M2, rather than the TETU form F1 >> M >> F2. These patterns thus might be dubbed "The Emergence of the Faithful": F emerges (i.e. is active despite domination) in cases where dominating Onset/σ1 is inactive.

Albright (2004) discusses patterns of this sort, using the term "The Emergence of the Marked" to describe their surface pattern. In Lakota (Siouan), codas are banned in roots but licensed elsewhere (e.g. affixes, reduplicants, function words). This pattern results from the positional markedness ranking NoCoda-root >> F >> NoCoda; as Albright explains, this pattern is a "mirror image of the TETU configuration: here, greater faithfulness emerges outside roots, when a higher-ranked markedness constraint (NoCoda-root) is inapplicable" (2004: 7). Here, markedness "emerges" in distributionally rare root-external contexts. To be clear, the distributional sense of "emerge" used here is different from the formal sense used by McCarthy and Prince: formally, effects of the constraint which is dominated yet active emerges; distributionally, whichever pattern is not generally permitted (markedness *vs.* unmarkedness) "emerges" in specific, restricted contexts.

5 Conclusion

TETU is a property of theories with violable constraints, and sets these theories apart from those with parameters or inviolable constraints. In TETU rankings, a markedness constraint is shown to be dominated in a language, yet active in situations where the dominating constraints are irrelevant. Three types of such situations are surveyed in §2. Active-yet-dominated markedness constraints have also been used in the analysis of gradient patterns, as discussed in §3. Finally, patterns mirroring TETU which result from active-yet-dominated faithfulness constraints are discussed in §4.

TETU effects, which demonstrate the violability of OT constraints, set OT apart from theories with inviolable constraints, also known as parameters in Principles and Parameters Theory (Chomsky 1981, 1986). In Principles and Parameters Theory, the learner starts with parameters set to their default, or unmarked, position; parameters can be switched off given evidence from the ambient language. The NoCoda parameter, for instance, will remain on for a speaker of Hawaiian, as this language doesn't allow codas. However, speakers of English or Nuu-chah-nulth will switch the NoCoda parameter off, as codas are generally allowed in these languages. Once off, however, the NoCoda parameter can no longer be used to account for the contexts in which these two languages prefer codaless syllables (see §1 and §2.2.1 above), causing a loss of generality in the analysis of these languages (McCarthy 2002: 131–132).

Interest in TETU effects initially brought attention to a variety of cases where constraints were shown to be active even in languages where they were roundly violated, e.g. NoCoda in English. This lent support to the view that there is a single, universal constraint set for all languages, which in turn led to fruitful research on how language-specific rankings of these universal constraints could be learned (see e.g. Tesar 1995; Tesar and Smolensky 1998; and much work since).

Early work in OT typically assumed that this universal constraint set was innate; assumptions of both innateness and constraint universality have begun to lose favor in recent years with the advent of proposals that some or all constraints are induced by learners (Flack 2007; Hayes and Wilson 2008; Moreton 2010).

TETU effects were a major focus of interest in the early days of Optimality Theory, when the concept of violable constraints was new to the linguistic community. With the increased acceptance of violable constraints in theoretical work, cases of TETU no longer attract special attention, even as interest turns to other theories that incorporate violable constraints, including OT-CC (McCarthy 2007), Harmonic Grammar (Legendre *et al.* 1990; Pater 2009), and MaxEnt (Goldwater and Johnson 2003; Hayes and Wilson 2008).

REFERENCES

Albright, Adam. 2004. The emergence of the marked: Root-domain markedness in Lakhota. Handout of paper presented at the 78th Annual Meeting of the Linguistic Society of America, Boston.

Becker, Michael. 2009. Phonological trends in the lexicon: The role of constraints. Ph.D. dissertation, University of Massachusetts, Amherst.

Beckman, Jill N. 1999. *Positional faithfulness: An optimality theoretic treatment of phonological asymmetries*. New York: Garland.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Ph.D. dissertation, University of Massachusetts, Amherst.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Coetzee, Andries W. 2009. Grammar is both categorical and gradient. In Steve Parker (ed.) *Phonological argumentation*, 9–42. London: Equinox.

Coetzee, Andries W. & Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26. 289–337.

Davis, Stuart. 1984. Some implications of onset–coda constraints for syllable phonology. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 20. 46–51.

Davis, Stuart. 1991. Coronals and the phonotactics of nonadjacent consonants in English. In Carole Paradis & Jean-François Prunet (eds.) *The special status of coronals: Internal and external evidence*, 49–60. San Diego: Academic Press.

Flack, Kathryn. 2007. The sources of phonological markedness. Ph.D. dissertation, University of Massachusetts, Amherst.

Flack, Kathryn. 2009. Constraints on onsets and codas of words and phrases. *Phonology* 26. 269–302.

Frisch, Stefan A. 1996. Frequency and similarity in phonology. Ph.D. dissertation, Northwestern University.

Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22. 179–228.

Goldwater, Sharon & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson & Osten Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–120. Stockholm: Stockholm University.

Gouskova, Maria. 2010. Unexceptional segments. Unpublished ms., New York University.
Harris, John. 1990. Derived phonological contrasts. In Susan Ramsaran (ed.) *Studies in the pronunciation of English: A commemorative volume in honour of A. C. Gimson*, 87–105. London: Routledge.
Hayes, Bruce & Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104.
Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440.
Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.
Lees, Robert B. 1961. *The phonology of Modern Standard Turkish*. Bloomington: Indiana University.
Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky. 1990. Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness – an application. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, 884–891. Mahwah, NJ: Lawrence Erlbaum.
Mascaró, Joan. 2004. External allomorphy as emergence of the unmarked. In John J. McCarthy (ed.) *Optimality Theory in phonology: A reader*, 513–522. Malden, MA & Oxford: Blackwell.
McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.
McCarthy, John J. 2005. Optimal paradigms. In Laura J. Downing, T. A. Hall & Renate Raffelsiefen (eds.) *Paradigms in phonological theory*, 170–210. Oxford: Oxford University Press.
McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.
McCarthy, John J. & Alan Prince. 1994. The emergence of the unmarked: Optimality in prosodic morphology. *Papers from the Annual Meeting of the North East Linguistic Society* 24. 333–379.
McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.
McCarthy, John J. & Alan Prince. 1999. Faithfulness and identity in prosodic morphology. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*, 218–309. Cambridge: Cambridge University Press.
Moreton, Elliott. 2010. Connecting paradigmatic and syntagmatic simplicity bias in phonotactic learning. Paper presented at MIT, April 2010.
Napikoğlu, Mine & Nihan Ketrez. 2006. Children's overregularizations and irregularizations of the Turkish aorist. In David Bamman, Tatiana Magnitskaia & Colleen Zaller (eds.) *Proceedings of the 30th Annual Boston University Conference on Language Development*, vol. 2, 399–410. Somerville, MA: Cascadilla Press.
Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33. 999–1035.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.
Salzman, Zdenek. 1956. Arapaho I: Phonology. *International Journal of American Linguistics* 22. 49–56.
Smith, Jennifer L. 1999. Noun faithfulness and accent in Fukuoka Japanese. *Proceedings of the West Coast Conference on Formal Linguistics* 18. 519–531.
Smith, Jennifer L. 2001. Lexical category and phonological contrast. In Robert Kirchner, Joe Pater & Wolf Wikely (eds.) *Papers in experimental and theoretical linguistics 6: Workshop on the Lexicon in Phonetics and Phonology*, 61–72. Edmonton: University of Alberta.

Smith, Jennifer L. 2002. Phonological augmentation in prominent positions. Ph.D. dissertation, University of Massachusetts, Amherst.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 219–250. San Diego: Academic Press.

Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Studies in honor of Paul Kiparsky*, 151–179. Cambridge, MA: MIT Press.

Tesar, Bruce. 1995. Computational Optimality Theory. Ph.D. dissertation, University of Colorado, Boulder.

Tesar, Bruce & Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29. 229–268.

Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30. 219–255.

Vaux, Bert. 2002. Consonant epenthesis and the problem of unnatural phonology. Handout from paper presented at Yale University.

Vaux, Bert. 2008. Why the phonological component must be serial and rule-based. In Bert Vaux & Andrew Nevins (eds.) *Rules, constraints, and phonological phenomena*, 20–60. Oxford: Oxford University Press.

Zuraw, Kie. 2000. Patterned exceptions in phonology. Ph.D. dissertation, University of California, Los Angeles.

59 Metathesis

Eugene Buckley

The term *metathesis* – Greek for 'transposition' – refers to a reordering of segments. This chapter outlines the range of phenomena that fall under this description, and theoretical perspectives on their insightful analysis. Other cross-linguistic surveys of this topic include Webb (1974), Ultan (1978), Hock (1985), Wanner (1989), Blevins and Garrett (1998, 2004), Becker (2000), and Hume (2001, 2004).

The term has traditionally been best known for the description of historical sound changes (CHAPTER 93: SOUND CHANGE), often described as sporadic. For example, Osthoff and Brugmann (1878: xiv, n. 1) cite metathesis, along with dissimilation (CHAPTER 60: DISSIMILATION), as lacking the "mechanical" character of regular sound change. Hock (1985), however, argues that diachronic metathesis is regular when it serves to enforce a structural constraint. For example, in early attestations of Persian, as well as in reconstructed forms, clusters of an obstruent or nasal plus a liquid can be found before a final vowel. Loss of that vowel leads to a final cluster with a rising sonority profile (CHAPTER 49: SONORITY; CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS); this configuration is repaired by metathesis of the two consonants, so that the more sonorous liquid is closer to the vowel. (The segments involved are underlined in (1).)

(1) *Persian liquid metathesis* (Hock 1985: 534)

suxra	>	surx	'red'
vafra	>	barf	'snow, ice'
asru	>	ars	'tear'
*namra	>	narm	'soft'

Although much of the literature discusses historical metathesis – where copious examples can be found – this chapter focuses on instances of metathesis that are active synchronically. By this I mean alternations in the ordering of segments that appear to be part of a speaker's productive grammatical knowledge, and therefore must be accounted for in theories of linguistic competence. There is of course an intimate connection between diachronic metathesis and the synchronic alternations that may persist in the grammar as a result, but I will take care to distinguish examples for which only diachronic change is well attested, and where the results of the change appear to be new underlying forms rather than a new phonological alternation. Similarly, although the emphasis is on phonologically defined patterns, some types of metathesis require reference to morphological context, even if the specific change is expressed in terms of phonological categories.

For most of the twentieth century, metathesis was described either in prose or, as formalisms became more sophisticated, as reorderings of indexed objects in a string. Chomsky and Halle (1968: 361) describe metathesis as "a perfectly common phonological process," and permit transformations that effect permutation. In their notation, /skt/ → [kst] metathesis in Faroese, shown below in (3), could be expressed as follows.

(2) *Metathesis as a transformation*

Structural description	s	k	t				
Structural change	1	2	3	→	2	1	3

The need for indexation distinguishes metathesis from most other processes, such as insertion (CHAPTER 67: VOWEL EPENTHESIS), deletion (CHAPTER 68: DELETION), and featural assimilation (CHAPTER 81: LOCAL ASSIMILATION). In those sorts of changes, whatever elements of the representation remain after the change maintain their relative ordering on their tier. A true featural equivalent to segmental metathesis would be a swap in feature values on the same tier (or at some non-root node), such as a change from LH to HL tone in a context where underspecification of the L with simple shift of H is not a plausible analysis. As noted in §1.4, there is limited evidence for tonal metathesis of this sort.

Following the most common modern usage, in this chapter I apply the term *metathesis* to permutations of segments regardless of intervening material. §1 deals with *local* metathesis, including the sequences CC, CV, and VV, followed by brief consideration of other types. §2 considers the *long-distance* metathesis of non-adjacent segments, as well as the displacement of a segment that is not exchanged with another. §3 considers the relation of metathesis to other phenomena with which it shares some formal properties, such as infixation.

1 Local metathesis

In local metathesis, two adjacent segments are swapped, without any necessary change in their features, although in some cases other processes may affect the outcome. These can be classified formally according to the segments involved in the reversal: two consonants, a consonant and a vowel (in either order), or two vowels.

1.1 CC metathesis

To organize this presentation, I group the processes according to the features of the segments involved. These include the special role of sibilants, place of articulation, and manner of articulation.

1.1.1 *Sibilants*

Sibilant consonants are often observed to reverse order with an adjacent stop consonant (Silva 1973; Hume 2001: 12–14; Seo and Hume 2001; Steriade 2001: 234f.; Blevins and Garrett 2004: 139f.; Hume and Seo 2004: 36–39). An example is found in Faroese, where /sk/ followed by /t/ is reversed (Lockwood 1955: 23f.).

(3) *Faroese metathesis of* /sk/ (Lockwood 1955: 24)

masculine	*neuter*	
fɛsk-ur	fɛks-t	'fresh'
rask-ur	raks-t	'energetic'
dansk-ur	daŋks-t	'Danish'

As noted above, metathesis has typically been considered a sporadic or irregular process, unlike phenomena such as assimilation that can often be described in very general and regular terms (see Hume 2001: 1f. for representative quotations). But the Faroese reversal illustrates that a process of metathesis can be fully regular while also quite restricted in scope, simply because the necessary configuration does not often arise. Thus the neuter noun suffix /t/ provides the environment for reversal of stem-final /sk/; but a similar environment in verbs also triggers the changes, as can be seen in /ʊʏŋks-ti/ 'wish (PAST SG)' compared to the present singular /uʏnsk-ir/ with the underlying ordering thanks to the following vowel (Hume 1999: 294).

It was traditionally claimed that metathesis yields sequences that are in some way better formed than the input ordering, usually in the sense of "ease of articulation" or satisfying a language's phonotactic constraints (Wechssler 1900: 497; Grammont 1933: 239; Ultan 1978: 390). More recent work has placed greater emphasis on the role of perception (CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY), and on historical explanations for how metathesis arises (CHAPTER 93: SOUND CHANGE). Faroese can be seen as *auditory metathesis* – the temporal decoupling of the noise of a fricative, especially a sibilant, from the surrounding signal, which can lead to a sibilant and an adjacent stop being reinterpreted as occurring in the opposite of the original order (Blevins and Garrett 2004: 120). A segment often moves to a position in which it is more easily perceptible, especially due to the formant transitions in an adjacent vowel (Hume 1999: 295f.; Seo and Hume 2001: 215–217). Thus Faroese metathesis places the stop /k/ in a more perceptible position, adjacent to the preceding vowel, while the sibilant remains perceptible without an adjacent vowel. This directionality suggests that confusibility in the ordering of the segments is not the sole factor, since symmetrical confusion predicts random reordering according to the two possible interpretations of an ambiguous auditory signal (Steriade 2001: 233–235); but see Blevins and Garrett (2004: 119f.) for a defense of the misperception account. The outcome in particular languages may depend on prosody, such as the location of stress, and phonetic detail, such as the release of final stops; such differences may explain the symmetrically opposite changes in Late West Saxon (/frosk/ → [froks] 'frog') and a certain variety of colloquial French (/fiks/ → [fisk] 'fixed') (Blevins and Garrett 2004: 139f.).

A transformational rule that reverses the order of segments does not make reference to the apparent motivations of the reordering, such as an improvement in markedness (CHAPTER 4: MARKEDNESS). But like other phonological processes, metathesis may operate in order to satisfy the phonotactic restrictions of a language. That is, just as the place assimilation in *anba* → *amba* satisfies a condition that nasal codas must agree in place with a following stop, so a metathesis such as *inma* → *imna* in (6) satisfies a condition on the sequencing of coronal and labial consonants. Recent approaches have attempted to capture this insight and to treat metathesis more on a par with other processes.

In the surface orientation of Optimality Theory (Prince and Smolensky 2004), the expected Faroese sequence [skt] can be penalized by a constraint against a stop that occurs between two other consonants (Hume 1999: 298), whether it is defined directly in terms of perceptibility or as a more abstract configuration. This pressure must dominate the correspondence constraint LINEARITY, which otherwise prevents reorderings of segments, and obviously plays a central role in the analysis of metathesis (Hume 1998: 149, 68f.; McCarthy and Prince 1995: 371f.; McCarthy 2000: 173). Metathesis occurs only when LINEARITY is ranked below faithfulness constraints such as MAX and DEP; these prevent deletion or insertion of material that otherwise might serve to remedy the surface constraint that metathesis addresses.

(4)

/raskt/	*STOP/C__C	MAX	DEP	LINEARITY
☞ a. raskt	*!			
b. rast		*!		
c. rask		*!		
d. raskit			*!	
e. rakst				*

Naturally, no violation of LINEARITY is required in a form such as [raskur], where the stop /k/ is adjacent to a vowel, and the sequence surfaces intact.

Another relatively restricted case of stop–sibilant metathesis is the Tiberian Hebrew *hitpa'el* verb form, where the /t/ of the prefix reverses with a stem-initial sibilant (Malone 1993: 52f.; Coetzee 1999: 106; see Malone 1971 for similar facts in other Semitic languages; see also CHAPTER 108: SEMITIC TEMPLATES). The examples in (5a) show the lack of metathesis with non-sibilants.

(5) *Tiberian Hebrew metathesis* (Coetzee 1999: 106)

a.	hit-pallel	→	hitpallel	'he prayed'
	hit-qaddeʃ	→	hitqaddeʃ	'he sanctified himself'
b.	hit-sappex	→	histappex	'he felt attached to'
	hit-ʃammer	→	hiʃtammer	'he protected himself'
	hit-śakker	→	hiśtakker	'he gave himself into service'
	hit-zakker	→	hizdakker	'he remembered'
	hit-sˤaddeq	→	hisˤtˤaddeq	'he considered himself righteous'

For Coetzee (1999: 122f.), the motivation for metathesis in exactly this context, when a /t/ would otherwise precede a sibilant, is that a [t] + sibilant sequence would be subject to reinterpretation as an affricate, a type of segment disfavored in Tiberian Hebrew. He proposes a constraint *t+SIBILANT against that sequence, again with relatively low-ranked LINEARITY. Hume (2004: 222f.), discussing the equivalent metathesis in Modern Hebrew, argues that the poor attestation of [t] + sibilant sequences in Hebrew sets the stage for a reinterpretation with the sibilant in first position: an ambiguous acoustic signal is likely to be interpreted sequentially according to the most commonly attested ordering of those segments, dependent not necessarily on universal principles, but on the lexicon and grammar of the language in question.

1.1.2 *Place of articulation*

Many instances of CC metathesis depend on place of articulation (CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION), with certain orderings of place favored over others. Permutation of this type is found in a range of Malayo-Polynesian languages (Blevins and Garrett 2004: 136). In Cebuano, for instance, a coronal stop or nasal followed by a labial or velar consonant is reversed, optionally in some cases (Blust 1979: 110).[1] The consonants come to be adjacent as the result of vowel syncope after a vowel-initial suffix is added.

(6) *Cebuano metathesis of coronal + non-coronal clusters* (Blust 1979: 110)

stem	*suffixed form*	
lutuk	lukt-un	'put the finger in'
gitik	gitk-anun ~ gikt-anun	'ticklish'
atup	atp-an ~ apt-an	'roof'
inum	imn-a	'drink'

Stems such as /lakat/ 'walk' that already have the preferred ordering maintain it ([lakt-un]), showing that the process is not simply an across-the-board reversal in consonant clusters. In this case, the favored ordering places the coronal in second position.

The two changes in Cebuano – deletion of the vowel and reversal in the resulting cluster – were likely ordered historical events, and this history can be modeled easily by ordered synchronic rules. But the same facts are also consistent with simultaneous satisfaction of two surface constraints in OT. The candidates *[lutukun] and *[lutkun] both violate one of these constraints – by the lack of syncope, or the disfavored consonant ordering – whereas [luktun] satisfies both, and wins under low ranking of LINEARITY and MAX-V.

(7)

/lutuk-un/	*VCVCV	*TK	LINEARITY	MAX-V
a. lutukun	*!			
b. lutkun		*!		*
☞ c. luktun			*	*

A phonetic explanation for this type of reordering is *co-articulatory metathesis*, which results from the overlap in adjacent consonant gestures (Blevins and Garrett 2004:

[1] Similar patterns are found in a number of other Philippine languages (Blust 1971: 85f.; 1979: 104f.; Crowhurst 1998: 597), including Tagalog, where, however, it is poorly attested and classified with irregular verbs (Schachter and Otanes 1972: 375–380).

136–138); for example, overlapping coronal (T) and non-coronal closures (K) are perceived as the non-coronal, which leads to reversals such as TK → KT in Cebuano (6). A general preference for apicals to follow non-apicals has been cited with regard to metathesis in other languages such as Greek, which may be related to the tendency for coronal codas to assimilate to following non-coronals (Bailey 1970: 348). One abstract phonological approach formalizes the licensing properties of different places of articulation (Rubin 2001: 194–199); unmarked Coronal is a *natural head* and licenses the place features of a preceding non-coronal, favoring KT over TK. See also Blust (1979: 102f.) and Winters (2001) on the general preference for coronals to occur second in a cluster.

Some reorderings have considerably more complex origins. In the Kondh branch of Dravidian, sequences of a velar /k g/ plus a labial /p b/ are reversed. The Pengo examples below illustrate two allomorphs /-pa/ and /-ba/ of the intensive-frequentative or plural action suffix, both of which also occur in contexts without metathesis as seen in (8a). For similar Kui examples, see Hume (2001: 8).

(8) *Pengo velar + labial metathesis* (Burrow and Bhattacharya 1970: 82f., 201)

a.	gru:t-pa-	→	gru:t-pa-	'fell'
	huz-ba-	→	huz-ba-	'roast'
b.	ḍrik-pa-	→	ḍripka-	'break'
	ku:k-pa-	→	ku:pka-	'call'
	ṛa:k-ba-	→	ṛa:bga-	'sacrifice'
	tog-ba-	→	tobga-	'be split'

According to Garrett and Blevins (2009: 538ff.), this metathesis pattern arose by re-analysis of complex allomorphy deep in the history of Dravidian. Briefly, causative /p/ could replace the last consonant of the stem, as in Kolami [melg-] 'grow (INTR)' and derived [mel-p-] 'rear'. This was interpreted as a rule deleting the velar before the labial, which was extended to other labial-initial suffixes, including the plural action containing [-p-]. This would yield an alternation between simple */ku:k-/ 'call' and plural action */ku:k-p/→ *[ku:-p-]. But there is another basic allomorph of the plural action suffix containing /-k-/; if this were added to the existing plural action in order to make the exponence of that category clearer, the result is the pair *[ku:k-] and *[ku:-p-k-], which then gives the appearance of metathesis of the suffixal /p/ and the stem-final /k/.

Whatever the historical origin of Pengo metathesis, it became part of the grammar thanks to learners treating it as an active synchronic process. The constraint encoding the Pengo alternation must penalize a velar + labial sequence; call it *KP. In addition to MAX and DEP, it is especially relevant here to include the constraint IDENT to prevent changes to the features targeted by the phonotactic constraint.

(9)

/togba/	*KP	MAX	DEP	IDENT	LINEARITY
a. togba	*!				
b. toba		*!			
c. togiba			*!		
d. tomba				*!	
☞ e. tobga					*

Perceptual factors may contribute to such re-analyses. Placing the labial first in the cluster, at least in related Kui, puts it in the stressed syllable, which may enhance its perceptibility; the weak bursts of labial stops reduce the benefit of being located in the onset (Hume 1999: 296). Experimental evidence indicates that labial place is more perceptible in codas than is velar, and that velars benefit more in perceptibility from being in the onset than do labials (Winters 2001: 238–241). This means that the ordering PK is overall more likely to be heard correctly than KP. Emphasizing historical origin, however, Blevins and Garrett (2004: 136) consider the perceptually based prediction to be PK → KP, as attested in other languages such as Mokilese /apkas/ → [akpas] 'now', and claim that the Dravidian pattern favoring PK could arise only by such means as re-analysis of a morphological pattern.

Segments undergoing metathesis may originate in different morphemes, as in Pengo, but may also occur inside a single morpheme. Across a morpheme boundary, the offending cluster is created by concatenation; within a morpheme, the context for metathesis may be created directly by a syncope rule that brings the consonants into contact (as in Cebuano (6)), or a triggering context introduced by concatenation but affecting two consonants that are underlyingly adjacent (as in Faroese (3)).

1.1.3 *Manner of articulation*

Classes of consonants defined by manner, such as liquids or sonorants (see CHAPTER 13: THE STRICTURE FEATURES), are often targeted specifically by metathesis. (One might also include here the sibilants discussed in §1.1.1.) Metathesis involving the class of liquids is found in a number of languages (Blevins and Garrett 2004: 128f.); a historical example from Persian was cited in (1). In Rendille (Cushitic, Kenya), an /r/ and a preceding obstruent or nasal reverse in order after they become adjacent upon deletion of the intervening vowel (Sim 1981: 7, 9f.; Hume 1998: 178; Blevins and Garrett 2004: 129).

(10) *Rendille metathesis of /r/ in clusters* (Sim 1981: 7, 9)

feminine	*masculine*	
údur-te	úrd-e	's/he slept'
ágar-te	árg-e	's/he saw'
hámar-te	hárm-e	's/he shivered'

In the framework of Blevins and Garrett (2004: 121–125), this *perceptual metathesis* arises when the cues for a sequence of sounds are perceived by the listener as reordered relative to the speaker's intention, which is possible when some feature is realized over a relatively long duration and therefore contains ambiguity of analysis. This is true for consonant clusters as well as vowel–consonant sequences (see §1.2). Besides liquids (CHAPTER 30: THE REPRESENTATION OF RHOTICS; CHAPTER 31: LATERAL CONSONANTS), other segment types with elongated cues include pharyngeals (CHAPTER 25: PHARYNGEALS), secondary labialization (CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION), palatalization (CHAPTER 71: PALATALIZATION), and glottalization or aspiration (Blevins and Garrett 2004: 123). Hume (2004: 220–227) argues similarly that ambiguity or indeterminacy in the auditory signal sets the stage for a reinterpretation of linear order, but places a special emphasis on the role of the specific attested sequences in the language, as discussed for Hebrew above.

In Kambata (East Cushitic, Ethiopia), a suffix-initial nasal transposes with a preceding obstruent, and is also subject to place assimilation in this position, as illustrated by [ŋk] and [mb] resulting from an /n/-initial suffix (Hudson 1980: 105); similarly the related language Sidamo (Vennemann 1988: 55) and several other East Cushitic languages (Garrett and Blevins 2009: 532f.).

(11) *Kambata metathesis of obstruent + nasal clusters* (Hudson 1980: 105)

it-ne:mmi	→	inte:mmi	'we have eaten'
t'u:d-na:mmi	→	t'u:nda:mmi	'we will see'
oros-na:mmi	→	oronsa:mmi	'we will take'
sok-ne:mmi	→	soŋke:mmi	'we have sent'
hab-no:mmi	→	hambo:mmi	'we forgot'

This metathesis is part of a conspiracy (CHAPTER 70: CONSPIRACIES) of changes (including complete assimilation and vowel epenthesis) that avoid ill-formed consonant clusters, in this case obstruent + sonorant; see Hume (1999: 300–302) for a perceptual-optimization account. In a novel strategy that anticipates Optimality Theory, Hudson (1980: 109) proposes that affixation generates two outputs with alternate orderings of juxtaposed consonants (such as [itne:mmi], [inte:mmi]), where the choice between the outputs is made according to conditions on phonotactics. This technique would not, however, generalize to examples such as Faroese and Cebuano, in which the relevant consonants are not juxtaposed across a morpheme boundary.

The obstruent + nasal reversal in East Cushitic has been cited as an example of a metathesis that does not result from a conventional source such as a misperception of the ordering of the cues (Garrett and Blevins 2009: 532–537). Rather, it appears to reflect the pressure of other consonant interactions; I illustrate with the facts of Kambata, but follow the argument of Garrett and Blevins, who use data from related Bayso. In some clusters that occur at stem boundaries, we find regressive assimilation, as in /rn/ → [nn] and /mt/ → [nt]; but in others, there is apparent progressive assimilation to create a geminate, as in /bt/ → [bb].

(12) *Kambata assimilation in clusters* (Hudson 1980: 105)

a.	im-to:ʔi	→	into:ʔi	'she dug'
	ful-na:mmi	→	funna:mmi	'we will go out'
	kam-no:mmi	→	kanno:mmi	'we forbade'
	mar-ni	→	manni	'we, going'
b.	ub-to:ʔi	→	ubbo:ʔi	'she fell'
	dag-tonti	→	daggonti	'you knew'
	t'u:d-tenti	→	t'u:ddenti	'you have seen'
	oros-ta:nti	→	orossa:nti	'you will go'

If the learner seeks to generalize to a single process of regressive assimilation, then an intermediate step of metathesis is necessary to create the right outcome: /bt/ → tb → [bb]. Extending this to instances of obstruent + nasal, such as /bn/

→ nb → [mb], also has the effect of preserving the features of the root-final consonant and yielding a nasal + obstruent sequence of the sort that is common in other concatenations. Interestingly, whereas Kambata assimilation and metathesis apply to all places of articulation, in Bayso both are restricted to coronals, which reinforces the connection; this correlation is found across the East Cushitic languages (Garrett and Blevins 2009: 536f.). From the perspective of synchronic phonology, an unavoidable conclusion is that metathesis processes are available to the learner, whether the pattern results from a misperception of the phonetic signal or a generalization of an existing pattern.

1.2 CV metathesis

Ordering reversals of a consonant and vowel involve many of the same principles of explanation and analysis as CC reversals – for example, the historical reinterpretation of an ambiguous signal, and a synchronic constraint that dominates LINEARITY. From the examples in the literature, however, synchronic CV metathesis appears to be strongly associated with specific morphological contexts, and the reordering may be the main exponence of a grammatical category, something that is not typical of CC metathesis. But before considering such cases, we examine a few more strictly phonological examples.

1.2.1 *Phonological reorderings*

A well-known case that has been treated as metathesis is Cayuga (Iroquoian), in which a laryngeal consonant /h ˀ/ transposes with a preceding vowel when it occurs in an odd-numbered, non-final syllable (Foster 1982: 69f.; Blevins and Garrett 1998: 509–512). The necessary prosodic context can be analyzed as the weak branch of an iambic foot (Hayes 1995: 222f.; see also CHAPTER 44: THE IAMBIC–TROCHAIC LAW).

(13) *Cayuga laryngeal metathesis* (Foster 1982: 69f.)

kahwistaˀeks	→	kʰḁ'wisdˀaes	'it strikes, chimes'
koˀnikõhaˀ	→	gˀo̥'nikhwaˀ	'her mind'
akekahaˀ	→	a'gekhaaˀ	'my eye'
ahanohaeˀ	→	a'hanhwaeˀ	'he washed it'

To some degree, however, it is uncertain whether this process is truly a reversal of segment order or instead a spreading of features across the vowel, resulting in overlap rather than reordering (Foster 1982: 70). Somewhat similar metathesis of vowel + /h/ occurs in Cherokee when a stop consonant precedes the vowel; the result is that the laryngeal is realized as aspiration on the stop (Flemming 1996; Blevins and Garrett 1998: 520f.).

In the framework of Blevins and Garrett (1998: 509f., 2004: 121–125), Cayuga and Cherokee show the results of *perceptual metathesis*. Just as with the CC metathesis involving liquids and other segment types, the spread of laryngealization or devoicing through the vowel leads to the possibility of reinterpretation. Diachronic instances of the same phenomenon include liquid metathesis in Slavic, as in *orbota 'work' > Polish /robota/; and reordering of /r/ with schwa in Le Havre French, such as [bərbi] 'ewe' compared to standard [brəbi] (Blevins and Garrett

1998: 513, 16f.). The Slavic example is somewhat unusual, in that it involves an initial sequence undergoing metathesis; reordering is cross-linguistically disfavored for root-initial segments, since a disruption in that position interferes with effective word recognition more than metathesis of other segments (Hume and Mielke 2001).

A clearer example of synchronic CV metathesis is found in the Austronesian language Leti, discussed in detail by Hume (1998) and Blevins and Garrett (1998: 541–547). Alternating stem forms in Leti are phonologically conditioned according to the following context; in particular, morpheme-final VC reverses to CV to avoid an illicit consonant cluster within a phrase.

(14) *Leti* VC *metathesis* (Hume 1998: 153)

a.	ukar lavan	→	ukarlavan	'thumb, big toe'
b.	ukar ppalu	→	ukrappallu	'index finger'
	ukar muani	→	ukramwani	'middle finger'

There is, additionally, the same reversal in phrase-final position, so that 'finger' appears as [ukra]; here, rather than serving general phonotactics, the metathesized form appears to mark the word as phrase-final (Bonthuis 2001: 37f.).

Because metathesis in Leti affects all consonant types – compare /ulit/ → [ulti] 'skin', /mɛtam/ → [mɛtma] 'black' – it cannot be attributed to the elongated phonetic realization of a class such as laryngeals, and is not perceptual metathesis. Instead, Blevins and Garrett (1998: 539–547) identify it as *pseudo-metathesis*. By this they mean an alternation in ordering that did not arise historically as a direct reinterpretation of segment order. In the case of Leti, two main steps are posited, with evidence from other patterns within Leti and in related languages. First, an epenthetic vowel was inserted after final consonants, /ulit/ → [ulitə̟]. Although the inserted vowel was not a copy of the preceding vowel, it nevertheless would have been subject to co-articulatory effects of the more palatal or labial quality of a preceding /i/ or /u/, as in [ulitə̟]. Second, syncope of medial vowels led to loss of that schwa preceding another word beginning CV ([ulitə̟] → [ulit]), but loss of the original medial vowel in other contexts ([ulitə̟] > [ultə̟]); here, however, the vowel quality of the deleted vowel is preserved in the final schwa due to the co-articulation ([ultə̟] > [ulti]). Words containing the low vowel, such as /ukar/, do not show palatal or labial co-articulation, but result from the fact that schwa more generally became /a/ in the history of Leti (thus /ukarə/ > /ukrə/ > /ukra/). From the point of view of synchronic phonology, the crucial point is that alternations such as [ulit] ~ [ulti] were successfully integrated into the grammar as learners re-analyzed the historical patterns.

1.2.2 *Morphological context*

As noted above, CV metathesis often appears to occur in the presence of a particular morphological trigger, even if the reordering that occurs can be defined phonologically. A famous example, also from Austronesian, is found in Rotuman. In this language, words appear in two different "phases," called complete and incomplete (Churchward 1940). The incomplete form is derived from the complete by a variety of means, but the default strategy is metathesis of the final CV to VC, often forming a short diphthong with the preceding vowel.

(15) *Metathesis in Rotuman phase alternations* (Churchward 1940: 14)

complete	*incomplete*	
ho.sa	hoas	'flower'
i.ʔa	iaʔ	'fish'
pu.re	puer	'to rule, decide'
ti.ko	tiok	'flesh'
se.se.va	se.seav	'erroneous'

The process applies to loanwords as well, such as /pe.pa/ → [peap] 'paper'. Consistent with many other languages, Rotuman short diphthongs must consist of two vowels with rising sonority (i.e. movement from a higher to a lower vowel). Where this condition is not met, the incomplete phase is realized in other ways: by dropping a final vowel, as in /to.ki.ri/ → [to.kir] 'to roll'; by fusing two vowels brought together by metathesis, as in /mo.se/ → [møs] 'to sleep'; or by directly changing a vowel sequence to a long diphthong, as in /ke.u/ → [keu] 'to push'.

Blevins and Garrett (1998: 527–529) categorize the Rotuman alternation as *compensatory metathesis*. Historically, this entails an anticipation or perseveration of vowel features across an intervening consonant toward the stressed vowel, leading to "extreme vowel-to-vowel coarticulation." In Rotuman there would have been anticipation of the final vowel in the direction of the preceding stressed syllable, followed by loss of the final vowel, essentially /hosa/ > /hoasa/ > /hoas/. In some vowel sequences, further changes occurred, such as /mose/ > /moese/ > /moes/ > /møs/. Metathesis of similar origin is also pervasive in the related language Kwara'ae, where final CV changes in most communicative contexts to VC to mark phrasal boundaries (Sohn 1980: 311f.); the details of Kwara'ae vowel realization lend particular support to the proposed historical origin as compensatory metathesis (Blevins and Garrett 1998: 530f.).

The phases of Rotuman were originally described by Churchward in complex syntactic and semantic terms, but some recent work has argued that their specific realization depends on prosody, and therefore that they are basically phonologically determined rather than triggered by a morphological or other grammatical context (Hale and Kissock 1998: 120–123). For example, it has been proposed that the desired outcome achieved by metathesis as well as the other processes is a word-final heavy syllable (Blevins and Garrett 1998: 531–534; McCarthy 2000: 159, 73f.). From this point of view, the reversal in order is just one way of satisfying the heavy-syllable constraint; there is no specific rule demanding metathesis. This analysis relies on the claim of Hale and Kissock (1998) that the complete phase occurs before monomoraic morphemes such as /-me/ 'hither' in [hoʔa-me] 'to bring', and the incomplete before bimoraic morphemes such as transitive /-kia/ in [hoaʔ-kia] 'to take (TRANS)'. The essential idea is that right-aligned prosodic structure in [ho(ʔa-me)] requires the stem-final CV syllable to be grouped with the CV suffix in a proper bimoraic foot, and the pressure for a stem-final heavy syllable is thwarted. But in [(hoaʔ)-(kia)], the stem and the suffix are footed independently, and the stem undergoes metathesis to ensure a stem-final heavy syllable. The same result is predicted in the absence of a suffix.

Kurisu (2001: 187) cites, in addition to certain exceptional suffixes, minimal pairs from Churchward (1940: 15) showing that the two phases can occur in an identical phonological context, such as complete [ʔepa la hoaʔ] 'the mats will be taken' and incomplete [ʔeap la hoaʔ] 'some mats will be taken'. This overlap indicates that the phase changes must in some way be triggered by the presence of a morphosyntactic category, the Incomplete Phase. For Kurisu, the high-ranked constraint REALIZEMORPHEME forces the incomplete to be phonologically distinct from the complete phase (CHAPTER 103: PHONOLOGICAL SENSITIVITY TO MORPHOLOGICAL STRUCTURE); the relative ranking of phonological constraints, including LINEARITY, determines exactly how the base form is modified. The metathesis outcome is favored by the constraint ranking, although particular configurations (such as vowel sequences with falling sonority) lead to other outcomes, including fusion of the vowel features. Notably, in this more morphologically oriented approach, there is still no specific morphological demand for metathesis; rather, the drive for distinctness of word forms interacts with phonological constraints to produce metathesis, among other results.

The examples presented so far involve underlying CV changing to VC, especially stem-finally. The converse change at the left edge, where initial VC changes to CV, is attested in some Northern Paman languages such as Ngkot̪, following the historical loss of initial consonants (Hale 1976: 17f., 23–28; Blevins and Garrett 1998: 537f., 2004: 135f.).

(16) *Ngkot̪ initial VC metathesis* (Hale 1976: 23–28)

*nʲipul-	>	*ipul-	>	pjul-	'you (NON-SG)'
*nˡiːna-	>	*ina-	>	n̪ja-	'to sit'
*kulan-	>	*ulan-	>	lwan-	'possum'
*puŋa-	>	*uŋa-	>	ŋwa-	'son'
*ŋali-	>	*ali-	>	laj-	'we (DUAL INCL)'
*kami-	>	*ami-	>	maj-	'mother's mother'

Unlike in Rotuman and a number of other Austronesian languages, however, this metathesis appears to be diachronic only.

A somewhat similar pattern is found synchronically for a number of verbs in the Nilo-Saharan language Fur (Jakobi 1990: 57f., 64–74; Hume and Mielke 2001: 141f.). These verbs, when preceded by a monoconsonantal person-marking prefix, undergo reversal of the initial CV.

(17) *Fur CV metathesis under prefixation* (Jakobi 1990: 57f., 64–74)

k-ba-	→	kab-	'we drink'
k-teer-	→	keter-	'we forge'
k-lat-	→	kald-	'we beat, hit'

Some alternations are quite irregular, such as /li-/ → [al-] 'wash' and /tii-/ → [ei-] 'catch', so that a plausible alternative is that the allomorphs are lexically listed. This account would also address the formal problems in alternations such as /bul-/ → [ulb-] (→ [ulm-]) 'find', which involve two apparent metatheses (Hume 2001: 18f.); see §2.3 below. Even if the allomorphs are listed, however, metathesis was a crucial historical source.

1.2.3 *Metathesis in templates*

Languages with *templatic morphology* express certain inflectional or derivational categories by changes to the syllable structure of the stem (see CHAPTER 105: TIER SEGREGATION; CHAPTER 108: SEMITIC TEMPLATES). If a particular paradigm includes different orderings of C and V elements, then the result is a form of metathesis. Templatically created metathesis generally does not derive from general phonological properties of a language, but rather from potentially arbitrary exponence of a morphological category. For example, a relatively productive metathesis applies to derived Classical Arabic nominal stems two syllables in length; initial /Ca/ is reversed to [aC] and a glottal stop onset is inserted (McCarthy and Prince 1990: 213f., 279f.). Examples include /kabar/ → [ʔakbar] 'greater, greatest' and /dʒanib-at/ → [ʔadʒnib-at] 'wings'; compare the underived forms [kabiːr] 'great' and [dʒanaːb-at] 'wing', without metathesis. The change cannot be treated as a general phonological process, because it is limited to certain morphological categories, and does not occur in verbs such as /katab/ 'he wrote' (*[ʔaktab]).

In Mutsun, a Costanoan language of northern California, templatic alternations include the reversal of a stem-final sequence of vowel and consonant; the primary stem is consonant-final and the derived stem is vowel-final (Okrand 1979: 126f.). The choice between alternate verb stem forms depends on what suffix is added; in other cases (18b) the primary stem is a noun, and the derived stem is a verb with related meaning. The derived stem has the uniform shape CVCCV, despite considerable variation in the primary stem shape.

(18) *Mutsun stem alternations* (Okrand 1979: 126f.)

	primary	*derived*	
a.	pasik-	paski-	'to visit'
	litʃːej-	litʃʼje-	'to stand'
	matːal-	matla-	'to be face down'
b.	lulːup-	lulpu-	'flute / to play the flute'
	toːher-	tohre-	'a cough / to cough'
	laːlak-	lalka-	'goose / to get geese'
	posoːl-	poslo-	'posole (stew) / to make posole'

Okrand observes that while the vowel-final derived stem is the form used with all suffixes that would create an illicit consonant cluster if added to the primary stem ([litʃʼje-hte] 'standing', *[litʃːej-hte]), it also occurs with some suffixes that would be phonotactically well-formed with a preceding consonant ([matla-nu] 'put (someone) face down', alongside [mattal-pu] 'put oneself face down'). Therefore, this morphologically defined reordering does not merely repair phonological violations, even if it sometimes conspires to avoid phonotactically problematic concatenations. It has been pointed out that for similar alternations in related Sierra Miwok, a representation with V/C segregation makes a specific metathesis rule unnecessary (Smith 1985: 366f.; Goldsmith 1990: 91; Stonham 1994: 157f.); more on this below.

In Tunisian Arabic, a stem-internal alternation is a cleaner example of metathesis than what we find in Classical Arabic (Kilani-Schoch and Dressler 1986; Becker 2000: 579f.). Historical changes to vowels within stems have led to minimal differences defined by ordering, such as Classical Arabic /malak-a, milk-u/ > Tunisian /mlək, məlk/; this pattern is now productive in relating triliteral surface forms.

(19) *Tunisian Arabic stem alternations* (Kilani-Schoch and Dressler 1986: 62, 65f.)

mlək	'he possessed'	məlk	'property'
fhəm	'he understood'	fəhm	'understanding'
ħrəm	'he forbade'	ħarm	'prohibition'
kfɔr	'he blasphemed'	kɔfr	'blasphemy'

A similar alternation is found in Alsea (Buckley 2007). In this coastal Oregon language, stems generally show at least two forms; the full stem contains a root vowel, while the short stem lacks it. For stems with a medial sonorant consonant, an additional distinction is found: the full stem occurs in two varieties, light and heavy, according to whether the root vowel follows or precedes the sonorant. The stem choice depends on the presence of particular suffixes as well as an aspectual distinction.

(20) *Alsea stems with a medial sonorant* (Buckley 2007: 8f.)

light (CV)	*heavy* (VC)	*short* (no V)	
stlak-	stalk-	stlk-	'slide'
twih-	tiwh-	twh-	'pour'
tmus-	tums-	tms-	'close'

In the analysis of Buckley (2007: 15–18), the light stem is the underlying form; the short stem is created by deletion of the root vowel, and the heavy stem results from VC metathesis. Since only sonorants undergo this potential reordering, they alone are treated as weight-bearing in the coda, and therefore only in that case can metathesis yield satisfaction of the heavy template requirement. The same approach might be applied to Tunisian Arabic, with the difference that all consonant classes are moraic, and therefore metathesis applies to stems regardless of the medial consonant. The larger point is that the requirement for a heavy syllable is morphologically determined, but the effect is generated phonologically.

Similar is the stem alternation found in Klallam and other Straits Salish languages, in two forms called the actual and non-actual aspect (Thompson and Thompson 1969: 215–217; Demers 1974: 17f.; Montler 1989: 96f.).

(21) *Klallam stem alternations* (Thompson and Thompson 1969: 216)

non-actual	*actual*	
ʧ'kʷu-	ṭ'ukʷ-	'shoot'
ɬq'i-	qiq'-	'restrain'
ɹ̣q'ɔ-	ɹ̣ɔq'-	'swallow'
mtəqʷ-	mətqʷ-	'put in water'

Anderson (2005: 9–11) argues that synchronically, Klallam and similar languages require a processual rule of metathesis to express this morphological category. Montler (1989: 93), however, expresses this effect for Saanich as a CVCC template that causes metathesis in a form such as /ˈsχət/ 'push it' → [ˈsəχt] 'pushing it'. In roots where the CVCC template cannot be satisfied by metathesis, other strategies are available, such as glottal stop insertion after the vowel in /ˈweqəs/ 'yawn' → [ˈweˀqəs] or reduplication to achieve this templatic result, as in /ˈqen'/ 'steal' → [ˈqeqn'] (→ [ˈqeqən'] by epenthesis). The same additional strategies occur in Klallam as well. In a prosodic version of the template approach, Stonham (1994: 173f.) proposes insertion of a mora in Klallam and Saanich that forces CCV to surface as CVC, a heavy syllable, and also causes the related effects of coda insertion and reduplication. The templatic and moraic approaches treat metathesis as one possible means of satisfying the morphologically determined, but phonologically expressed, restriction on shape. As with Rotuman, metathesis is one change among several, and not necessarily the direct goal of the morphological category.

In a more strictly phonological approach for closely related Lummi, Demers (1974: 16) proposes a rule that deletes unstressed schwa between obstruents. In this view, the actual and non-actual forms are both based on a CəCə root, but have different stress placement. Schwa deletion yields apparent metathesis in pairs with the surface shapes seen in/ˈCəCə/ → [ˈCəC] and /CəˈCə/ → [ˈCCə]. Although the synchronic evidence for exactly this derivation is missing in Klallam, the Lummi pattern suggests the likely diachronic origin of metathesis as a reinterpretation of vowel deletion. Such a historical origin can explain why these templatic changes normally involve reorderings of consonants and vowels, but not of consonants. For example, suppose that (similar to Lummi) the Alsea stem 'to close' that alternates between [tmus] and [tums] derives from original *tumus, with deletion of the unstressed vowel in forms with distinct stress patterns due to different suffixation: *ˈtumus-a > /ˈtums-a/ 'door' and *tuˈmus-χ > /ˈtmus-χ/ 'is closed' (Buckley 2007: 22f.). The alternate forms that preserve different vowels are subject to reinterpretation as a stem with a single underlying vowel that is reordered with the adjacent consonant in different suffixal contexts; but vowel deletion by itself will not result in the reordering of consonants. Given the frequency of vowel harmony and syncope, patterns like this can be expected to arise rather often.

Despite the crucial role of morphological context in conditioning these reorderings, phonological techniques can often be used to generate the necessary effects. One important tool has been the segregation of vowels and consonants onto different tiers (see CHAPTER 105: TIER SEGREGATION), so that they have no underlying ordering and no actual metathesis occurs in the derivation (McCarthy 1989: 5, 22f.). The advent of Optimality Theory, with its emphasis on output constraints rather than restricted input representations, makes V/C segregation "superfluous" (McCarthy 2000: 180f.). Even in an approach that does not treat the consonantal root as a morpheme listed independent of any vowels, derivational and inflectional morphemes often consist of vowels that overwrite the underlying vowels of the stem (Ussishkin 2005). Apparent VC metathesis among surface forms is merely the result of different overwriting patterns, as when the elements /h i i/ are imposed on Modern Hebrew /gadal/ 'grow' to form [h-igdil] 'enlarge'. Constraints on the realization of affixal material in the stem lead to particular overwriting patterns, but the vowels of the affixes still have no underlying ordering relation to the consonants of the input word.

It is less clear how a vowel-overwriting approach for Semitic can extend to language families such as Miwok-Costanoan and Yokuts, where the vowels and consonants can be reordered, but the vowels do not have the status of separate morphemes (McCarthy 1989: 74, 78). Thus in Mutsun, the verb 'to visit' is lexically specified with not only the consonants /psk/ but also the vowels /ai/, combined in different ways, including [paski-] and [pasik-] (18). The overwriting operation would have to be available for subparts of one lexical entry, rather than independent morphemes, in order to account for languages like Mutsun.

1.3 VV metathesis

Although CV and CC metathesis are robustly attested, there is weak evidence for VV metathesis. Webb (1974: 8) states that "[e]ven as a sporadic change metathesis of vowels appears to be quite uncommon." Kiparsky and O'Neil (1976: 531, n. 7) believe "there are few if any rules that metathesize contiguous syllabic segments in any language." McCarthy (2000: 176) observes that the few synchronic analyses that posit VV reversals "involve very abstract analyses, in which the underlying representations and/or the consequences of metathesis are by no means apparent." The rarity of such reversals may be related to the much longer typical duration of vowel gestures compared to consonants, so that a considerable temporal shift would be required for re-analysis of the ordering of two vowels (Steriade 1990: 390f.; McCarthy 2000: 176).

A classic example is VV reversal in Kasem, to which Chomsky and Halle (1968: 361) first applied the transformational rule format for metathesis. In particular, the vowel sequence /ia/ is reversed to [ai] when followed by the plural suffix /i/; but the first /i/ deletes and then the remaining vowels coalesce, as in /pia-i/, which surfaces as [pe] 'sheep (PL)'. Needless to say, on first inspection /piai/ → [pe] is not an obvious example of metathesis. Phelps (1975: 303f., 10f., 25, 1979: 56f.) argues against the Chomsky and Halle VV metathesis rule, but in favor of an entirely different CV metathesis, in derivations such as /boa:l-u/ → [bola:-u] (→ [bolo] 'valley'). This derivation is again complex, although with different assumptions about underlying forms. Both of Phelps's general conclusions regarding Kasem metathesis – CV is transposed but not VV – are endorsed, in a more modern framework, by Haas (1988: 241–253, 45f.); see also Burton (1989: 29f.) for an analysis of vowel coalescence without an intermediate reordering.

Similar re-analyses have been proposed for other languages with apparent VV metathesis. Keyser (1975: 404) posits a rule for Old English that reverses vowels in order to feed a vowel elision rule, as in /lufa-i/ → lufia → [lufa] 'love!'; Kiparsky and O'Neil (1976: 535f.) argue that a revised formulation of vowel elision makes metathesis unnecessary. A rule of VV metathesis has been claimed to play "a central role" in Latvian phonology; it reverses the order of elements in the diphthongs /ai au æi æu/, although under restricted conditions (Halle and Zeps 1966: 108). In a more recent treatment of Latvian vowels, although not focusing on metathesis, Anderson and Durand (1988: 34, n. 7) reject some of the synchronic abstractness assumed by Halle and Zeps; instead they assume raising of a monophthongal vowel that then undergoes breaking to form a diphthong, where no metathesis is required.

A few diachronic examples of VV metathesis can be cited, especially if we include vowel/glide reversals in this category, since the same set of segmental features may serve as a glide or vowel before and after the metathesis (Ultan 1978: 375f.). Two examples from Portuguese are /genukulum/ > /geoʎo/ > /ʒoeʎo/ 'knee' and /dehonestaːre/ > /deostar/ > /doestar/ 'to insult' (Williams 1962: 111); these reversals may have occurred "on the analogy of the more familiar sequence *oe*" (Ultan 1978: 376).

1.4 Other types of metathesis

Permutations involving something larger than a segment may fall under broader definitions of metathesis (Ultan 1978: 370). These include syllable reversals in language games or *ludlings*, such as Chasu /i.ku.mi/ → [i.mi.ku] 'ten' (Bagemihl 1995: 704). Metathesis has also been proposed for elements such as *location* in sign language phonology (Sandler 1993: 246).

Hyman and Leben (2000: 590) state that there are "sporadic reports of tonal metathesis in the literature"; some examples include Bamileke-Dschang (Pulleyblank 1986: 41, 50), Mixtec (Goldsmith 1990: 25), and Dangme (Holscher *et al.* 1992: 126). These processes typically involve the movement of a floating tone, originating at the edge of a word or other domain, past a single linked tone. But – like VV metathesis in §1.3 – they are also embedded in complex derivations, and depend on multiple assumptions about how the pieces of the analysis fit together. Under other assumptions, metathesis may not be required. For example, Pulleyblank (1986: 41) proposes that in Bamileke-Dschang a floating L tone moves leftward across a H, and remains floating to represent downstep; Hyman (1985: 71, 73), on the other hand, links the L directly to the H on a second tonal tier as a direct representation of a downstepped H. In essence, the new rule is a merger rather than a reordering, similar to Zoque palatalization in §3 below.

It should be kept in mind that "metathesis" of syllabicity, as in French /oj/ > /wɛ/ > modern /wa/ and Proto-Slavic *ew > /ju/ (Ultan 1978: 376), does not involve transposition of segmental features but rather a shift in affiliation relative to the head of the syllable. Thus in French /oj/ > /wɛ/, the round vocoid continues to precede the front vocoid; in Slavic *ew > /ju/, the round element is second, but remains there. The same observation can be made for English /iw/ > /ju/, found in words such as *few* (Jespersen 1949: 101), which is quite similar to the Proto-Slavic change. None of these represents segmental metathesis.

2 Non-local effects

Often grouped with local metathesis is the exchange of segments that are not adjacent, called *long-distance* or *non-contiguous* metathesis (Ultan 1978: 380–383). In fact, *metathesis* or its equivalent in another language has been used, especially by earlier writers, specifically for such long-distance effects (Blevins and Garrett 1998: 525). Grammont (1933: 239ff., 339ff.) devotes separate chapters to long-distance *métathèse* and local *interversion*, a terminology found more recently, for example, in Pierret (1994: 61); but Wechssler (1900: 496) already uses *Metathese* for both local and long-distance transpositions. As noted at the beginning of this chapter, metathesis here refers to either type of reordering.

2.1 *Diachronic*

A famous diachronic example of transposition over intervening segments is the Spanish metathesis of *r . . . l* > *l . . . r*, observable in a few modern words (Ultan 1978: 381; Penny 2002: 36).

(22) *Spanish liquid metathesis* (Penny 2002: 36)

Latin		*Spanish*	
mi:ra:kulum	>	milagro	'miracle'
peri:kulum	>	peligro	'danger'
parabola	>	palabra	'word'

These pronunciations were probably influenced by the greater frequency of consonant + *r* in the lexicon; various sound changes had eliminated many inherited instances of consonant + *l* (Ultan 1978: 391; Penny 2002: 70–72); the change can also be viewed as two steps, first the change of /l/ to /r/ in a cluster, and then the well-attested dissimilation of identical liquids (Wanner 1989: 444f.).

Comparison of cognate words in the Yuman family of the American Southwest reveals a variety of historical metathesis processes, including root consonants in Walapai /'pil/ 'burns' ~ Cocopa /'lip/ 'flames up', or Havasupai /ka'to/ ~ Walapai /ta'ko/ 'chin' (Langdon 1976). There are also variant forms within languages, such as Ipai Diegueño /maxa'tun/ ~ /xama'tun/ 'knee'. These alternations are widespread, but remain lexically specific. Swapping of the consonants in largely CVC roots is also common in the Salish family, as seen in apparent cognate pairs such as Shuswap /xʷej/ ~ Twana /jəxʷ/ 'disappear', and Klallam /ts'əq'ʷ/ ~ Upper Chehalis /q'ʷəts'/ 'dirty' (Noonan 1997: 482). The pervasiveness of this pattern in Salish is unusual, and is possibly best explained by historical processes of reduplication and consonant deletion rather than direct metathesis (Hume and Mielke 2001: 143, n. 4; Noonan 1997: 513).

Prunet (2006: 57–61) discusses examples of consonant metathesis within Semitic roots. These are said to be particularly common in the Hebrew lexicon, as in synonymous variants such as [keveś] ~ [keśev] 'lamb' and related meanings such as [ʔa:raz] 'tie packages' ~ [ʔa:zar] 'bind, girdle' (Horowitz 1960: 228–234). More dramatic examples of non-contiguous consonant metatheses are found in language games in Bedouin Hijazi and Moroccan Arabic, in which the root consonants are scrambled (Prunet *et al.* 2000: 623f.); in Hijazi, /kattab/ 'caused to write' can be realized as [battak], [takkab], [kabbat], [tabbak], and [bakkat]. Although this radical permutation is not part of the basic grammar, such language games show an impressive computational capacity for synchronically active metathesis (Bagemihl 1995: 703f.; Anderson 2005: 11f.).

2.2 *Synchronic*

Typological surveys have claimed that permutation of non-adjacent segments does not occur as a regular synchronic process (Webb 1974: 5; Wanner 1989: 445; Hume and Mielke 2001: 145f.). Certainly, the permutation of non-adjacent segments is common in speech errors, such as classic Spoonerisms, but such errors also involve strings of segments such as complex onsets and rhymes (Fromkin 1971: 31f.).

(23) *Metathesis in speech errors* (Fromkin 1971: 31f.)

intended		*error*	
kip ə teɪp	→	tip ə keɪp	'keep a tape'
fɑr mɔr	→	mɑr fɔr	'far more'
peɪ skeɪl	→	skeɪ peɪl	'pay scale'
swɛrə drajɪŋ	→	drɛrə swajɪŋ	'sweater drying'
hip ɔv ʤʌŋk	→	hʌŋk ɔv ʤip	'heap of junk'

These form part of a larger phenomenon of anticipation, perseveration, deletion, and so forth. Speech errors may, however, be a source of sporadic metathesis in historical change (Wanner 1989: 445). The outputs of speech errors, like more systematic metathesis, overwhelmingly respect the existing phonotactics of the language (Wells 1951: 26; Fromkin 1971: 40–42; Dell 1995: 200); but transposition of adjacent consonants, so common in regular metathesis, is "exceptionally rare" as a speech error, such as *whipser* for *whisper* (Berg 1987: 9). Elements that transpose by error usually occupy parallel syllable positions, which is not the case for adjacent consonants; instead, as discussed above, metathesis of such segments normally arises historically by misperception rather than production or planning errors.

An interesting comparison is an optional metathesis reported for a few words in Turkana (Dimmendaal 1983: 48f.; Hume and Mielke 2001: 139f.; Hume 2004: 218). Here two consonants with the same value of [sonorant] that serve as onsets to successive syllables, and are adjacent to identical vowels, are optionally transposed in fast speech.

(24) *Turkana onset metathesis* (Dimmendaal 1983: 48f.)

preferred	*alternate*	
ŋa-kɛ̀mɛ̀r-a	ŋa-kɛ̀rɛ̀m-a	'mole'
ŋi-kwaŋɔ̀rɔmɔk-à	ŋi-kwaŋɔ̀mɔrɔk-à	'a kind of tree'
ɛ-sɪkɪn-aˋ	ɛ-kɪsɪn-aˋ	'breast'

These alternations have the appearance of a common speech error that has become somewhat conventionalized. In particular, it has been widely observed that exchange (and other) errors are more likely when the sounds in question are found in similar phonological environments, so that for example *left hemisphere* → *heft lemisphere*, where the initial consonants are both followed by /ɛ/, is more likely than the parallel error in *right hemisphere*, where the vowels are different (MacKay 1970: 325–328; Dell 1984: 222).

Morphologically restricted metathesis (§1.2.2) can apply synchronically to surface non-adjacent segments. For example, Akkadian has a /t/ infixed in reciprocal verbs; it surfaces there in most cases, exemplified by [pitrus-] (25a), which motivates the stem-internal position as basic. But this stop is transposed to word-initial position when the root has an initial coronal obstruent as in (25b) (Caldwell *et al.* 1977: 118; McCarthy 1981: 381; Buccellati 1996: 233f.; Huehnergard 2005: 390, 531, 611).

(25) *Akkadian long-distance metathesis* (Caldwell *et al.* 1977; Huehnergard 2005)

	root	*infinitive*	*reciprocal*		
a.	/prs/	paraːs-	pitrus-		'divide'
	/rkb/	rakaːb-	ritkub-		'mount, ride'
	/kmr/	kamaːr-	kitmur-		'heap up'
b.	/sˤbt/	sˤabaːt-	tisˤbut-	*sˤitbut-	'seize'
	/snq/	sanaːq-	tisnuq-	*sitnuq-	'be near'
	/zkr/	zakaːr-	tizkur-	*zitkur-	'declare'
	/dkʃ/	dakaːʃ-	tidkuʃ-	*ditkuʃ-	'swell'

The same metathesis occurs in iterative stems: [pitarrus-] but [tisˤabbut-] (*[sˤitabbut-]). In the analysis of Łubowicz (2009), Akkadian metathesis serves to move the /t/ outside the stem domain, where it would cause a violation of the Obligatory Contour Principle (OCP) penalizing two tier-adjacent coronal consonants; thus [tizkur] does not violate this constraint, whereas *[zitkur] would, because the /t/ is located within the stem. This approach gives a relatively prominent role to phonology (the OCP constraint on coronals) while maintaining a crucial morphological component, due to the role of the stem domain.

2.3 *Displacement*

A related phenomenon, which is also called metathesis by many authors (Grammont 1933: 339; Ultan 1978: 372), involves the shift or *displacement* of a segment over more than one intervening segment. A famous example comes from the Occitan dialect of Bagnères-de-Luchon in southwestern France (Grammont 1905–6: 74, 85, 1933: 341; Blevins and Garrett 1998: 526). Among other processes, a liquid following a stop shifts leftward to form a cluster in the preceding syllable.

(26) *Bagnères-de-Luchon long-distance shift of liquids* (Grammont 1905–6: 74, 85, 1933: 341)

*ˈkabra	>	ˈkrabo	'goat'
*ˈbespras	>	ˈbrespes	'vespers'
*ˈpawpro	>	ˈprawbe	'poor'
*ˈtendro	>	ˈtrende	'tender'
*ˈkambra	>	ˈkrambo	'room'
*kumˈpra	>	krumˈpa	'to buy'
*eˈspingla	>	eˈsplingo	'pin'

A shift like this is formally identical to metathesis when just one segment is skipped. However, with intervening material that includes non-constituent strings such as /esp/, it must be movement of /r/ rather than exchange. A similar shift of /r/ to the initial syllable is attested in South Italian Greek (Rohlfs 1930; Blevins and Garrett 2004: 130f., 34f.).

If it is to be classified with long-distance metathesis, displacement might be expected to be absent from synchronic grammars. But synchronically active long-distance displacement is attested at least for laryngeal and pharyngeal features (Blevins and Garrett 2004: 132–134; see also CHAPTER 25: PHARYNGEALS). For example, in the Interior Salish language Colville (Nxilxcín), the pharyngeal consonant of a root is displaced to a stressed suffix, where it lowers the adjacent vowel to [a] (Mattina 1979).

(27) *Colville pharyngeal displacement* (Mattina 1979: 17f.)

ˈq’ʷˤaj	→	ˈq’ʷˤaj	‘black’
ˈq’ʷˤaj-ˈ**us**	→	q’ʷəjˈˤas	‘black man’
ˈq’ʷˤaj-ˈ**lstˣu**t	→	q’ʷəjlstˣˈˤat	‘his clothes are dirty’

Both pharyngealization and laryngealization can be seen as suprasegmental features in Salish (Mattina 1979: 19f.). Displacement of these features appears to reflect the spread of features over multiple syllables that may then be localized to a salient position (Blevins and Garrett 2004: 122f.). Such displacement resembles the *mobility* of a tone that shifts from the morpheme with which it is underlyingly affiliated to some phonologically defined position such as the penultimate syllable (Yip 2002: 65f., 89f., 132). A reasonable synchronic analysis is a [pharyngeal] feature affiliated with the root, which is attracted to the stressed syllable and possibly realized there as a segment. By the same token, the displacement of /r/ to the initial syllable in South Italian Greek reflects the salience and perceptual prominence of such syllables due to their location in the word (Blevins and Garrett 2004: 134); in Luchonnais, both stress and initial position favor the first syllable.

Patterns of this sort have similar historical origins to simple exchanges of segments – in particular, phonetic cues that are relatively long in the temporal dimension and therefore subject to re-analysis, as discussed above for perceptual metathesis. But since they are displacements rather than exchanges, they are not formally equivalent to true metathesis as the exchange of positions. In particular, if local metathesis is seen as a minimal displacement (across a single segment), then long-distance metathesis would have to involve two simultaneous displacements, one leftward and one rightward, as in /abXcdYef/ → [abYcdXef]. This extra formal complexity may account for the rarity of synchronic non-local metathesis, which seems to be restricted to limited examples such as the optional reversal in Turkana (24) and the morphologically defined environment in Akkadian (25).

An ordering alternation in the form of two suffixes is reported for several Costanoan languages, including Mutsun (Okrand 1979; Hume 1998: 170f., 1999: 300f., 2004: 223f.). Both suffixes have the shape CCV after a vowel-final stem, and CVC after a consonant-final stem, which makes phonotactic sense insofar as a three-consonant cluster at the stem boundary would be ill-formed.

(28) *Mutsun suffix alternations* (Okrand 1979: 127f., n. 17)

CCV		CVC	
pire-tka	‘on the ground’	ʔurkan-tak	‘in the mortar’
rukka-kma	‘houses’	wimmah-mak	‘wings’

Although the locative [tka] ~ [tak] can be treated as local metathesis, in the plural [kma] ~ [mak], the [k] appears to move across two other segments. If Linearity is gradiently violable, with one violation for each segment over which another is displaced, the minimal change (one ordering reversal) is generally optimal (Hume 1998: 168–171, 2001: 17–19). Gradient violation does still permit Mutsun /mak/ → [kma] when other constraints force multiple violations of

LINEARITY, and the alternation can be seen as part of the synchronic grammar; but such cases seem to be quite rare and limited in scope. In the more recent version of OT that incorporates candidate chains (OT-CC), changes to the representation occur by minimal steps, and non-contiguous metathesis is subject to the requirement that each change in linear order increase well-formedness (McCarthy 2007: 87f.). In a suffix alternation requiring the synchronic derivation /mak/ → mka → [kma], the first step might be motivated by a preference for sonorant codas, but that change does not appear to be motivated more generally in Mutsun; in fact, Hume (1998: 170f.) specifically gives the constraint *m]$_{coda}$. It might therefore be that the architecture of OT-CC forces the Mutsun alternation to be treated as listed allomorphy (CHAPTER 99: PHONOLOGICALLY CONDITIONED ALLOMORPH SELECTION).

3 Related processes

In some cases, a pattern that was originally considered to be metathesis was later seen as non-metathesis – and occasionally vice versa. A good example is the Zoque 3rd person singular prefix /j-/, which never surfaces as a strict prefix, but has been described as permuting with the stem-initial consonant, as in /j-pata/ → [pʲata] 'his mat' (Wonderly 1951: 117f.; Dell 1973: 110). Sagey (1986: 105–111) argues that the glide /j/ actually merges featurally with the following consonant to produce a palatalized segment (CHAPTER 71: PALATALIZATION), which may be realized with an offglide, as implied by transcriptions such as [pʲ]. This position is supported by the independent need for a non-metathesis source of palatalized segments found at compound boundaries; cases such as /kuj-tʌm/ → [kujtʲʌm] 'avocado' show that the glide spreads, rather than reversing in order. A similar pattern is found in several languages of Nigeria and Cameroon. Prefixes that can be reconstructed as the high vowels *i and *u are realized as a glide – or a secondary articulation on the consonant – after the stem-initial consonant, as in Noni /k-w-en/ 'firewood' from the base /ken/ (Blevins and Garrett 1998: 514–516). See also the cases in Webb (1974: 12f.).

Another phenomenon that has a certain affinity to metathesis is *infixation*, since it likewise requires a reordering from the expected position. In particular, "infixation and metathesis commonly show the potential mobility of full segments" rather than just subsegments such as features or nodes (Zoll 2001: 51). The closest analogy can be found in the infixation of a single consonant across one other consonant, as in the active neutral infix /-m-/ of Atayal /t-m-apeh/ 'beckon' (Egerod 1965: 265f.); this is formally similar to the metathesis of adjacent consonants. But infixation encompasses a broader set of phenomena that can include multiple segments in the item that undergoes reordering, as well as multiple segments in the span over which the infix is displaced; both are illustrated by the Tagalog actor focus /-um-/ that (optionally) moves over complex onsets in borrowed words such as /gr-um-adwet/ ~ /g-um-radwet/ 'graduate' (Orgun and Sprouse 1999: 204). On the other hand, Halle (2001) argues that the apparent Tagalog infixes appearing as /-um-/ and /-in-/ are actually CV underlyingly, with non-local metathesis of the two leftmost onsets, as /mu-tawag/ → [tu-mawag]. Theoretical and empirical problems with this approach are discussed by Klein (2005: 989–991), who advocates an infixation analysis within Optimality Theory.

Another phenomenon that might be seen as involving either metathesis or infixation (or even other possibilities) is *imbrication* in Bantu languages such as Cibemba (Hyman 1995). In this process, the perfective suffix /-il/ combines with a polysyllabic stem, such that the /l/ of the suffix disappears and the /i/ combines with the rightmost vowel in the stem according to the usual coalescence rules of the language, as in /sákat-il-e/ → [sákeete] 'seize'. The striking fact is that the suffixal vowel appears to skip over the stem-final consonant; in principle, this could be handled a variety of ways, including either CV metathesis, /sákat-il-e/ → sákaitle → [sákeete]; or infixation of the suffix inside the final consonant, /sáka-il-t-/ → sákailt-e → [sákeete]. These approaches assume subsequent simplification of the consonant cluster, as well as vowel coalescence. Hyman (1995: 11–16) argues in favor of infixation, which he relates to the positioning of the perfective (and the applicative) before the passive and causative suffixes.

Diachronically, metathesis is the origin of some instances of infixation (Ultan 1975: 178f.; Yu 2007: 139–148). Another point of comparison is found in Horwood (2002: 170, 2004: 11), who uses Linearity to control the displacement of prefixes and suffixes to infixed positions. A crucial difference is that infixation of this sort (that is, excluding infixation tied to prosodically prominent constituents) is inherently edge-oriented; the infixed material remains as close to the left or right edge of the stem as possible, subject to the phonotactic constraints or other pressures that force deviation from simple prefixation or suffixation (McCarthy and Prince 1993; Prince and Smolensky 2004: 40–43; Yu 2007: 67–71). Metathesis, on the other hand, often occurs at stem edges as the result of morpheme concatenation, but in principle can occur anywhere in a word – recall the stem-medial cases in Cebuano and Rendille (§1.1). In addition, the infix has the status of a morpheme, which may happen to consist of a single segment; but in metathesis the single-segment status is fundamental, and not necessarily correlated with a particular morpheme.

It can be noted finally that metathesis as a phenomenon is important evidence in favor of the category *segment*, however it may be formalized (chapter 54: the skeleton). Whether one considers the category of segment to be innate in the language faculty or something that emerges from the coordination of phonological gestures (Bybee 2001: 85f.), it is impossible to describe reorderings coherently in terms of disparate features or phonetic cues: the essential property of metathesis is that it moves all features associated with a segment, and the cues that instantiate these features are affected as a group. Indeed, the features may be implemented by rather different cues in the new position. For example, the Alsea alternation [stlak] ~ [stalk] affects just two of the five segments in this root. Even if /la/ were to be described as a core syllable, which is then reversed in some sense, the notion of "reversal" makes covert reference to the segments within the CV syllable. Otherwise there must be a claim that the prevocalic /l/ has the same phonetic realization as when it occurs in the coda, and that the release of the /t/ into the /a/ is no different from that into the /l/ in the nonmetathesized form. The need to refer to discrete segments even to characterize metathesis, and even more so to provide a theoretical analysis, presents particularly good evidence against suggestions that segments have no psychological reality, and are a mere artifact of an alphabetic writing system (Ladefoged 2005: 191; Silverman 2006: 6, 203).

REFERENCES

Anderson, John M. & Jacques Durand. 1988. Underspecification and dependency phonology. In Pier Marco Bertinetto & Michele Loporcaro (eds.) *Certamen phonologicum: Papers from the 1987 Cortona Phonology Meeting*, 3–36. Turin: Rosenberg & Sellier.

Anderson, Stephen R. 2005. Morphological universals and diachrony. *Yearbook of Morphology 2004*. 1–17.

Bagemihl, Bruce. 1995. Language games and related areas. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 697–712. Cambridge, MA & Oxford: Blackwell.

Bailey, Charles-James N. 1970. Toward specifying constraints on phonological metathesis. *Linguistic Inquiry* 1. 347–349.

Becker, Thomas. 2000. Metathesis. In Booij *et al.* (2000), 576–581.

Berg, Thomas. 1987. A cross-linguistic comparison of slips of the tongue. Bloomington: Indiana University Linguistics Club.

Blevins, Juliette & Andrew Garrett. 1998. The origins of consonant–vowel metathesis. *Language* 74. 508–556.

Blevins, Juliette & Andrew Garrett. 2004. The evolution of metathesis. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 117–156. Cambridge: Cambridge University Press.

Blust, Robert. 1971. A Tagalog consonant cluster conspiracy. *Philippine Journal of Linguistics* 2. 85–91.

Blust, Robert. 1979. Coronal–noncoronal consonant clusters: New evidence for markedness. *Lingua* 47. 101–117.

Bonthuis, Fiorieneke. 2001. Metathesis in Leti. In Hume *et al.* (2001), 26–52.

Booij, Geert, Christian Lehmann & Joachim Mugdan (eds.) 2000. *Morphologie: Ein internationales Handbuch zur Flexion und Wortbildung*. Berlin & New York: Mouton de Gruyter.

Buccellati, Giorgio. 1996. *A structural grammar of Babylonian*. Wiesbaden: Harrassowitz.

Buckley, Eugene. 2007. Vowel–sonorant metathesis in Alsea. *International Journal of American Linguistics* 73. 1–39.

Burrow, Thomas & Sudhibhushan Bhattacharya. 1970. *The Pengo language: Grammar, texts, and vocabulary*. Oxford: Clarendon Press.

Burton, Strang. 1989. Kasem coalescence and metathesis: A particle analysis. *Toronto Working Papers in Linguistics* 10. 21–32.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Caldwell, Thomas A., John N. Oswalt & John F. X. Sheehan. 1977. *An Akkadian grammar: A translation of Riemschneiders Lehrbuch des Akkadischen*. 3rd edn. Milwaukee: Marquette University Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Churchward, C. Maxwell. 1940. *Rotuman grammar and dictionary, comprising Rotuman phonetics and grammar and a Rotuman–English dictionary*. Sydney: Australasian Medical Publishing Company.

Coetzee, Andries W. 1999. Metathesis in Tiberian Hebrew: A perspective from Optimality Theory. *Theoretical Linguistics* 25. 99–131.

Crowhurst, Megan J. 1998. *Um* infixation and prefixation in Toba Batak. *Language* 74. 590–604.

Dell, François. 1973. *Les règles et les sons*. Paris: Hermann.

Dell, Gary S. 1984. The representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory and Cognition* 10. 222–233.

Dell, Gary S. 1995. Speaking and misspeaking. In Lila R. Gleitman & Mark Liberman (eds.) *An invitation to cognitive science*, vol. 1: *Language*, 183–208. 2nd edn. Cambridge, MA: MIT Press.

Demers, Richard. 1974. Alternating roots in Lummi. *International Journal of American Linguistics* 40. 15–21.

Dimmendaal, Gerrit Jan. 1983. *The Turkana language*. Dordrecht: Foris.

Egerod, Søren. 1965. Verb inflexion in Atayal. *Lingua* 15. 251–282.

Flemming, Edward. 1996. Laryngeal metathesis and vowel deletion in Cherokee. *UCLA Occasional Papers in Linguistics* 16. 23–44.

Foster, Michael. 1982. Alternating weak and strong syllables in Cayuga words. *International Journal of American Linguistics* 48. 59–72.

Fromkin, Victoria A. 1971. The non-anomalous nature of anomalous utterances. *Language* 47. 27–52.

Garrett, Andrew & Juliette Blevins. 2009. Analogical morphophonology. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Essays in honor of Paul Kiparsky*, 527–545. Cambridge, MA: MIT Press.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Oxford & Cambridge, MA: Blackwell.

Grammont, Maurice. 1905–6. La métathèse dans le parler de Bagnères-de-Luchon. *Mémoires de la societé de linguistique de Paris* 13. 73–90.

Grammont, Maurice. 1933. *Traité de phonétique*. Paris: Delagrave.

Haas, Wim G. de. 1988. Phonological implications of skeleton and feature underspecification in Kasem. *Phonology* 5. 237–254.

Hale, Kenneth. 1976. Phonological developments in particular Northern Paman languages. In Peter Sutton (ed.) *Languages of Cape York*, 7–40. Canberra: Australian Institute of Aboriginal Studies.

Hale, Mark & Madelyn Kissock. 1998. The phonology–syntax interface in Rotuman. In Matthew Pearson (ed.) *Recent papers in Austronesian linguistics*, 115–128. Los Angeles: Department of Linguistics, University of California, Los Angeles.

Halle, Morris. 2001. Infixation versus onset metathesis in Tagalog, Chamorro, and Toba Batak. In Michael Kenstowicz (ed.) *Ken Hale: A life in language*, 153–168. Cambridge, MA: MIT Press.

Halle, Morris & Valdis Zeps. 1966. A survey of Latvian morphophonemics. *MIT Research Laboratory of Electronics Quarterly Progress Report* 83. 104–113.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hock, Hans Henrich. 1985. Regular metathesis. *Linguistics* 23. 529–546.

Holscher, Dan, Monica Macaulay & Marnie Jo Petray. 1992. Tone metathesis in the Dangme imperative. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 17. 20–33.

Horowitz, Edward. 1960. *How the Hebrew language grew*. New York: Jewish Educational Committee Press.

Horwood, Graham. 2002. Precedence faithfulness governs morpheme position. *Proceedings of the West Coast Conference on Formal Linguistics* 21. 166–179.

Horwood, Graham. 2004. Order without chaos: Relational faithfulness and position of exponence in Optimality Theory. Ph.D. dissertation, Rutgers University.

Hudson, Grover. 1980. Automatic alternations in nontransformational phonology. *Language* 56. 94–125.

Huehnergard, John. 2005. *A grammar of Akkadian*. 2nd edn. Winona Lake: Eisenbrauns.

Hume, Elizabeth. 1998. Metathesis in phonological theory: The case of Leti. *Lingua* 104. 147–186.

Hume, Elizabeth. 1999. The role of perceptibility in consonant/consonant metathesis. *Proceedings of the West Coast Conference on Formal Linguistics* 17. 293–307.

Hume, Elizabeth. 2001. Metathesis: Formal and functional considerations. In Hume *et al.* (2001), 1–25.

Hume, Elizabeth. 2004. The indeterminacy/attestation model of metathesis. *Language* 80. 203–237.

Hume, Elizabeth & Jeff Mielke. 2001. Consequences of word recognition for metathesis. In Hume *et al.* (2001), 135–158.

Hume, Elizabeth & Misun Seo. 2004. Metathesis in Faroese and Lithuanian: From speech perception to Optimality Theory. *Nordic Journal of Linguistics* 27. 35–60.

Hume, Elizabeth, Norval Smith & Jeroen van de Weijer (eds.) 2001. *Surface syllable structure and segment sequencing*. Leiden: Holland Institute of Generative Linguistics.

Hyman, Larry M. 1985. Word domains and downstep in Bamileke-Dschang. *Phonology Yearbook* 2. 47–83.

Hyman, Larry M. 1995. Minimality and the prosodic morphology of Cibemba imbrication. *Journal of African Languages and Linguistics* 16. 3–39.

Hyman, Larry M. & William R. Leben. 2000. Suprasegmental processes. In Booij *et al.* (2000), 587–594.

Jakobi, Angelika. 1990. *A Fur grammar: Phonology, morphophonology and morphology*. Hamburg: Buske.

Jespersen, Otto. 1949. *A Modern English grammar on historical principles*. Part I: *Sounds and spellings*. London: George Allen & Unwin.

Keyser, Samuel J. 1975. Metathesis and Old English phonology. *Linguistic Inquiry* 6. 377–411.

Kilani-Schoch, Marianne & Wolfgang Dressler. 1986. Métathèse et conversion morphologiques en arabe tunisien. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39. 61–75.

Kiparsky, Paul & Wayne O'Neil. 1976. The phonology of Old English inflections. *Linguistic Inquiry* 7. 527–557.

Klein, Thomas B. 2005. Infixation and segmental constraint effects: UM and IN in Tagalog, Chamorro, and Toba Batak. *Lingua* 115. 959–995.

Kurisu, Kazutaka. 2001. The phonology of morpheme realization. Ph.D. dissertation, University of California, Santa Cruz.

Ladefoged, Peter. 2005. *Vowels and consonants: An introduction to the sounds of languages*. 2nd edn. Malden, MA & Oxford: Blackwell.

Langdon, Margaret. 1976. Metathesis in Yuman languages. *Language* 52. 866–883.

Lockwood, W. B. 1955. *An introduction to modern Faroese*. Copenhagen: Munksgaard.

Łubowicz, Anna. 2009. Infixation as morpheme absorption. In Steve Parker (ed.) *Phonological argumentation: Essays on evidence and motivation*, 261–284. London: Equinox.

MacKay, Donald G. 1970. Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia* 8. 323–350.

Malone, Joseph L. 1971. Systematic metathesis in Mandaic. *Language* 47. 394–415.

Malone, Joseph L. 1993. *Tiberian Hebrew phonology*. Winona Lake, IN: Eisenbrauns.

Mattina, Anthony. 1979. Pharyngeal movement in Colville and related phenomena in the interior Salishan languages. *International Journal of American Linguistics* 45. 17–24.

McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12. 373–418.

McCarthy, John J. 1989. Linear order in phonological representation. *Linguistic Inquiry* 20. 71–99.

McCarthy, John J. 2000. The prosody of phase in Rotuman. *Natural Language and Linguistic Theory* 18. 147–197.

McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.

McCarthy, John J. & Alan Prince. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory* 8. 209–283.

McCarthy, John J. & Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology 1993*. 79–153.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.

Montler, Timothy. 1989. Infixation, reduplication, and metathesis in the Saanich actual aspect. *Southwest Journal of Linguistics* 9. 92–107.

Noonan, Michael. 1997. Inverted roots in Salish. *International Journal of American Linguistics* 63. 475–515.

Okrand, Marc. 1979. Metathesis in Costanoan grammar. *International Journal of American Linguistics* 45. 123–130.

Orgun, Cemil Orhan & Ronald L. Sprouse. 1999. From MParse to Control: Deriving ungrammaticality. *Phonology* 16. 191–224.

Osthoff, Hermann & Karl Brugmann. 1878. *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*. Leipzig: Hirzel.

Penny, Ralph J. 2002. *A history of the Spanish language*. 2nd edn. Cambridge: Cambridge University Press.

Phelps, Elaine. 1975. Simplicity criteria in generative phonology: Kasem nominals. *Linguistic Analysis* 4. 297–332.

Phelps, Elaine. 1979. Abstractness and rule ordering in Kasem: A refutation of Halle's maximizing principle. *Linguistic Analysis* 5. 29–69.

Pierret, Jean-Marie. 1994. *Phonétique historique du français et notions de phonétique générale*. Louvain-la-Neuve: Peeters.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Prunet, Jean-François. 2006. External evidence and the Semitic root. *Morphology* 16. 41–67.

Prunet, Jean-François, Renée Béland & Ali Idrissi. 2000. The mental representation of Semitic words. *Linguistic Inquiry* 31. 609–648.

Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology*. Dordrecht: Reidel.

Rohlfs, Gerhard. 1930. *Etymologisches Wörterbuch der unteritalienischen Gräzität*. Halle, Saale: Niemeyer.

Rubin, Dominic. 2001. Place-licensing: Opposite orderings are cognitive, not phonetic. In Hume *et al.* (2001), 189–209.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.

Sandler, Wendy. 1993. A sonority cycle in American Sign Language. *Phonology* 10. 243–279.

Schachter, Paul & Fe Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.

Seo, Misun & Elizabeth Hume. 2001. A comparative account of metathesis in Faroese and Lithuanian. In Hume *et al.* (2001), 210–229.

Silva, Clare M. 1973. Metathesis of obstruent clusters. *OSU Working Papers in Linguistics* 14. 77–84.

Silverman, Daniel. 2006. *A critical introduction to phonology: Of sound, mind, and body*. London & New York: Continuum.

Sim, Ronald J. 1981. Morphophonemics of the verb in Rendille. *Afroasiatic Linguistics* 8. 1–33.

Smith, Norval. 1985. Spreading, reduplication and the default option in Miwok nonconcatenative morphology. In Harry van der Hulst & Norval Smith (eds.) *Advances in nonlinear phonology*, 363–380. Dordrecht: Foris.

Sohn, Ho-min. 1980. Metathesis in Kwaraae. *Lingua* 52. 305–323.

Steriade, Donca. 1990. Gestures and autosegments: Comments on Browman and Goldstein's paper. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 382–397. Cambridge: Cambridge University Press.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 219–250. San Diego: Academic Press.

Stonham, John T. 1994. *Combinatorial morphology*. Amsterdam & Philadelphia: John Benjamins.

Thompson, Laurence C. & M. Terry Thompson. 1969. Metathesis as a grammatical device. *International Journal of American Linguistics* 35. 213–219.

Ultan, Russell. 1975. Infixes and their origins. In Hansjakob Seiler (ed.) *Linguistic workshop III*, 157–205. Munich: Fink.

Ultan, Russell. 1978. A typological view of metathesis. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 2: *Phonology*, 367–402. Stanford: Stanford University Press.

Ussishkin, Adam. 2005. A fixed prosodic theory of nonconcatenative templatic morphology. *Natural Language and Linguistic Theory* 23. 169–218.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin: Mouton de Gruyter.

Wanner, Dieter. 1989. On metathesis in diachrony. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 25. 434–450.

Webb, Charlotte. 1974. Metathesis. Ph.D. dissertation, University of Texas, Austin.

Wechssler, E. 1900. Giebt es Lautgesetze? *Forschungen zur romanischen Philologie: Festgabe für Hermann Suchier*, 349–538. Halle: Niemeyer.

Wells, Rulon. 1951. Predicting slips of the tongue. *Yale Scientific Magazine*, December, 9–30.

Williams, Edwin B. 1962. *From Latin to Portuguese*. 2nd edn. Philadelphia: University of Pennsylvania Press.

Winters, Steve. 2001. VCCV perception: Putting place in its place. In Hume *et al.* (2001), 230–247.

Wonderly, William L. 1951. Zoque II: Phonemes and morphophonemes. *International Journal of American Linguistics* 17. 105–123.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Yu, Alan C. L. 2007. *A natural history of infixation*. Oxford: Oxford University Press.

Zoll, Cheryl. 2001. Constraint and representations in subsegmental phonology. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 46–78. Cambridge: Cambridge University Press.

60 Dissimilation

Patrik Bye

1 Introduction

Dissimilation prototypically refers to a situation in which a segment becomes less similar to a nearby segment with respect to a given feature. As a synchronic alternation, it can be exemplified by liquid dissimilation in Georgian, where the ethnonym-forming suffix {-uri} becomes [uli] when an /r/ precedes it anywhere within the word (Fallon 1993; Odden 1994). The resulting pattern of alternation is shown in (1).

(1) *Georgian r-dissimilation*

a.	p'olon-uri	'Polish'
	somχ-uri	'Armenian'
b.	sur-uli	'Assyrian'
	p'rusi-uli	'Prussian'
c.	avst'ral-uri	'Australian'
	kartl-uri	'Kartvelian'

In (1a), the suffix surfaces in its basic (non-dissimilated) form. The forms in (1b) illustrate the result of unbounded dissimilation within the word. In the word meaning 'Prussian' it takes place despite the presence of the intervening consonant. If a lateral /l/ intervenes between the two rhotics, however, dissimilation does not apply. This is shown in (1c).

A very similar example of liquid dissimilation comes from Latin (e.g. Kent 1936, 1945; Steriade 1987), where the alternation is reversed. The adjectival suffix *-ālis*, as in *nāvālis* 'naval', dissimilates to *-āris* whenever another /l/ precedes it in the word, e.g. *lūnāris* 'lunar'. Dissimilation is similarly blocked whenever /r/ intervenes between the trigger and the target, e.g. *flōrālis* 'floral', **flōrāris*.

As a diachronic change, dissimilation is most often sporadic, applying to random lexical items (Posner 1961). The historical development of Latin and the Romance languages furnish several examples of sporadic liquid dissimilation, e.g. Latin *arbor* > Spanish *arbol* 'tree', *peregrīnus* > Late Latin *pelegrīnus* 'pilgrim'. Regular synchronic alternations involving dissimilatory processes are far more rare and, as a result, dissimilation has been afforded somewhat less systematic attention than other more common segmental patterns like assimilation. Nevertheless, the study of dissimilation phenomena offers a valuable source of insights into the fundamental questions phonologists ask. These questions include (a) the nature of rules and representations, and the relation between the two, (b) the division of labor between the grammar and the lexicon, and (c) whether phonological patterns reflect possibly innate cognitive biases or extralinguistic factors operating during acquisition.

The organization of the remainder of this chapter is as follows. §2 sets out the major parameters of dissimilation, explaining which features participate, along with restrictions on the interaction of context and focus determined by locality and domain of application. §3 addresses the contribution that the study of dissimilation phenomena has made to phonological theory, assessing how it has shaped our understanding of both representations and rules. §4 provides an overview of the motivations for dissimilatory patterns proposed in the literature. Conclusions and questions for future research are given in §5.

2 Dissimilatory patterns and their parameters

2.1 *Participating features*

Suzuki (1998) presents a comprehensive survey of cross-linguistic dissimilatory patterns. His survey includes 39 dissimilatory alternations. Table 60.1 provides a somewhat revised summary with a total of 46 alternations, adducing a few additional cases not covered in Suzuki's original survey, and suppressing cases that on closer inspection turn out not to be true dissimilation.[1] The second column of the table specifies the locality condition on the process (for illustration

[1] For example, Suzuki's cases 54–57 are grouped under "polarity", but on closer inspection they appear to have little in common. The reasons for reclassifying or not including these cases here are, briefly, as follows. In Russian *jakan'e* (54) a pre-tonic non-high vowel reduces to [i] or [a], depending on the quality of the following stressed vowel. The high or low quality of the reduced vowel gives the impression of maximizing the contrast in vowel height, e.g. /sʲeˈmʲju/ → [sʲaˈmʲju] 'seven (INST)', but /dʲiˈsʲatka/ → /dʲiˈsʲatka/ 'tenfold'. Crosswhite (1999: 79–83), however, argues that the dissimilatory effect is only epiphenomenal, and actually has nothing to do with dissimilation or enhancement of vowel height contrast. Based on work by Alderete (1995), she argues that what is at issue is actually a difference in foot structure. Syllables with prominent nuclei may constitute feet on their own. In [dʲi(ˈsʲa$_{\mu\mu}$t)ka], the stressed syllable forms a foot on its own, whereas in [(sʲaˈmʲju)] the pre-tonic syllable must also be incorporated because the stressed nucleus is not sufficiently prominent. The choice of raising or lowering thus comes to depend on whether the focus is parsed into a foot or not. Dinka (55) represents a morphological exchange rule, which is highly controversial in linguistic theory. See Wolf (2007) for an alternative analysis of exchange rules in terms of featural affix allomorphs. Huamelultec Chontal (56) appears to be a case of [spread glottis] dissimilation, not polarity. Margi (57) represents a case of allomorphy. Also not included is Thurneysen's Law in Gothic, which recent research by Woodhouse (1998) shows to be a case of analogical relexicalization rather than a phonological rule. Suzuki also includes Finnish consonant gradation (Keyser and Kiparsky 1984; Alderete 1997), but it is excluded here, since it is neither synchronically properly phonological nor obviously a dissimilation rule. One language cited as evincing low vowel dissimilation is the Chadic language Kera (Ebert 1974; Kenstowicz and Kisseberth 1979), where /a/ is claimed to dissimilate to [ə] preceding another /a/. Recent work by Pearce (2008), however, shows that the effect is due to reduction in unstressed syllables.

Table 60.1 Dissimilatory alternations (Rt = root adjacency; σ = syllable adjacency; unbounded (within the word); P = progressive; R = regressive)

Feature	*Locality*	*Example*
labial	Rt	Tashlhiyt Berber (P: Elmedlaoui 1985, 1995; Jebbour 1985; Boukous 1987; Larsi 1991; Odden 1994)
	σ	Cantonese (P: Yip 1982, 1988), Palauan (R: Josephs 1975, 1990; Finer 1986)
	unbounded	Akkadian (R: Soden 1969; McCarthy 1979; Yip 1988; Hume 1992; Odden 1994), Tashlhiyt Berber (2xR; see references above), Palauan (R; see references above)
coronal	Rt	Dakota (R: Shaw 1976, 1985)
lateral	unbounded	Kuman (R: Trefry 1969; Walsh Dickey 1997), Latin (P: Kent 1936, 1945; Posner 1961; Johnson 1973; Steriade 1987, 1995; Odden 1994; Walsh Dickey 1997), Yidiɲ (R: Dixon 1977; Steriade 1995; Walsh Dickey 1997), Yimas (P: Foley 1991; Odden 1994; Walsh Dickey 1997)
rhotic	Rt	Ainu (R: Maddieson 1984; Shibatani 1990)
	unbounded	Georgian (P: Fallon 1993; Odden 1994), Modern Greek (R: Newton 1971; Walsh Dickey 1997), Sundanese (R: Cohn 1992; Holton 1995), Yindjibarndi (P: Wordick 1982)
voice	σ	Bantu (R: Bennett 1967; Davy & Nurse 1982; Odden 1994; Lombardi 1995), Japanese (Itô & Mester 1986; Alderete 1997)
spread glottis	σ	Ancient Greek (R: Grassmann 1863), Huamelultec Chontal (P: Waterhouse 1949, 1962; Kenstowicz & Kisseberth 1979), Inari Saami (P: Itkonen 1986–91)
constricted glottis	σ	Seri (R: Marlett & Stemberger 1983; Yip 1988), Cuzco Quechua (R: Parker 1997)
nasal	Rt	Chukchi (R: Odden 1987)
NC	σ	Gooniyandi (P: McGregor 1990; Odden 1994; Evans 1995)
	unbounded	Gunindji (P: McConvell 1988; Odden 1994; Evans 1995), Yindjibarndi (P: Wordick 1982; Odden 1994)
continuant	Rt	Modern Greek (P: Kaisse 1988), Northern Greek (R: Newton 1971), Tsou (P: Szakos 1994), (P: Quintero 2004), North Central Spanish (R: González 2008)
high	σ	Guere (R: Paradis & Prunet 1989)
	Rt	Arusa (R: Levergood 1987), Wintu (Pitkin 1984)
low	σ	Marshallese (R: Bender 1968, 1969; Kenstowicz & Kisseberth 1977), Woleaian (R: Sohn 1975; Sohn & Tawerilmang 1976; Poser 1982)
length	σ	Gidabal (P: Geytenbeek & Geytenbeek 1971), Latin (R: Leumann 1977; Sihler 1995), Oromo (P: Gragg 1976; Lloret 1988; Alderete 1997), Slovak (P: Kenstowicz & Kisseberth 1977, 1979; Rubach 1993)
H	Rt	Bantu (P: Goldsmith 1984; Odden 1994)
	σ	Bantu (P: Goldsmith 1984; Odden 1994)
	unbounded	Arusa (P: Levergood 1987; Odden 1994)
L	unbounded	Peñoles Mixtec (P: Daly 1993; Odden 1994)

of dissimilations with different locality conditions, see §2.2). Also indicated is the direction of dissimilation and, in case the language has more than one, the number of dissimilative patterns. The numbers of progressive and regressive dissimilations are more or less evenly split, with 21 and 24 cases respectively.

Major class features such as [consonantal], [sonorant], and [approximant] do not appear to participate in dissimilation. The existence of some of these features is indeed contested in the literature. While there is something of a consensus that [sonorant] is necessary,[2] Hume and Odden (1996) propose that [consonantal] may be dispensed with. There is also a widespread assumption that the feature [approximant] is not contrastive in language, although see Levi (2008) for evidence to the contrary.

Beyond the major class features, all classes of feature may be involved in dissimilation, including place of articulation, laryngeal state, manner (continuancy, liquid, nasality), vowel height, and suprasegmental properties such as length and tone. We shall illustrate some of these in this section; other alternations that raise particular theoretical issues will be illustrated in the relevant sections. Thus, see §3 for examples of nasal dissimilation, and §4.3 for continuant dissimilation.

Of the place of articulation features, only [labial] dissimilation is common. Labial dissimilation is illustrated in (2) with data from Tashlhiyt Berber (Odden 1994). The labial nasal /m/ in a prefix dissimilates to [n] if the stem contains a labial consonant anywhere within it.

(2) *Labial dissimilation in Tashlhiyt Berber*

a.	las	'shear'	am-las	'shearer'
	agur	'remain'	am-agur	'abandoned'
b.	ˤrmi	'be tired'	an-ˤrmi	'tired person'
	bur	'remain celibate'	an-bur	'bachelor'
	ˤazum	'fast'	an-ˤazum	'faster'

Dissimilation of [coronal] is only attested in a single case, Dakota (Shaw 1976, 1985). Underlying coronal non-continuants /t ʧ n d/ are all neutralized to [k] (or, with regressive voicing assimilation, [g]) before another coronal consonant. The examples in (3) are from Shaw (1985: 184); see also Shaw (1976: 337).

(3) *Coronal dissimilation in Dakota reduplication*

a.	/ʧek/	ʧek-ˈʧeka	'stagger'
	/ʧap/	ʧap-ˈʧapa	'trot'
	/tʼĩs/	tʼĩs-ˈtĩza	'draw tight'
	/khuʃ/	khuʃ-ˈkhuʒa	'lazy'
b.	/sut/	suk-ˈsuta	'strong'
	/ʒat/	ʒag-ˈʒata	'curved'
	/tʰeʧ/	tʰek-ˈtʰeʧa	'be new'
	/ʧʰeʧ/	ʧʰek-ˈʧʰeʧʼa	'to look like'
	/nin/	nig-ˈnina	'very'

2 This consensus naturally does not extend to those representational theories like Government Phonology, where the elements of representation must have autonomous interpretations (see especially Kaye *et al.* 1985). Obviously, [sonorant] has no phonetic interpretation independent of the place and manner features with which it is associated.

There are apparently no attested examples of dissimilation involving the feature [dorsal].

Alternations involving laterals and rhotics are relatively common. We shall provide examples of liquid dissimilation in §2.2 in connection with the discussion of locality parameters.

Several Australian languages show dissimilation of prenasalized stops or nasal + stop clusters (NC). In Gurindji (Pama-Nyungan, Northern Territory; McConvell 1988; Odden 1994), this process is unbounded, e.g. /lutcu-ŋka/ 'ridge', /pinka-ŋka/ → [pinka-ka] 'river-LOC', /kankula-mpa/ → [kankula-pa] 'high ground-LOC'.

Several languages have restrictions on consecutive heavy nuclei that do not appear reducible to prosodic structure. In Slovak, for example, a long nucleus becomes short following a long nucleus, according to a rule known as the Rhythmic Law (Rubach 1993: 172–175). Thus, the suffixes {-aː} (FEM SG) and {-eːmu} (DAT SG) shorten their first vowel following a long vowel, e.g. [ma'l-a] 'small-FEM SG' *vs.* ['mlːkv-a] 'silent-FEM SG'; [ma'l-emu] 'small-MASC DAT SG' *vs.* ['mlːkv-emu] 'silent-MASC DAT SG'. The alternation is apparently unrelated to stress (Rubach 1993: 41–42). At least in Western Slovak, the main stress falls on the initial syllable of the word, and sources report a binary stress pattern, some with the possibility of ternary alternation. The Rhythmic Law nevertheless applies in odd-numbered syllables, where we would expect resumption of secondary stress on a binary alternating pattern. This is shown by derivations with the agentive suffix {-niːk} and the diminutive {-ik}, e.g. [hutniːk] 'steelworker' *vs.* [ʧaluːnik] 'wallpaperer', [xlebiːk] 'bread' *vs.* ['ʤbaːnik] 'pot'.

Several languages of Vanuatu have productive Low Vowel Dissimilation (Lynch 2003). In Maskelynes (Malayo-Polynesian), the nominalizer prefix is realized as [nə-] when the following vowel is low /e a o/, and [na-] following a high vowel /i ə u/.

(4)	a.	na-vis	'banana'
		na-xəmar	'men's house'
		na-xut	'louse'
	b.	nə-matu	'right (hand)'
		nə-gor	'green coconut'

Dissimilation is occasionally also used to refer to the deletion of one of a pair of similar neighboring sounds. Hall (2009), for example, describes this phenomenon with reference to /r/ in American English, in principle giving alternations like [fɑɻm] *farm* *vs.* [fɑmɚ] *farmer*, and [istən] *eastern* *vs.* [istənɚ] *easterner*.

All the cases we have looked at so far involve the elimination of sequences of similar sounds. Preventive dissimilation is when the creation of new sequences of similar sounds is blocked. One example is provided by Inari Saami (Itkonen 1986–91), which has a morphologically conditioned process of consonant gradation. An overlong obstruent in the "strong" grade generally alternates with the corresponding singleton in the "weak" grade, as shown in (5). In each of the examples below, the strong grade form on the left represents the nominative singular, the weak grade form on the right the accusative-genitive singular. Examples have been adapted from Finno-Ugric transcription into IPA, according to the conventions set out in Bye *et al.* (2009) ([ʌ] is a somewhat low central vowel; [ʌ̆] is "ultrashort").

(5) *Inari Saami consonant gradation (obstruents)*

tsuop:pʰă	tsuopʰʌ	'meat of fish'
fat:tă	faatʌ	'yard'
ɲeʃ:ʃi	ɲeeʃi	'mud, slush'

Under normal circumstances, the overlong aspirated velar stop /k:kʰ/ alternates with /h/, as shown in (6). This may be taken to reflect a general process that debuccalizes /kʰ/, leaving bare [h].

(6) *Inari Saami consonant gradation: Debuccalization of aspirated velar stop*

kak:kʰu	kaahu, *kaakʰu	'unleavened rye-bread'
ʧok:kʰi	ʧohii, *ʧokʰii	'peak, summit'
kal:kkʰă	kaalhʌ, *kaalkʰʌ	'chalk'

However, there is one situation where debuccalization to [h] fails to take place. As (7) shows, this is when the onset of the preceding syllable is also /h/.

(7) *Inari Saami consonant gradation: Debuccalization blocked*

hik:kʰi	hiikʰi, *hiihi	'hay-basket'
hak:kʰă	haakʰʌ, *haahʌ	'canon'
hul:kkʰă	huulkʰʌ, *huulhʌ	'knife-sheath'

2.2 Locality and domains

Dissimilation may be associated with one of three locality conditions listed in (8). This parameter was first addressed in detail by Odden (1994). Suzuki's (1998) survey largely confirms this picture.

(8) *Locality conditions*

a. Root adjacency
b. Syllable adjacency
c. Unbounded

Liquid dissimilation may be used to exemplify all three locality conditions. Ainu, a language isolate of Japan, illustrates the root-adjacency condition (Shibatani 1990: 13). Given an underlying cluster /rr/, the first /r/ dissimilates to [n], as shown in (9).

(9) *Ainu* r-*dissimilation*

kukor kur	'my husband'	kukon rusuj	'I want to have (something)'
kor mat	'his wife'	kon rametok	'his bravery'

Yimas (Foley 1991: 54), a Sepik-Ramu language of Papua New Guinea, illustrates liquid dissimilation operating under syllable adjacency. An /r/ dissimilates to [t] if there is an /r/ in the immediately preceding syllable. The examples in (10) show variation in the shape of the inchoative suffix {-ara} (1991: 290).

(10) *Yimas ɾ-dissimilation*

tuak-ara-	'break open'
kamprak-ara-	'snap'
apr-ata-	'open, spread'

Dissimilation may also be unbounded within the word, as we have already seen in the Georgian example in (1) with which we opened this chapter.

3 Dissimilation in the grammar

One of the fundamental issues in generative phonology has always been whether linguistically significant generalizations should be assigned to particular designated levels of representation, such as the underlying or surface level, or to the rules that map one representation onto another (McCarthy 2007; Bye, forthcoming). In earlier approaches in the style of *SPE* (Chomsky and Halle 1968), dissimilations were described in terms of feature-changing rules of the general form shown in (11).

(11) $X \rightarrow [-F] \; / \; \_\_ \; [+F]$

Rules of this kind were criticized because the pairing of structural change and environment was arbitrary. Because of this, they were unable to distinguish between natural assimilations like (12a) and arbitrary rules like (12b) (example from Odden 1987).

(12) a. [+consonantal] → [+voice] / __ [+voice]
b. [+consonantal] → [+voice] / __ [+continuant]

Concerns about the generative power of feature-changing rules thus motivated the development in the mid-1970s and 1980s of non-linear approaches to phonological representation (Goldsmith 1976) that permitted greater elegance and simplicity in the statement of natural rules. Assimilation rules were remodeled as feature-filling spreading (see CHAPTER 81: LOCAL ASSIMILATION and CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS). In non-linear terms, dissimilation is simply the deletion or delinking of a feature and, in accounts that retain a view of features as binary, independently motivated insertion of a default value (Odden 1987, 1994; CHAPTER 27: THE ORGANIZATION OF FEATURES). For example, Chukchi has a process changing underlying /ŋ/ to [ɣ] before another nasal, shown in (13).

(13)

taraŋ-ək	'build a dwelling'	nə-taraɣ-mori	'we built a dwelling'
inawrəŋ-ək	'to give as a gift'	inawrəɣ-nin	'he gave it'
pitˀiŋ	'cold'	pitˀiɣ-ŋinqij	'boy with a cold'

Odden (1987: 242) provides an analysis of this alternation as delinking of [+nasal] before another [+nasal], as shown in (14). Subsequently, redundancy rules fill in the feature [−nasal] by default.

```
(14)          denasalization              default
    inawrəŋ-ən    →   inawrəK-ən     →   inawrəɣ-ən
          |   |                 |                |   |
      [+N][+N]               [+N]            [−N][+N]
```

Accompanying the development of non-linear representations was a return to the idea that at least certain phonological generalizations are best stated as constraints on surface forms. This conception made it feasible to explain how it was possible that certain rules seemed to share the same functional teleology (the "conspiracy" problem; see Kisseberth 1970 and CHAPTER 70: CONSPIRACIES). Such constraints could trigger the application of repairs, such as the deletion of the first of the two [+nasal] features, or block the application of rules that would otherwise apply (see examples of preventive dissimilation in §2). The first such constraint on output representations was the Obligatory Contour Principle (Leben 1973; Goldsmith 1976; McCarthy 1979, 1981, 1986, 1988; Odden 1988; Yip 1988, 1989), one formulation of which is provided in (15).

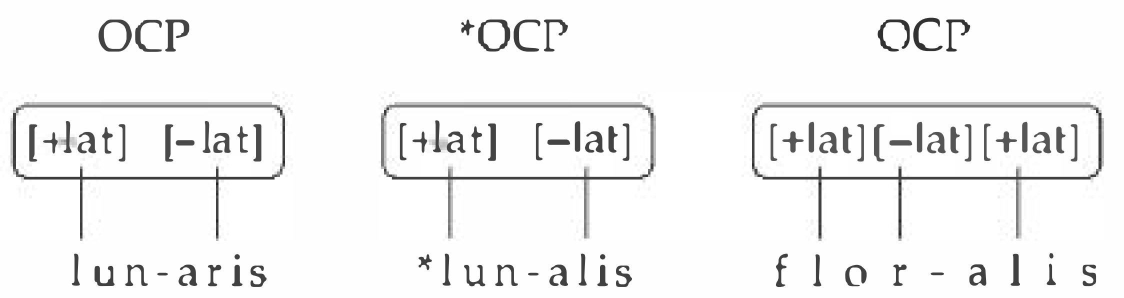
(15) *Obligatory Contour Principle (OCP)* (McCarthy 1986, 1988)

At the melodic level, adjacent identical elements are prohibited.

The OCP was originally used in accounting for tonal phenomena, especially adjacency of high tones, but it was subsequently extended to include other features. The OCP specifies a negative output target, and dissimilation only represents one strategy for satisfying it. Other repair strategies include merger of adjacent identical nodes, blocking of syncope (McCarthy 1986), and the insertion of epenthetic segments (Yip 1988). The OCP was incorporated into work couched in the framework of Optimality Theory (OT: McCarthy and Prince 1993; Prince and Smolensky 1993), where it became a violable constraint. Alderete (1997) and Itô and Mester (2003) propose that the OCP may represent a local self-conjunction of more primitive markedness constraints (Smolensky 1995, 1997). $\text{OCP}_{[F]}$ is violated precisely when *[F] is violated more than once within some local domain.

Another major theoretical concern during the 1980s especially was locality conditions on the application of rules, and dissimilation played a major part in this debate. The autosegmentalization of representations into tiers permitted the elimination of many kinds of apparent long-distance effects. Sounds that are non-adjacent on the level of the segmental root may nevertheless dissimilate, provided that the relevant features are adjacent on the same autosegmental tier. Steriade (1987) argues that the Latin facts mentioned at the beginning of this chapter may be accounted for by a version of the OCP with jurisdiction over the [lateral] tier, over which interactions between tier-adjacent liquids may be described. The diagrams in (16) show how the liquids in the words *lūnāris* and *flōrālis* are projected onto separate tier-specifying values of [lateral]. This allows us to explain the ungrammaticality of the counterfactual form **lūnālis* as a result of the two occurrences of [+lateral] being adjacent on the [lateral] tier, in violation of $\text{OCP}_{[\text{lateral}]}$. In *flōrālis*, on the other hand, there is an intervening [−lateral] between the two occurrences of [+lateral], so the OCP is not violated.

(16) *Lateral dissimilation*



Even with the possibility of factoring the representation into tiers, though, there is still an empirical residue that poses a problem for a strict interpretation of locality. In some theories, vocalic place (V-Place) and consonantal place (C-Place) are represented on separate planes (Clements 1991; Clements and Hume 1995; Morén 2003; see also CHAPTER 19: VOWEL PLACE and CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION). This organization implies that non-adjacent consonants and vowels should not display any interaction, but this expectation is not borne out. Akkadian (Soden 1969: 64ff.), for example, has a nominalizer prefix |ma-| that dissimilates to [na-] if followed by a labial consonant in the stem, e.g. /ma-šʔal-t-u/ 'question' but /ma-rkab-t/ → [na-rkab-t] 'chariot'. If a labial vowel or glide intervenes between the trigger and the target, however, dissimilation is blocked, e.g. /ma-wmii-t-um/ → *[ma-amii-t-um] 'oath'. Odden (1994: 319) argues for an additional adjacency parameter, transplanar locality, to cover these cases, but it is unclear how this is to be formalized.

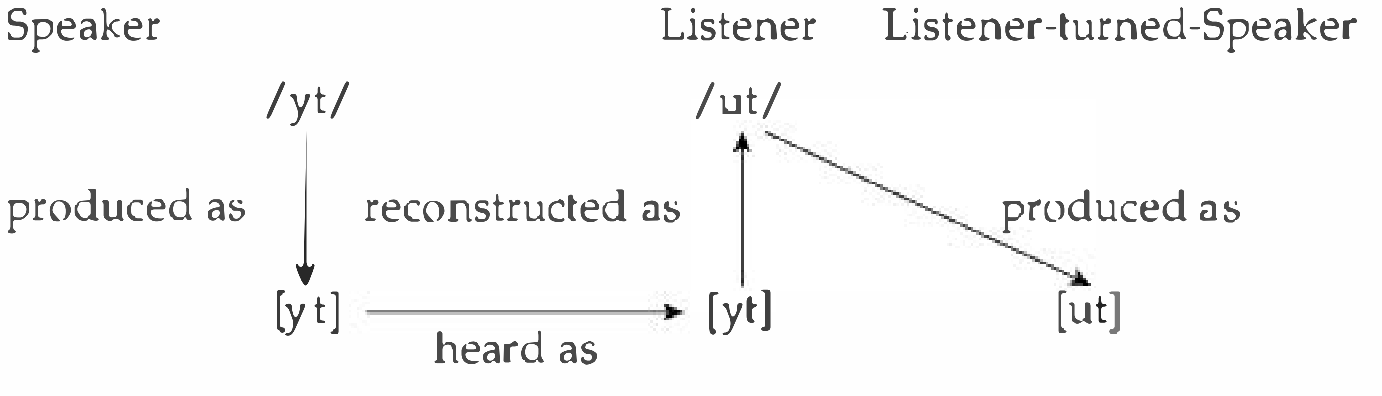
4 Motivations for dissimilation

There are a number of theories of what causes dissimilation. The purpose of this section is to review the major proposals as well as some others of more limited applicability. Our point of departure will be Ohala's Co-articulation-Hypercorrection Theory (CHT; Ohala 1981, 1993, 2003), which is presented in §4.1. According to the CHT, dissimilation results when the listener reverses a perceived co-articulation. The central prediction of the CHT is that dissimilation should only occur with features that have cues that are significantly extended in time. Other theories assume a processing motivation. Frisch *et al.* (2004) argue that similarity avoidance effects are due to the difficulties associated with processing the sequencing of similar segments. This bias is reflected in the statistical structure of the lexicon and is described in §4.2. Following on from this, §4.3 considers the possibility that dissimilation in manner between pairs of adjacent fricatives or stops may be understood in terms of the enhancement of place cues. In §4.4 we look at dissimilation-like phenomena in certain kinds of reduplication ("echo" reduplication) and language games, which exploit non-identity for aesthetic, ludic, or secret purposes.

4.1 Dissimilation as listener reversal of co-articulation

The phonetic realization of certain features may extend over long temporal domains. Long-domain features are interesting because they create an ambiguity for the listener faced with the task of reconstructing the feature's place in phonological structure. This ambiguity creates conditions favorable to reanalysis, which – in the case of temporally extended features – may take one of three forms: assimilation, metathesis, or dissimilation (Blevins 2004). In dissimilation, when one instance of a distinctive feature occurs within the phonetic domain of another instance of the same distinctive feature, there is an ambiguity as to whether the phonetic effects should be ascribed to the first or second instance of the feature, or both. In a series of highly influential publications, Ohala (1981, 1993, 2003) argues that dissimilation as a sound change is the result of reversal by the listener of perceived co-articulation (see CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). The driving force of the change on this view is the overzealous application of reconstructive rules, with the result that long-domain effects that are actually intended by the speaker become reversed. This is known as hypercorrection. The mechanism of Ohala's Co-articulation-Hypercorrection Theory is schematized in (17).

(17) *Dissimilation as sound change by the listener* (after Ohala 1981: 187)



In this example, the speaker intends to say [yt], which is also the form that the listener actually hears. However, the listener is in possession of tacit phonetic knowledge that coronal consonants raise the value of F2 on neighboring vowels. Drawing on this knowledge, he concludes that the intended quality of the vowel has been distorted due to its proximity to the coronal consonant. The perceived distortion is then eliminated by reconstructing the intended form as /ut/: the /y/ dissimilates.

There are three important entailments of the CHT. The first follows from the assumption that dissimilation involves co-articulation. Segments are only expected to dissimilate to the extent that they entail overlapping articulations. Many dissimilations involve segments that are not phonologically adjacent on the level of the segmental root node. A well-known example is Grassmann's Law in Indo-European (Grassmann 1863). In Ancient Greek, which provides one instantiate of the sound law, there cannot be more than one aspirated stop in a pair of adjacent syllables (Smyth 1956 [1920]: 31). Thus, earlier /t^hrik^h-os/ 'hair (GEN SG)' became [trik^{h}os] τρίχος (cf. [t^hriks] θρίξ (NOM SG)). Grassmann's Law apparently represents an interaction between two non-adjacent consonants. Once we take into account the co-articulatory effect of the aspiration on the following vowel, the apparent action-at-a-distance effect evaporates because the aspiration overlaps phonetically with the dissimilation target. Following release of the stop closure, aspiration persists into the following vowel for 60 msecs or so, presenting the listener with an ambiguity as to whether the aspiration represents post-aspiration of the first stop or pre-aspiration of the second. Segments that are outside of each other's co-articulatory range are not expected to dissimilate according to the CHT.

The second consequence of the CHT is that dissimilation cannot take the form of a quantitative change within the same category. Dissimilation should always be limited to phonologically contrastive features (cf. Grammont 1895; Kiparsky 2003). This follows directly from the assumption that what listeners are doing, when they hypercorrect, is reconstructing what they believe is the intended form, which must be a distinctive segment of the language. Assimilative changes, on the other hand, may give rise to novel structures or segments.

The third consequence of the CHT is that it should not matter which direction the perceived distortion is resolved in. The CHT is neutral with respect to whether the dissimilation is progressive or regressive. In (17), an equally valid outcome would have been dissimilation of the consonant, e.g. to /yk/.

The empirical substance of Ohala's proposal consists of the following predictions.

(18) *Ohala's predictions*

a. The likelihood that a given consonantal feature participates in dissimilation depends on whether the associated perceptual cues have a short or long domain.
b. The domain of dissimilation should be linked to the temporal extension of the perceptual cues.
c. Features whose cues are localized on the segment should not show dissimilatory behavior.

On this basis, an up-to-date list of the features shown to have temporal extension, and therefore likely to dissimilate according to the CHT, is shown in Table 60.2, adapted from a corresponding table in a paper on the evolution of metathesis by Blevins and Garrett (2004: 123). Examples are incorporated from the surrounding discussion in their text.

Features not likely to dissimilate according to the CHT are fricative, affricate, stop, and voice. The phonetic cues for each of these segment types are localized on the segment itself. For example, stops are cued by high amplitude bursts on release of the closure. These bursts are very short, of the order of 5 msecs to 10 msecs. The temporal extent of voicing and fricative noise is limited by the extent of the segment's articulation phase. Examples of continuancy and voicing dissimilation nevertheless exist. Examples of continuancy dissimilation are discussed in §4.3, along with a possible phonetic motivation. When Ohala (1981) initially framed his CHT, the existence of liquid dissimilation appeared to present a problem, since at that time no work had been done on temporally extended cues for liquids. Far from being occasional, liquids are, after labials, the most likely to dissimilate. Moreover, they show pronounced action-at-a-distance, as the Georgian example at the beginning of this chapter shows. This is surprising if it is all down to the formant transitions onto neighboring vowels. Starting with Kelly and Local (1986) and Kelly (1989), however, much research has shown that liquids have temporally extended acoustic-perceptual cues. Tunley (1999) demonstrated experimentally that /l/ causes raising in F2 and F3 on neighboring high vowels, while /r/ results in lowering. These effects are moreover observable up to five syllables away from the lateral segment itself (Hawkins and Smith 2001; see also CHAPTER 30: THE REPRESENTATION OF RHOTICS). West (1999b) found that when the liquid and its phonetic context were masked with white noise, speakers were nonetheless able

Table 60.2 Temporally extended features. References to acoustic properties are from Ladefoged (1993; L), Ladefoged *et al.* (1988; LMJ) and Ladefoged and Maddieson (1996; LM)

Feature	*Acoustic property*	*Examples*
rounding	lowering of all formants (LM 356–358)	French, English (Benguerel & Cowan 1974; Lubker & Gay 1982)
velarization	lowered F2 (LM 361–362)	Arabic (Ghazali 1977; Card 1983)
pharyngealization	lowered F3, raised F1 (LM 307)	Interior Salish (Bessell 1998a, 1998b)
palatalization	raised F2 (LM 364)	Catalan (Recasens 1984, 1987), English (Hawkins & Slater 1994), Japanese (Magen 1984), Marshallese (Choi 1992), Russian (Keating 1988), Bantu (Manuel 1987)
retroflection	lowered F3, F4, clustering of F2, F3, F4 (E 203, LM 28)	Gooniyandi (McGregor 1990), Gujarati (Dave 1977), Hindi (Stevens & Blumstein 1975), Malayalam (Dart 1991), Tiwi (Anderson & Maddieson 1994)
laryngealization	more energy in F1, F2, more jitter (LMJ)	Cayuga (Dougherty 1993)
aspiration	more energy in F0, more noise (LMJ)	Cayuga (Dougherty 1993)
nasalization	spectral zero, nasal anti-resonance (LM 116)	English (Cohn 1990)
jaw lowering	raised F1	English (Amerman *et al.* 1970)
rhoticity	lowered F3 (LM 244, 313)	
laterality	lateral formants (LM: 193–197), raised/lowered F2/F3	English (Kelly & Local 1986; Kelly 1989; Tunley 1999; West 1999a, 1999b, 2000; Hawkins & Smith 2001)

to reconstruct the intended liquid from the resonances in vowels up to three syllable nuclei away. These recent findings on the phonetics of liquids thus square well with the prediction that the phonological domain of the dissimilating feature should mirror the temporal extension of the corresponding cues.

Ohala does not consider dissimilation between vowels. Öhman (1966) showed that vowels may co-articulate across intervening consonants. Dissimilation of vowels across syllables is thus consistent with Ohala's broader claims. Interestingly, though, all of the known examples of vowel dissimilation involve vowel height (CHAPTER 21: VOWEL HEIGHT). Vowel height dissimilation is certainly consistent with the experimental finding that lowering of the jaw co-articulates (Amerman *et al.* 1970), but the existence of co-articulation is apparently not a sufficient predictor of dissimilation. Indeed there is a striking complementarity of phonological patterning between vowel height, on the one hand, and vowel color (roundness and backness), on the other. To date, no examples of dissimilation involving the labial or front–back dimensions have come to light. Conversely, systems of vowel harmony in which backness or rounding (or both) are active are richly attested in the literature, but the feature [low] is not frequent in vowel harmony (see Krause 1979 for a possible example from Chukchee). Further discussion of this problem may be found in Alderete and Frisch (2007).

There is also still a residue of dissimilatory patterns for which the CHT does not seem to offer an explanation, including NC, long vowel, continuancy (see §4.3 below), and voicing. In the next section we will consider an alternative theory of the origins of dissimilation that does not seem to make any prediction about which features participate.

4.2 Similarity avoidance in the lexicon

Several recent studies have examined statistical asymmetries in the lexicon, pointing to a preference for phonetic dissimilarity between neighboring consonants in roots (see also CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS). Berkley (2000) studied English monosyllabic words and found evidence of gradient similarity avoidance effects. Focusing on words of the shape C_1VC_2, she found that there are significantly fewer such words containing homorganic consonants than would be expected if consonants combined randomly, i.e. had the same probability of occurring as two independent events. Under- or overrepresentation is the ratio of observed to expected frequency (see CHAPTER 90: FREQUENCY EFFECTS). A pair is underrepresented if the observed-to-expected (or O/E) ratio is less than 1, overrepresented if greater than 1. Words in which C_1 and C_2 are homorganic – such as *mop*, *lull*, and *king* – are underrepresented in the English lexicon. For CVVC words with a long vowel intervening between C_1 and C_2, the homorganic similarity avoidance effect is also present but weaker. These results, adapted from Berkley (2000), are shown in Table 60.3. Shaded cells indicate underrepresented combinations.

Frisch *et al.* (2004) studied similarity avoidance in the lexicon of Arabic triradical verb roots. They found a very strong effect for adjacent pairs of consonants

Table 60.3 Similarity avoidance in English monosyllabic roots

			Labial p b f v m w	Cor obs t d θ ð s z ʃ ʒ ʧ ʤ	Cor son n l r j	Dorso-guttural k g h ŋ w
CVC	Labial	p b f v m w	0.52	1.15	1.33	0.85
	Cor obs	t d θ ð s z ʃ ʒ ʧ ʤ		0.73	1.03	1.09
	Cor son	n l r j			0.59	1.21
	Dorso-guttural	k g h ŋ w				0.72
CVVC	Labial	p b f v m w	0.61	1.11	1.11	1.14
	Cor obs	t d θ ð s z ʃ ʒ ʧ ʤ		0.75	1.17	1.03
	Cor son	n l r j			0.70	1.01
	Dorso-guttural	k g h ŋ w				0.71

(C_1 and C_2, C_2 and C_3), tending to categorical, as shown in Table 60.4. For non-adjacent C_1 and C_3 the effect was still strong, although somewhat weaker. Under both adjacency and non-adjacency the similarity avoidance effect is far stronger than the one observed by Berkley for English. Gray cells are homorganic in terms of major class (Labial, Coronal obstruent, Dorsal, Guttural, Coronal sonorant). They also found that the avoidance was stronger the more similar the consonant. Within the major class of coronals, for example, an adjacent pair of coronals was significantly more frequent if they had different values for [continuant].

Frisch *et al.* argue that when observed co-occurrence deviates from expected co-occurrence, the learner posits a gradient phonological constraint, which they dub the gradient Obligatory Contour Principle that encodes the generalization "roots with repeated homorganic consonants are unusual." Statistical generalizations like these form the basis of metalinguistic judgments of relative acceptability of novel words ("word-likeness"), influencing which words are actually used, and the phonological forms of novel and borrowed words (cf. Frisch 2004: 346).

For Frisch *et al.*, the gradient OCP represents a statistical generalization over a static lexicon; it does not encode tacit phonetic knowledge directly. Despite this, Frisch *et al.* do propose a functional explanation for the distributional asymmetries in the lexicon. Repetition of similar consonants is difficult to process (Frisch 2004). This finds the beginnings of an explanation in neural network models that encode linearization of segments. Nodes in the network must be excited and inhibited so as to fire in the right sequence. If there is a sequence of similar segments, the periods of excitation and inhibition may overlap, whether or not there is a corresponding overlap in the acoustic signal. Given two segments, C_1 and C_2, in linear sequence that activate the same distinctive feature node, if the node encoding C_1 is still firing when C_2 is perceived, this may result in simultaneous

Table 60.4 Similarity avoidance in Arabic verb roots

			Labial	Cor obs		Dorsal	Guttural		Cor son
			b f m	t d tˤ dˤ	θ ð s z sˤ zˤ ʃ	k g q	χ ʁ	ħ ʕ h ʔ	l r n
Adjacent	Labial	b f m	0.00	1.37	1.31	1.15	1.35	1.17	1.18
	Cor obs	t d tˤ dˤ		0.14	0.52	0.80	1.43	1.25	1.23
		θ ð s z sˤ zˤ ʃ			0.04	1.16	1.41	1.26	1.21
	Dorsal	k g q				0.02	0.07	1.04	1.48
	Guttural	χ ʁ					0.00	0.07	1.39
		ħ ʕ h ʔ						0.06	1.26
	Cor son	l r n							0.06
Non-adjacent	Labial	b f m	0.30	1.08	1.02	1.26	1.25	1.28	1.11
	Cor obs	t d tˤ dˤ		0.38	1.06	1.24	1.05	1.02	0.97
		θ ð s z sˤ zˤ ʃ			0.24	1.16	1.35	1.14	1.23
	Dorsal	k g q				0.07	0.68	1.19	1.03
	Guttural	χ ʁ					0.25	0.12	1.10
		ħ ʕ h ʔ						0.34	1.13
	Cor son	l r n							0.67

perception of C_1 and C_2. The resulting blend of the two percepts may result in the same kind of ambiguity that results from co-articulation in the CHT, and is presumably consistent with the same re-analytic strategies. An alternative source of dissimilation in processing may be a refractory period during which the node must be reset in order to detect a second stimulus of the same type. Unlike the CHT and the blending scenario we have just sketched, the refractory period would seem to predict asymmetries in the direction of the resulting pattern. If C_2 occurs within this refractory period of a node that has just fired for C_1, the relevant feature will be perceived on C_1 but may not be perceived on C_2. The effective result is progressive dissimilation.

This processing bias may further help explain the pronounced difference in the strength of the effect in English and Arabic. Arabic has non-concatenative morphology and psychologically real abstract consonant roots like /ktb/ 'write'. In phonological analysis, this has motivated analyses in which consonants and vowels occupy separate tiers in the representation. Roots are vulnerable to speech errors involving misordering of radical consonants.

These functional mechanisms and their implications for linguistic patterning provide much fertile ground for further research. First, a greater range of languages must be studied to determine to what extent root co-occurrence constraints have a gradient character or not. Suzuki (1998) presents 16 examples of root co-occurrence restriction, which are summarized in Table 60.5, along with a couple of additional cases.

A second challenge concerns the difference between similarity and identity. Arabic shows avoidance of identical radicals. In English, however, segmental identity provides an escape hatch to the OCP. CVC roots where both C_1 and C_2 are labial are generally dispreferred, but not if C_1 and C_2 are identical. For example, selecting /p/ for C_1 and C_2 and permuting the possible nuclei, almost every cell of the paradigm corresponds to an actually occurring word of English: *pip, pep, pap, pop, pup, peep, poop, parp, pipe, Pape, pope*. See Idsardi and Raimy (2008) for a relevant proposal that segmental identity may be represented in a data structure they call a "linked list."

Finally, the similarity avoidance approach raises anew the question of which features are expected to participate in dissimilation. Are certain features associated with longer periods of excitation in perception than others? And if that turns out to be so, is there a systematic correlation between the length of a feature's temporal domain in the speech signal and the duration of excitation? To date, place and laryngeal features (MacEachern 1997) have been studied. These studies must therefore be extended to short-domain features, such as stops, fricatives, and voicing.

4.3 *Dissimilation and cue robustness*

Manner dissimilation is predicted not to occur by the CHT. Despite this, a small number of languages display dissimilation of pairs of adjacent stops or fricatives. For example, Osage (Quintero 2004), a Siouan language spoken in Oklahoma, has a rule dissimilating /ð/ to [t] following /s/, e.g. /ʃkɔ̃ʃða/ → [ʃkɔ̃ʃta] 'you want'. Tsou (Szakos 1994), an Austronesian language of Taiwan, has a rule that hardens /h/ to [k] following /s/, giving alternations such as [s-in-uhnu] 'send someone to do something (ACTOR VOICE)' ~ [skuna] (PATIENT VOICE), [s-in-ohpici] 'pinch

Table 60.5 Root co-occurrence restrictions

Feature	*Case*
place	Amharic (Bender & Fulass 1978), Arabic (Greenberg 1950; McCarthy 1979, 1981; Mester 1986; Yip 1989; Padgett 1991; Pierrehumbert 1993; Frisch *et al.* 2004), Cambodian (Yip 1989), French (Plénat 1996), Hawai'ian (McKay 1970), Hebrew (Koskinen 1964), Javanese (Uhlenbeck 1949; Mester 1986; Padgett 1991), Russian (Padgett 1991), Serbian (McKay 1970), Yucatec Mayan (Yip 1989; Lombardi 1991)
labial	Cantonese (Yip 1982, 1988), Ponapean (Rehg & Sohl 1981; Goodman 1995), Yao (Purnell 1965; Ohala 1981), Zulu (Doke 1926)
coronal	Akan (Welmers 1946; McCarthy & Prince 1995)
pharyngeal	Moses-Columbia Salish (Czaykowska-Higgins 1993)
liquid	Javanese (Uhlenbeck 1949; Mester 1986)
rhotic	American English (Hall 2009)
voice	Japanese (Itô & Mester 1986, 2003; Steriade 1987, 1995; Ishihara 1991; Archangeli & Pulleyblank 1994; Alderete 1997; Pater 1999), Bakairi (Gussenhoven & Jacobs 2005)
spread glottis	Sanskrit (Grassmann 1863; Langendoen 1966; Anderson 1970; Phelps & Brame 1973; Sag 1974, 1976; Phelps 1975; Schindler 1976; Borowsky & Mester 1983; Kaye & Lowenstamm 1985; Lombardi 1991)
high	Ngbaka (Thomas 1963; Chomsky & Halle 1968; Mester 1986)
back	Ainu (Itô 1984; Mester 1986; Archangeli & Pulleyblank 1994), Tzeltzal (Slocum 1948; Itô 1984)
length	Japanese

(ACTOR VOICE)' ~ [skopica] (PATIENT VOICE). Non-sibilant fricatives such as [θ x] have diffuse spectra. Harris (1958) shows that the F2 transition is required for reliable identification of the fricative. In a cluster of fricatives, one of the transitions, C–V or V–C, is missing. Dissimilating one of the fricatives to the corresponding stop has the effect of sharpening the F2 transition and adding a stop release burst, rendering the place of articulation more easily identifiable.[3] In Chontal (Waterhouse 1949, 1962; Kenstowicz and Kisseberth 1979), a Hokan language of Mexico, the imperative suffix is {-laʔ} after voiceless segments, and {-ɬaʔ} after voiced ones, e.g. [fuʃ-lʲaʔ] 'blow it!', [panx-laʔ] 'sit down!', [ko-ɬaʔ] 'say it!', [kan-ɬaʔ] 'leave it!'. The pattern seems to involve deleting the second [spread glottis] feature in a

[3] One of the apparent cases of continuant assimilation mentioned by Ohala as a potential counterexample to the CHT may in fact turn out to be best understood as an instance of it. Dyen (1972) shows that Proto-Austronesian */s . . . s . . ./ was dissimilated across an intervening vowel in Ngaju-Dayak to /t . . . s . . ./. The evidence, however, only consists of the two words PA *sisik > ND [tisik] 'fish-scale' and PA *susu > ND [tuso] 'breast'. It is perhaps relevant that the vowel immediately following the initial *s is a high vowel. High vowels are known to increase the degree of post-aspiration of a preceding voiceless stop and affrication of a preceding /t/. The initial sibilant may thus have been interpreted as the co-articulatory affrication of an intended /t/.

cluster of voiceless fricatives – assuming the claim of Vaux (1998) that voiceless fricatives are universally [spread glottis] – allowing the lateral to be more clearly identified as such.

Dissimilation between two stops is far more rare, but González (2008) supplies an example from North-Central Spanish, which shows dissimilation of coda /k/ to [θ] (and other realizations) before another stop (generally /t/, e.g. [doθ.'tor] 'doctor'. González also proposes an explanation in terms of cue robustness, noting that the cues for the first stop are not as salient before another stop as before other segments, due to weaker, or absent, stop release burst and formant transitions. Similar considerations have been argued to condition metathesis in other languages, e.g. Faroese, where final /skt/ metathesizes to [kst] (Hume and Seo 2004; see CHAPTER 59: METATHESIS).

4.4 The dissimilation game

In a different vein, people also apply tacit knowledge of similarity to a variety of ludic and poetic ends. Indeed, the term "dissimilation" entered the field in the 19th century from rhetoric, where it had been in use to describe the variation in style required for good public speaking (cf. Brugmann 1909). The criterion of a perfect rhyme in English, such as *pet – bet*, is not only that the material following the onset of each stressed syllable is identical, but that the onset of each stressed syllable is different. In considering rhyme in English, we do not appear to count features; we are merely interested in contrastive segments. The pair *pet – bet* is thus as good a rhyme as the pair *pet – set*. The same requirement of non-identity turns up in echo reduplication (Alderete *et al.* 1999; Nevins 2005; CHAPTER 100: REDUPLICATION), where the base is reduplicated with an onset determined by convention (fixed segmentism). In Hindi, this kind of reduplication gives a meaning 'X and the like' (Singh 1969; Nevins 2005). The fixed segment is /v/ unless the base also begins with a /v/, in which case the echo reduplicant begins with /ʃ/. Examples from Nevins (2005: 280) are shown in (19).[4]

(19) *Hindi echo reduplication with fixed segmentism*

paanii-vaani	'water and the like'
aam-vaam	'mangoes and the like'
tras-vras	'grief and the like'
yaar-vaar	'friends and the like'
vakil-ʃakil, *vakil-vakil	'lawyers and the like'

Similar facts are observed in English *shm-* reduplication, e.g. *potato–shmotato*, but *shmaltz-shpaltz* (Nevins and Vaux 2003), Kannada (Lidz 2001), and Javanese (Yip 1995). Yip (1995, 1998) proposes that these are due to a constraint against the repetition of identical elements, *REPEAT, ultimately due to Menn and MacWhinney (1984). Similar facts also turn up in secret languages. In the Kunshan secret language Mo-pa (Yip 1982: 652ff.), a base of the shape $C_1V_1(C_2)$ is mapped to a template C_1[o]-GV_1C_2, where G is a consonant whose value for the feature

[4] The forms in (19) represent Nevins's own fieldwork. The original source on Hindi echo words is Singh (1969).

[continuant] is the opposite of that of C_1. Examples, not glossed in the source, are given in (20).

(20) *Kunshan secret language Mo-pa*

təw	tʊ ləw
k'ɛ	k'o ɦe
d'oŋ	d'o loŋ
tsa	tso za
vã	vo pã
sja	so tsja
nɔw	no tɔw
ɲən	ɲa tɕan

Oral stops are replaced by voiced continuants, while nasals and continuants are replaced by the corresponding voiceless unaspirated stop. This continuant dissimilation is covered neither by the CHT nor cue robustness; the reason for the alternation seems to be purely ludic.

5 Conclusions

There are a number of theories of the origin of dissimilation, and dissimilation may apparently have one of several motivations. According to the Co-articulation-Hypercorrection Theory (Ohala 1981, 1993, 2003), dissimilation results when the listener reverses a perceived co-articulation. The central prediction of the CHT is that dissimilation should only occur with features that have elongated cues. Other theories assume a functional motivation. Frisch *et al.* (2004) argue that similarity avoidance effects are due to the difficulties associated with processing the sequencing of similar segments. This bias is reflected in the statistical structure of the lexicon in many languages. However, the predictions of processing-based accounts with respect to the observed featural asymmetries are not yet clear. It was suggested here that manner dissimilation in pairs of adjacent fricatives or stops is best understood as maximizing cues for place of articulation, while dissimilatory phenomena in language games fulfill an aesthetic role. Future research will hopefully extend the empirical base for the study of dissimilation phenomena, and determine more precisely what the division of labor and synergies between the factors discussed here should be.

ACKNOWLEDGMENTS

I would like to thank Beth Hume, Marc van Oostendorp, and two anonymous reviewers for helpful feedback on this chapter.

REFERENCES

Alderete, John. 1995. Faithfulness to prosodic heads. Unpublished ms., University of Massachusetts, Amherst (ROA-94).

Alderete, John. 1997. Dissimilation as local conjunction. *Papers from the Annual Meeting of the North East Linguistic Society* 27. 17–32.

Alderete, John & Stefan A. Frisch. 2007. Dissimilation in grammar and the lexicon. In de Lacy (2007), 379–398.

Alderete, John, Jill Beckman, Laura Benua, Amalia Gnanadesikan, John J. McCarthy & Suzanne Urbanczyk. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30. 327–364.

Amerman, James, Raymond Daniloff & Kenneth Moll. 1970. Lip and jaw coarticulation for the phoneme. *Journal of Speech and Hearing* 13. 147–161.

Anderson, Stephen R. 1970. On Grassmann's Law in Sanskrit. *Linguistic Inquiry* 1. 387–396.

Anderson, Victoria B. & Ian Maddieson. 1994. Acoustic characteristics of Tiwi coronal stops. *UCLA Working Papers in Phonetics* 87. 131–162.

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Aronoff, Mark & Richard T. Oehrle (eds.) 1984. *Language sound structure*. Cambridge, MA: MIT Press.

Beckman, Jill N., Laura Walsh Dickey & Suzanne Urbanczyk (eds.) 1995. *Papers in Optimality Theory*. Amherst: GLSA.

Bender, Byron W. 1968. Marshallese phonology. *Oceanic Linguistics* 7. 16–35.

Bender, Byron W. 1969. Vowel dissimilation in Marshallese. *University of Hawai'i Working Papers in Linguistics* 1. 88–95.

Bender, M. Lionel & Hailu Fulass. 1978. *Amharic verb morphology: A generative approach*. East Lansing: Michigan State University.

Benguerel, A.-P. & H. Cowan. 1974. Coarticulation of upper lip protrusion in French. *Phonetica* 30. 41–55.

Bennett, Patrick R. 1967. Dahl's Law and Thagicũ. *African Language Studies* 8. 127–159.

Berkley, Deborah M. 2000. Gradient Obligatory Contour Principle effects. Ph.D. dissertation, Northwestern University.

Bessell, Nicola J. 1998a. Local and non-local consonant-vowel interaction in Interior Salish. *Phonology* 15. 1–40.

Bessell, Nicola J. 1998b. Phonetic aspects of retraction in Interior Salish. In Ewa Czaykowska-Higgins & M. Dale Kinkade (eds.) *Salish languages and linguistics: Theoretical and descriptive perspectives*, 125–152. Berlin & New York: Mouton de Gruyter.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette & Andrew Garrett. 2004. The evolution of metathesis. In Hayes *et al.* (2004), 117–156.

Borowsky, Toni & Armin Mester. 1983. Aspiration to roots: Remarks on the Sanskrit diaspirates. In *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 19. 52–63.

Boukous, A. 1987. Phonotactiques et domaines prosodiques en berbère (parler tachelhit d'Agadi, Maroc). Thèse d'état, Université de Paris VIII.

Brugmann, Karl. 1909. Das Wesen der lautlichen Dissimilation. *Abhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften* 57. 139–178.

Bye, Patrik. Forthcoming. Derivations. In Nancy C. Kula, Bert Botma & Kuniya Nasukawa (eds.) *The Continuum handbook of phonology*. London: Continuum.

Bye, Patrik, Elin Sagulin & Ida Toivonen. 2009. Phonetic duration, phonological quantity and prosodic structure in Inari Saami. *Phonetica* 66. 199–221.

Card, Elizabeth. 1983. A phonetic and phonological study of Arabic emphasis. Ph.D. dissertation, Cornell University.

Choi, John D. 1992. Phonetic underspecification and target interpolation: An acoustic study of Marshallese vowel allophony. Ph.D. dissertation, University of California, Los Angeles.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1991. Place of articulation in consonants and vowels: A unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5. 77–123.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

Cohn, Abigail C. 1990. Phonetic and phonological rules of nasalization. Ph.D. dissertation, University of California, Los Angeles.

Cohn, Abigail C. 1992. The consequences of dissimilation in Sundanese. *Phonology* 9. 199–220.

Crosswhite, Katherine. 1999. Vowel reduction in Optimality Theory. Ph.D. dissertation, University of California, Los Angeles.

Czaykowska-Higgins, Ewa. 1993. Cyclicity and stress in Moses-Columbia Salish (Nxa'amxcin). *Natural Language and Linguistic Theory* 11. 197–278.

Daly, John P. 1993. The role of tone sandhi in tone analysis. *Notes on Linguistics* 61. 5–24.

Dart, Sarah N. 1991. Articulatory and acoustic properties of apical and laminal articulation. Ph.D. dissertation, University of California, Los Angeles.

Dave, Radhekant. 1977. Retroflex and dental consonants in Gujarati: A palatographic and acoustic study. *Annual Report of the Institute of Phonetics, University of Copenhagen* 11. 27–156.

Davy, J. L M. & Derek Nurse. 1982. The synchronic forms of Dahl's Law, a dissimilation process, in Central and Lacustrine Bantu languages of Kenya. *Journal of African Languages and Linguistics* 4. 159–195.

de Lacy, Paul (ed.) 2007. *The Cambridge handbook of phonology.* Cambridge: Cambridge University Press.

Dixon, R. M. W. 1977. *A grammar of Yidiɲ.* Cambridge: Cambridge University Press.

Doke, Clement M. 1926. *The phonetics of the Zulu language.* Johannesburg: University of the Witwatersrand Press.

Dougherty, Brian. 1993. The acoustic-phonetic correlates of Cayuga word stress. Ph.D. dissertation, Harvard University.

Dyen, Isidore. 1972. Non-gradual regular phonetic changes involving sibilants. In Jacques Barrau, Lucien Bernot, George Condominas, Mariel Jean Brunhes Delamare, Francis Leroy, Alexic Rygaloff & Jacqueline M. C. Thomas (eds.) *Langues et techniques natures et société,* vol. 1: *Approche linguistique,* 95–99. Paris: Klincksieck.

Ebert, Karen. 1974. Partial vowel harmony in Kera. *Studies in African Linguistics,* Supplement 5. 75–80.

Elmedlaoui, Mohammed. 1985. Le parler berbère chleuh d'Imdlawn: Segments et syllabation. Doctorat de troisième cycle, Université de Paris VIII, Saint-Denis.

Elmedlaoui, Mohammed. 1995. *Aspect des representations phonologiques dans certaines langues chamitosemitiques.* Rabat: Faculté des lettres et des sciences humaines.

Evans, Nicholas D. 1995. Current issues in the phonology of Australian languages. In Goldsmith (1995), 723–761.

Fallon, Paul D. 1993. Liquid dissimilation in Georgian. *Proceedings of the 10th Eastern States Conference on Linguistics (ESCOL),* 105–116.

Finer, Daniel. 1986. Reduplication and verbal morphology in Palauan. *The Linguistic Review* 6. 99–130.

Foley, William A. 1991. *The Yimas language of New Guinea.* Stanford: Stanford University Press.

Frisch, Stefan A. 2004. Language processing and segmental OCP effects. In Hayes *et al.* (2004), 346–371.

Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22. 179–228.

Geytenbeek, Brian B. & Helen Geytenbeek. 1971. *Gidabal grammar and dictionary.* Canberra: Australian Institute of Aboriginal Studies.

Ghazali, Salem. 1977. Back consonants and backing coarticulation in Arabic. Ph.D. dissertation, University of Texas, Austin.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.
Goldsmith, John A. 1984. Meeussen's Rule. In Aronoff & Oehrle (1984), 245–259.
Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.
González, Carolina. 2008. Assimilation and dissimilation of syllable-final /k/ in North-Central Spanish. In Joyce Bruhn de Garavito & Elena Valenzuela (eds.) *Selected Proceedings of the 10th Hispanic Linguistics Symposium*, 170–183. Somerville, MA: Cascadilla Press.
Goodman, Beverley. 1995. Features in Ponapean phonology. Ph.D. dissertation, Cornell University.
Gragg, Gene. 1976. Oromo of Wellegga. In M. Lionel Bender (ed.) *The non-Semitic languages of Ethiopia*, 166–195. East Lansing: Michigan State University Press.
Grammont, Maurice. 1895. *La dissimilation consonantique dans les langues indo-européennes et dans les langues romanes*. Dijon: Darantière.
Grassmann, Hermann. 1863. Ueber die Aspiraten und ihr gleichzeitiges Vorhandensein im An- und Auslaute der Wurzeln. *Zeitschrift für Vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 12. 81–138.
Greenberg, Joseph. 1950. The patterning of root morphemes in Semitic. *Word* 6. 162–181.
Gussenhoven, Carlos & Haike Jacobs. 2005. *Understanding phonology*. London: Hodder Arnold.
Hall, Nancy. 2009. R-dissimilation in American English. Unpublished ms., California State University, Long Beach.
Harris, K. S. 1958. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1. 1–7.
Hawkins, Sarah & Andrew Slater. 1994. Spread of CV and V-to-V coarticulation in British English: Implications for the intelligibility of synthetic speech. In *Proceedings of the 3rd International Conference on Spoken Language Processing*, 57–60. Yokohama: Acoustical Society of Japan.
Hawkins, Sarah & R. Smith. 2001. An acoustical study of long-domain /r/ and /l/ coarticulation. In Sebastian Heid & Sarah Hawkins (eds.) *Proceedings of the 5th Seminar on Speech Production: Models and Data*, 77–80. Bavaria: Kloster Seeon.
Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.
Holton, David. 1995. Assimilation and dissimilation of Sundanese liquids. In Beckman *et al.* (1995), 167–180.
Hume, Elizabeth. 1992. Front vowels, coronal consonants and their interaction in non-linear phonology. Ph.D. dissertation, Cornell University.
Hume, Elizabeth & David Odden. 1996. Reconsidering [consonantal]. *Phonology* 13. 345–376.
Hume, Elizabeth & Misun Seo. 2004. Metathesis in Faroese and Lithuanian: From speech perception to Optimality Theory. *Nordic Journal of Linguistics* 27. 35–60.
Idsardi, William J. & Eric Raimy. 2008. Reduplicative economy. In Bert Vaux & Andrew Nevins (eds.) *Rules, constraints, and phonological phenomena*, 149–184. Oxford: Oxford University Press.
Ishihara, Masahide. 1991. A lexical prosodic phonology of Japanese verbs. Ph.D. dissertation, University of Arizona.
Itkonen, Erkki. 1986–91. *Inarilappisches Wörterbuch*. 4 vols. Helsinki: Suomalais-ugrilainen Seura.
Itô, Junko. 1984. Melodic dissimilation in Ainu. *Linguistic Inquiry* 15. 505–513.
Itô, Junko & Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17. 49–73.
Itô, Junko & Armin Mester. 2003. *Japanese morphophonemics: Markedness and word structure*. Cambridge, MA: MIT Press.
Jebbour, Abdelkrim. 1985. La labio-vélarization en berbère dialecte tachelhit (parler de tiznit). In *Mémoire de phonologie*. Rabat: Université Mohammed V, Faculté des letters et des sciences humaine.

Johnson, Lawrence. 1973. Dissimilation as a natural process in phonology. *Stanford Occasional Papers in Linguistics* 3. 45–56.
Joseph, Brian D. & Richard Janda (eds.) 2003. *The handbook of historical linguistics.* Malden, MA & Oxford: Blackwell.
Josephs, Lewis S. 1975. *Palauan reference grammar.* Honolulu: University of Hawai'i Press.
Josephs, Lewis S. 1990. *New Palauan–English dictionary.* Honolulu: University of Hawai'i Press.
Kaisse, Ellen M. 1988. Modern Greek continuant dissimilation and the OCP. Unpublished ms., University of Washington.
Kaye, Jonathan & Jean Lowenstamm. 1985. A metrical treatment of Grassmann's Law. *Papers from the Annual Meeting of the North East Linguistic Society* 15. 220–233.
Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1985. The internal structure of phonological elements: A theory of charm and government. *Phonology Yearbook* 2. 305–328.
Keating, Patricia. 1988. Underspecification in phonetics. *Phonology* 5. 275–292.
Kelly, John. 1989. On the phonological relevance of some non-phonological elements. In Tamás Szende (ed.) *Proceedings of the Speech Research '89 International Conference,* 56–59. Budapest: Linguistic Institute of the Hungarian Academy of Sciences.
Kelly, John & John Local. 1986. Long-domain resonance patterns in English. In *Proceedings of the International Conference on Speech Input/Output: Techniques and Applications,* 304–309. London: Institution of Electrical Engineers.
Kenstowicz, Michael & Charles W. Kisseberth. 1977. *Topics in phonological theory.* New York: Academic Press.
Kenstowicz, Michael & Charles W. Kisseberth. 1979. *Generative phonology: Description and theory.* New York: Academic Press.
Kent, Ronald G. 1936. Assimilation and dissimilation. *Language* 12. 245–258.
Kent, Ronald G. 1945. *The sounds of Latin: A descriptive and historical phonology.* Baltimore: Waverley Press.
Keyser, Samuel J. & Paul Kiparsky. 1984. Syllable structure in Finnish phonology. In Aronoff & Oehrle (1984), 7–31.
Kiparsky, Paul. 2003. The phonological basis of sound change. In Joseph & Janda (2003), 313–342.
Kisseberth, Charles W. 1970. On the functional utility of phonological rules. *Linguistic Inquiry* 1. 291–306.
Koskinen, K. 1964. Kompatibilität in den dreikonsonantigen hebräischen Wurzeln. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 144. 16–58.
Krause, Scott. 1979. Topics in Chukchee phonology and morphology. Ph.D. dissertation, University of Illinois, Urbana-Champaign.
Ladefoged, Peter. 1993. *A course in phonetics.* 3rd edn. New York: Harcourt Brace Jovanovich.
Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages.* Oxford & Malden, MA: Blackwell.
Ladefoged, Peter, Ian Maddieson & Michel T. T. Jackson. 1988. Investigating phonation types in different languages. In Osamu Fujimura (ed.) *Vocal physiology: Voice production, mechanisms and functions,* 292–317. New York: Raven.
Langendoen, D. Terence. 1966. A restriction on Grassmann's Law in Greek. *Language* 42. 7–9.
Larsi, Ahmed. 1991. Aspects de la phonologie non-lindaire du parler berbère chleuh de Tidli. Thèse de doctorat, Université de la Sorbonne Nouvelle, Paris III.
Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT.
Leumann, Manu. 1977. *Lateinische Grammatik von Leumann-Hofmann-Szantyr,* vol. 1: *Lateinische Laut- und Formenlehre.* Munich: Beck.
Levergood, Barbara. 1987. Topics in Arusa phonology and morphology. Ph.D. dissertation, University of Texas, Austin.

Levi, Susannah V. 2008. Phonemic *vs.* derived glides. *Lingua* 118. 1956–1978.
Lidz, Jeffrey. 2001. Echo reduplication in Kannada and the theory of word-formation. *The Linguistic Review* 18. 375–394.
Lloret, Maria-Rosa. 1988. Gemination and vowel length in Oromo morphophonology. Ph.D. dissertation, Indiana University.
Lombardi, Linda. 1991. Laryngeal features and laryngeal neutralization. Ph.D. dissertation, University of Massachusetts, Amherst.
Lombardi, Linda. 1995. Dahl's law and privative [voice]. *Linguistic Inquiry* 26. 365–372.
Lubker, J. & T. Gay. 1982. Anticipatory labial coarticulation: Experimental, biological and linguistic variables. *Journal of the Acoustical Society of America* 71. 437–448.
Lynch, John. 2003. Low vowel dissimilation in Vanuatu languages. *Oceanic Linguistics* 42. 359–406.
MacEachern, Margaret R. 1997. Laryngeal co-occurrence restrictions. Ph.D. dissertation, University of California, Los Angeles. Published 1999, New York: Garland.
Maddieson, Ian. 1984. *Patterns of sounds.* Cambridge: Cambridge University Press.
Magen, Harriet S. 1984. Vowel-to-vowel coarticulation in English and Japanese. *Journal of the Acoustical Society of America* 75. Supplement S41.
Manuel, Sharon Y. 1987. Acoustic and perceptual consequences of vowel-to-vowel co-articulation in three Bantu languages. Ph.D. dissertation, Yale University.
Marlett, Stephen A. & Joseph Paul Stemberger. 1983. Empty consonants in Seri. *Linguistic Inquiry* 14. 617–639.
McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.
McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12. 373–418.
McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17. 207–263.
McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.
McCarthy, John J. 2007. Derivations and levels of representation. In de Lacy (2007), 99–117.
McCarthy, John J. & Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University.
McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Beckman *et al.* (1995), 249–384.
McConvell, Patrick. 1988. Nasal cluster dissimilation and constraints on phonological variables in Gurindji and related languages. In Nicholas D. Evans & Steve Johnson (eds.) *Aboriginal Linguistics* 1, 135–165. Armidale: Department of Linguistics, University of New England.
McGregor, William. 1990. *A functional grammar of Gooniyandi.* Amsterdam: John Benjamins.
McKay, D. G. 1970. Phoneme repetition in the structure of languages. *Language and Speech* 13. 199–213.
Menn, Lise & Brian MacWhinney. 1984. The repeated morph constraint: Toward an explanation. *Language* 60. 519–541.
Mester, Armin. 1986. Studies in tier structure. Ph.D. dissertation, University of Massachusetts, Amherst.
Morén, Bruce. 2003. The parallel structures model of feature geometry. *Working Papers of the Cornell Phonetics Laboratory* 15. 194–270.
Nevins, Andrew. 2005. Overwriting does not optimize in nonconcatenative morphology. *Linguistic Inquiry* 36. 275–287.
Nevins, Andrew & Bert Vaux. 2003. Metalinguistic, shmetalinguistic: The phonology and morphology of *shm-* reduplication. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 39. 702–722.

Newton, Brian. 1971. Sibilant loss in Northern Greek. *Canadian Journal of Linguistics* 17. 1–15.

Odden, David. 1987. Dissimilation as deletion in Chuckchi. *Proceedings of the 3rd Eastern States Conference on Linguistics (ESCOL)*. 235–246.

Odden, David. 1988. Anti antigemination and the OCP. *Linguistic Inquiry* 19. 451–475.

Odden, David. 1994. Adjacency parameters in phonology. *Language* 70. 289–330.

Ohala, John J. 1981. The listener as a source of sound change. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 17. 178–203.

Ohala, John J. 1993. The phonetics of sound change. In Charles Jones (ed.) *Historical linguistics: Problems and perspectives*, 237–278. London: Longman.

Ohala, John J. 2003. Phonetics and historical phonology. In Joseph & Janda (2003), 669–686.

Öhman, Sven E. G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39. 151–168.

Padgett, Jaye. 1991. Stricture in feature geometry. Ph.D. dissertation, University of Massachusetts, Amherst.

Paradis, Carole & Jean-François Prunet. 1989. On coronal transparency. *Phonology* 6. 317–348.

Parker, Steve 1997. An OT account of laryngealization in Cuzco Quechua. *Work Papers of the Summer Institute of Linguistics, University of North Dakota* 41. 47–57.

Pater, Joe. 1999. Austronesian nasal substitution and other NÇ effects. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*, 310–343. Cambridge: Cambridge University Press.

Pearce, Mary. 2008. Vowel harmony domains and vowel undershoot. *UCL Working Papers in Linguistics* 20. 115–140.

Phelps, Elaine. 1975. Sanskrit diaspirates. *Linguistic Inquiry* 6. 447–464.

Phelps, Elaine & Michael Brame. 1973. On local ordering of rules in Sanskrit. *Linguistic Inquiry* 4. 387–400.

Pierrehumbert, Janet B. 1993. Dissimilarity in the Arabic verbal roots. *Papers from the Annual Meeting of the North East Linguistic Society* 23. 367–381.

Pitkin, Harvey. 1984. *Wintu grammar*. Berkeley: University of California Press.

Plénat, Marc. 1996. De l'interaction des contraints: Une étude de case. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, 585–617. Salford: ESRI.

Poser, William J. 1982. Phonological representations and action-at-a-distance. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part II, 121–158. Dordrecht: Foris.

Posner, Rebecca R. 1961. Consonantal dissimilation in the Romance languages. *Publications of the Philological Society* 18. Oxford: Blackwell.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Purnell, Herbert C. 1965. *Phonology of a Yai dialect spoken in the province of Chiengrai, Thailand*. Hartford, CT: Hartford Seminary Foundation.

Quintero, Carolyn. 2004. *Osage grammar*. Lincoln, NE: University of Nebraska Press.

Recasens, Daniel. 1984. Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America* 76. 1624–1635.

Recasens, Daniel. 1987. An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics* 15. 299–312.

Rehg, Kenneth L. & Damian G. Sohl. 1981. *Ponapean reference grammar*. Honolulu: University of Hawai'i Press.

Rubach, Jerzy. 1993. *The Lexical Phonology of Slovak*. Oxford: Clarendon Press.

Sag, Ivan. 1974. The Grassmann's Law ordering pseudoparadox. *Linguistic Inquiry* 5. 591–607.

Sag, Ivan. 1976. Pseudosolutions to the pseudoparadox: Sanskrit diaspirates revisited. *Linguistic Inquiry* 7. 609–622.

Schindler, Jochem. 1976. Diachronic and synchronic remarks on Bartholomae's and Grassmann's Laws. *Linguistic Inquiry* 7. 622–637.

Shaw, Patricia A. 1976. Theoretical issues in Dakota phonology and morphology. Ph.D. dissertation, University of Toronto.

Shaw, Patricia A. 1985. Modularisation and substantive constraints in Dakota Lexical Phonology. *Phonology Yearbook* 2. 173–202.

Shibatani, Masayoshi. 1990. *The languages of Japan.* Cambridge: Cambridge University Press.

Sihler, Andrew L. 1995. *New comparative grammar of Greek and Latin.* Oxford: Oxford University Press.

Singh, Amar Bahadur. 1969. On echo words in Hindi. *Indian Linguistics* 30. 185–195.

Slocum, Marianna C. 1948. Tzeltzal (Mayan) noun and verb morphology. *International Journal of American Linguistics* 14. 77–86.

Smolensky, Paul. 1995. On the structure of the constraint component *Con* of UG (ROA-86).

Smolensky, Paul. 1997. Constraint Conjunction II. Handout from paper presented at the Hopkins Optimality Theory Workshop/University of Maryland Mayfest 1997.

Smyth, Herbert Weir. 1920. *Greek grammar.* Cambridge, MA: Harvard University Press. Revised 1956 by Gordon M. Messing.

Soden, Wolfram von. 1969. *Grundriß der akkadischen Grammatik.* Rome: Pontificum Institutum Biblicum.

Sohn, Ho-min. 1975. *Woleaian reference grammar.* Honolulu: University of Hawai'i Press.

Sohn, Ho-min & Anthony P. Tawerilmang. 1976. *Woleaian–English dictionary.* Honolulu: University of Hawai'i Press.

Steriade, Donca. 1987. Redundant values. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 339–362.

Steriade, Donca. 1995. Underspecification and markedness. In Goldsmith (1995), 114–174.

Stevens, Kenneth N. & Sheila Blumstein. 1975. Quantal aspects of consonant production and perception: A study of retroflex consonants. *Journal of Phonetics* 3. 215–233.

Suzuki, Keiichiro. 1998. A typological investigation of dissimilation. Ph.D. dissertation, University of Arizona.

Szakos, József. 1994. Die Sprache der Cou: Untersuchungen zur Synchronie einer Austronesischen Sprache auf Taiwan. Ph.D. dissertation, University of Bonn.

Thomas, Jacqueline. 1963. *Le parler ngbaka de Bokanga.* The Hague: Mouton.

Trefry, David. 1969. *A comparative study of Kuman and Pawaian: Non-Austronesian languages of New Guinea.* (*Pacific Linguistics* B13.) Canberra: Australian National University.

Tunley, Alison. 1999. Coarticulatory influences of liquids on vowels in English. Ph.D. dissertation, University of Cambridge.

Uhlenbeck, Eugenius M. 1949. *De structuur van het javaanse morpheem.* Bandoeng: Nix.

Vaux, Bert. 1998. The laryngeal specifications of fricatives. *Linguistic Inquiry* 29. 497–511.

Walsh Dickey, Laura. 1997. The phonology of liquids. Ph.D. dissertation, University of Massachusetts, Amherst.

Waterhouse, Viola. 1949. Learning a second language first. *International Journal of American Linguistics* 15. 106–109.

Waterhouse, Viola. 1962. *The grammatical structure of Oaxaca Chontal.* Bloomington: Indiana University Research Center in Anthropology.

Welmers, William E. 1946. *A descriptive grammar of Fanti.* Baltimore: Linguistic Society of America.

West, Paula. 1999a. The extent of coarticulation in English liquids: An acoustic and articulatory study. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee Bailey (eds.) *Proceedings of the 14th International Congress of the Phonetic Sciences,* 1901–1904. Berkeley: Department of Linguistics, University of California.

West, Paula. 1999b. Perception of distributed coarticulatory properties of English /l/ and /ɹ/. *Journal of Phonetics* 27. 405–426.

West, Paula. 2000. Long-distance coarticulatory effects of English /l/ and /r/. Ph.D. dissertation, University of Oxford.

Wolf, Matthew. 2007. For an autosegmental theory of mutation. In Leah Bateman, Michael O'Keefe, Ehren Reilly & Adam Werle (eds.) *Papers in Optimality Theory III*, 315–404. Amherst: GLSA.

Woodhouse, Robert. 1998. Verner's and Thurneysen's Laws in Gothic as evidence for obstruent development in Early Germanic. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 120. 194–222.

Wordick, Frank. 1982. *The Yindjibarndi language*. (*Pacific Linguistics* C71.) Canberra: Australian National University.

Yip, Moira. 1982. Reduplication and CV skeleta in Chinese secret languages. *Linguistic Inquiry* 13. 637–661.

Yip, Moira. 1988. The Obligatory Contour Principle and phonological rules: A loss of identity. *Linguistic Inquiry* 19. 65–100.

Yip, Moira. 1989. Feature geometry and cooccurrence restrictions. *Phonology* 6. 349–374.

Yip, Moira. 1995. Repetition and its avoidance: The case of Javanese. In Keiichiro Suzuki & Dirk Elzinga (eds.) *Proceedings of the 1995 Southwestern Workshop on Optimality Theory (SWOT)*. Coyote Papers: Working Papers in Linguistics. Conference Proceedings vol. 5, 238–262. Tucson: University of Arizona Linguistics Circle.

Yip, Moira. 1998. Identity avoidance in phonology and morphology. In Steven Lapointe, Diane Brentari & Patrick Farrell (eds.) *Morphology and its relation to phonology and syntax*, 216–246. Stanford: CSLI.

61 Hiatus Resolution

Roderic F. Casali

1 Overview of hiatus resolution

The term *vowel hiatus* is commonly used to refer to a sequence of adjacent vowels belonging to separate syllables, as in the following Hawaiian examples from Senturia (1998: 26). (Periods indicate syllable boundaries.)

(1) [ko.a.na] 'space'
 [Ii.le.a] (name of a shell)
 [ku.a] 'back'
 [hu.e.lo] 'tail'
 [hu.i.na] 'sum'
 [ko.e.na] 'remainder'

In some languages, vowel hiatus is permitted quite freely. Other languages place much stricter limits on the contexts in which heterosyllabic vowel sequences can occur, while some disallow them entirely. Languages that do not permit vowel hiatus may employ any of several processes that eliminate it in cases where it would otherwise arise (e.g. where an underlying vowel-final morpheme directly precedes a vowel-initial morpheme).

One of the most common forms of hiatus resolution involves the elision of one of the two vowels. (See CHAPTER 68: DELETION.) Vowel elision is illustrated below with examples from Yoruba, adapted from Pulleyblank (1988).

(2) /bu ata/ → [ba.ta] 'pour ground pepper'
 /gẹ́ olú/ → [gó.lú] 'cut mushrooms'
 /ta epo/ → [te.po] 'sell palm oil'
 /kɔ́ ɛ̀kɔ́/ → [kɛ́.kɔ̌] 'learn'
 /ra ɔwɔ̀/ → [rɔ.wɔ̀] 'buy a broom'

In all of these examples, it is the first of the two adjacent vowels (V_1) that deletes. Though this is the more common pattern cross-linguistically, cases in which the second vowel (V_2) deletes are also attested (and indeed, some instances of V_2 deletion are found in Yoruba itself).

In another very common hiatus resolution process, *glide formation*, V_1 is converted to a semivowel (see also CHAPTER 15: GLIDES). One well-known case, illustrated in (3), is Ganda (Tucker 1962; Katamba 1985; Clements 1986).[1]

(3)	/mu-iko/	→	[mwiː.ko]	'trowel'	cf. [mu-leːnzi]	'boy'
	/li-ato/	→	[ljaː.to]	'boats'	cf. [li-ggwa]	'thorn'
	/mu-ezi/	→	[mweː.zi]	'moon'		
	/mu-ogezi/	→	[mwoː.ge.zi]	'talker'		
	/mi-ezi/	→	[mjeː.zi]	'moons'		
	/mu-ana/	→	[mwaː.na]	'child'		

In general (we will look at an exception in §2.4 below), glide formation in Ganda applies only where V_1 is high. Non-high V_1's are elided before another vowel, with compensatory lengthening of V_2 (e.g. /ka-oto/ 'small fireplace' > [koːto]).

A third common pattern, *coalescence*, involves the merger of V_1 and V_2 to form a third vowel that combines features of both. This is illustrated in the Attic Greek examples below (de Haas 1988: 126). In these examples, various underlying sequences that combine a non-high [–ATR] vowel /a ɛ ɔ/ with a mid [+ATR] vowel /e o/ are realized phonetically as a long mid [–ATR] vowel that retains the backness and roundness of the original [+ATR] vowel.

(4)	/gene-a/	→	[gé.nɛː]	(/ea/ > [ɛː])	'race (NOM ACC PL)'
	/tiːma-omen/	→	[tiː.mɔ̂ː.men]	(/ao/ > [ɔː])	'honor (1PL PRES IND)'
	/ajdo-a/	→	[aj.dɔ̂ː]	(/oa/ > [ɔː])	'shame (ACC SG)'
	/dɛːlo-ɛːte/	→	[dɛ.lɔ̂ː.te]	(/oɛː/ > [ɔː])	'manifest (2PL PRES SUBJ)'
	/zdɛː-omen/	→	[zdɔ̂ː.men]	(/ɛːo/ > [ɔː])	'live (1PL PRES SUBJ)'

Note that for the pairs /a o/ and /ɛː o/, coalescence in Attic Greek is symmetric; the phonetic result is the same for both orders of input vowels.[2] Other languages with symmetric coalescence include Quebec French, Korean, Rotuman, Old Portuguese, and Classical Sanskrit (all discussed in de Haas 1988), and Afar (Bliese 1981). Symmetric coalescence is relatively uncommon, however. Much more frequently, coalescence applies only when the vowels occur in one of the two possible orders (see §2.3 below).

Other languages avoid hiatus by retaining both vowels but syllabifying them into the nucleus of a single syllable, a process generally known as *diphthong formation* or *diphthongization*. This occurs in Ngiti, as illustrated in the following examples, adapted from Kutsch Lojenga (1994: 90–91).

(5)	/abvo àji/	→	[a.bvoà.ji]	'widow'
	/tìtɔ̀ akpà/	→	[tì.tɔ̀a.kpà]	'liar'
	/opi àji/	→	[o.pìà.ji]	'Lendu woman'
	/ìndɽì akpà/	→	[ìn.dɽla.kpà]	'male goat'
	/fà ɔ́ɲò/	→	[fàɔ́.ɲò]	'our food'
	/fʊ̀kʊ́ oɓi/	→	[fʊ̀.kʊ́ó.ɓi]	'your (PL) knives'

[1] It is also common to find cases in which the second of two vowels becomes non-syllabic, e.g. /gene-i/ 'race (DAT SG)' > [ge.nej] in Attic Greek (de Haas 1988: 126). Generally, such cases are potentially analyzable as diphthong formation. See Senturia (1998: 12–15) for examples and related discussion.

[2] This is not the case in Attic for the pair /a e/: /a+e/ yields [ɛː], while /e+a/ yields [ea].

Kutsch Lojenga states that "both vowels must be realised as a short complex vowel nucleus on one V timing slot." She further notes (personal communication) that the first vowel in each sequence is shorter in duration than the second vowel, though not to the point where any auditory distinctions among vowels in V_1 position are neutralized. This argues against an analysis (i.e. glide formation) in which V_1 is syllabified as a consonantal onset.

Other languages that exhibit diphthong formation include Haitian Creole (Picard 2003), Indonesian (Rosenthall 1997), Attic Greek (Senturia 1998 and references therein), Obolo (Faraclas 1982), Bakossi (Hedinger and Hedinger 1977), Eastern Ojibwa (Howard 1973), Margi (Tranel 1992), and Larike (Rosenthall 1997).

Finally, an obvious means of eliminating vowel hiatus is to epenthesize a consonant between the two vowels. One language in which this occurs is Washo, as illustrated in the examples below, adapted from Midtlyng (2005); in each example a semivowel [j] is inserted between the initial vowel of a suffix and the final vowel of a preceding morpheme.

(6)	a.	'laːdu-a 'my hand-LOC'	→	['laː.du.ja]	'in my hand'
	b.	le'guʔu-iʔ '(1SG OBJ) mother's mother-ATTRIB-AG'	→	[le.'gu.ʔu.ji]	'my daughter's child'
	c.	'lemts'iha-i 'I cause to awake-IMP'	→	['lém.ts'i.ha.ji]	'I am waking him up'
	d.	'lemlu-'eːs-i 'I eat-NEG-IMP'	→	['lém.lu.'jeː.si]	'I am not eating'

Though these hiatus resolution strategies have been presented independently using data from different languages, it is common to find two or more different strategies at work in the same language (see §2.5).

It is also common to find that languages tolerate hiatus in some contexts but not others. A number of factors are capable of blocking or influencing hiatus resolution, including the nature of the prosodic or morphosyntactic boundary at which hiatus arises (Kaisse 1977; Baltazani 2006), prominence factors such as stress (Senturia 1998), vowel length and tone (Casali 1998: 73), minimal word length or weight conditions, the lexical or functional status of particular morphemes, rate of speech, and sensitivity to particular lexical items. Hiatus resolution also sometimes shows *derived environment effects* (see CHAPTER 88: DERIVED ENVIRONMENT EFFECTS), in which hiatus is tolerated in vowel sequences internal to a morpheme, but is eliminated in cases where two vowels come together across a morpheme boundary. Finally, morphemes consisting of just a single vowel are sometimes resistant to loss through elision, presumably due to the loss of semantic content that could result (Casali 1997).

Hiatus resolution can also arise in cases where *three* (or more) underlying vowels occur in sequence. Such cases are considerably less common, and it is difficult to make many strong generalizations about the resolution of $/V_1V_2V_3/$ sequences. Attested outcomes include gliding of V_2 (e.g. Eastern Ojibwa [Howard 1973]; Ganda [Clements 1986: 75]), and elision of both V_1 and V_2 (Baka [Parker 1985]).

The remainder of this paper is organized as follows. In §2, I describe some major respects in which hiatus resolution processes vary across languages. §3 discusses the treatment of hiatus resolution within various theoretical models and some associated challenges and issues. The paper concludes with a brief summary in §4.

2 Typological variation

Hiatus resolution patterns show considerable variation across languages, and any survey of this variation in a work of the present paper's scope will necessarily be selective.[3] Here we will look at certain aspects of variation involving consonant epenthesis (§2.1), vowel elision (§2.2), coalescence (§2.3), and glide formation (§2.4), as well as the co-occurrence of multiple processes within a single language (§2.5).

2.1 *Consonant epenthesis*

A question that naturally arises in looking at hiatus resolution by consonant epenthesis is *which* consonants can function epenthetically as hiatus interrupters. Three possibilities seem reasonably well attested:

(i) A semivowel, usually one that is homorganic with (i.e. shares the same frontness or roundness as) V_1 or V_2.
(ii) A glottal stop ([ʔ]) or fricative ([h]).
(iii) A coronal consonant, generally [t] or a rhotic.

By far the most common pattern (Picard 2003; Uffmann 2007) is the first one. This is sometimes explained (see for example Uffmann 2007) by assuming that homorganic glide epenthesis is in some sense different from (and less costly than) epenthesis of an entirely new segment, since the glide might be interpreted as a prolongation of phonological content that is already present. However, there are also languages – e.g. Ait Seghrouchen Berber (Senturia 1998), Galacian (Picard 2003), and Washo (Midtlyng 2005) – that consistently epenthesize [j], regardless of the featural content of adjacent vowels, and at least one language, Chamicuro (Parker 1989; de Lacy 2006), with consistent [w]-epenthesis.

An example of a language with glottal stop epenthesis is Malay (Ahmad 2001). The examples below show insertion of a glottal stop between a CV prefix and vowel-initial root:

(7)	/di-ubah/	→	[diʔubah]	'to change (PASS)'
	/sə-indah/	→	[səʔindah]	'to be as beautiful as'
	/sə-elok/	→	[səʔeloʔ]	'to be as pretty as'
	/di-olah/	→	[diʔolah]	'to beguile (PASS)'
	/di-aŋkat/	→	[diʔaŋkat]	'to lift (PASS)'

Other languages that epenthesize [ʔ] in at least some hiatus contexts include Ilokano, Selayarese, Tunica, and Indonesian (see Lombardi 2002 and references therein).

[3] One topic that is not treated, for reasons of space, is the typology of diphthong formation. See Schane (1987), Sohn (1987), Rosenthall (1994), and Senturia (1998) for some discussion.

A well-known case of epenthesis of a coronal consonant in hiatus contexts is Axininca Campa (Payne 1981; Lombardi 2002; Baković 2003), illustrated in the examples below (Payne 1981):

(8)	/i-N-koma-i/	→	[iŋkomati]	'he will paddle'
	/i-N-koma-aa-i/	→	[iŋkomataati]	'he will paddle again'

These examples show an epenthetic [t] interrupting vowel hiatus in suffixal contexts; hiatus in prefixal contexts is resolved in Axininca Campa by eliding one of the vowels instead.

The problem of predicting the range of possible epenthetic consonants has received significant attention in recent theoretical work. This is discussed further in §3.3.3 below.

2.2 *Vowel elision*

A natural question that arises in connection with vowel elision is *which* of two adjacent vowels elides. Cross-linguistically, elision of V_1 is far more common than elision of V_2 (Bergman 1968; Lamontagne and Rosenthall 1996; Casali 1997, 1998). Interestingly, it turns out that the contexts in which V_2 elision is well attested are not random. Clear cases of V_2 elision are largely confined to two contexts: (i) the boundary between a lexical (content) word and a following function word, and (ii) stem–suffix boundaries.[4]

Examples of the former type, from Etsako (Elimelech 1976), are shown in (9). Note that the latter also display V_1 elision of the final vowel of a *preceding* function word, suggesting rather strongly that it is lexical or non-lexical status, and not simple linear order, that is relevant in this case (see also CHAPTER 104: ROOT–AFFIX ASYMMETRIES).

(9)	/ɔna	aru	ɔli/	→	[ɔnaruli]	'that louse'
	the	louse	that			
	/ɔna	eɣi	ɔna/	→	[ɔneɣina]	'this tortoise'
	the	tortoise	this			

Examples of the latter type, adapted from Okpe (Pulleyblank 1986), are shown in (10).

(10)	/è-sé-ó/	→	[èsé]	'to fall'
	INF-fall-INF			
	/ɛ̀-dɛ́-ɔ́/	→	[ɛ̀dɛ́]	'to buy'
	INF-buy-INF			

Compare these forms with the additional Okpe words in (11), where the final V suffix is retained following an underlying high vowel, which undergoes glide formation.

[4] In addition to the more common cases in which the elided vowel is one that occupies a particular position, there are also cases (see Casali 1996, 1998; Causley 1999b) in which the vowel targeted depends on the featural makeup of the two vowels.

(11) /è-tí-ó/ → [ètjó] 'to pull'
INF-pull-INF
/ɛ̀-sú-ɔ́/ → [ɛ̀swɔ́] 'to sing'
INF-sing-INF

At other kinds of morphosyntactic boundaries, such as that between a prefix and following root or between two content words, elision regularly targets V_1. The cross-linguistically well-attested possibilities are summarized below. (See Casali 1997 for more discussion.)

(12) *Vowel elision and morphosyntactic position*

Context	*Robustly attested possibilities*
Between two content words	V_1 elision
Content word before function word	V_1 elision or V_2 elision
Prefix + root	V_1 elision
Root + suffix	V_1 elision or V_2 elision

2.3 *Coalescence*

As noted previously, symmetric coalescence, as in the Attic Greek data in (4), is relatively rare. By far the most common form of coalescence is a directionally asymmetric pattern, termed *height coalescence* in Casali (1998) (see also Lamontagne and Rosenthall 1996; Parkinson 1996), in which a non-high V_1 and a high V_2 coalesce to form a non-high vowel otherwise identical to V_2, e.g. /a+i/ > [e], /a+u/ > [o], as in the Xhosa examples below (Aoki 1974).

(13) /wa-inkosi/ → [wenkosi] 'of the chiefs'
/wa-umfazi/ → [womfazi] 'of the woman'

The reverse sequences /i+a/ and /u+a/ are not subject to coalescence in Xhosa, but are resolved instead by vowel elision and glide formation, respectively (see §2.5 below).

Languages in which the feature [ATR] is contrastive sometimes show a slightly more elaborate form of asymmetric height coalescence, in which the [ATR] value of a non-high V_1 is preserved in some cases as well. Such languages divide into two types: those in which [−ATR] is systematically preserved (e.g. /a+i/ > [ɛ], /ɛ+o/ > [ɔ]), and those in which [+ATR] is preserved (e.g. /a+i/ > [e], /o+ɪ/ > [e]). Languages of the former type include Owon Afa (Awobuluyi 1972) and Anufo (Adjekum *et al.* 1993). Languages of the latter type include several North Guang languages and Southern Sotho (Casali 1998, 2003 and references therein).[5]

Though asymmetric height coalescence most commonly applies to sequences in which V_1 is lower than V_2, cases of "reverse height coalescence" also exist in which a higher V_1 followed by a lower V_2 yields a lowered version of V_1 (e.g. /i+a/ > [e], /u+a/ > [o]), while the opposite sequences do not trigger coalescence. This occurs in Foodo (Kwa; Ghana; Plunkett 1991: 68), as shown below. (The initial

[5] The particular [ATR] value that is preserved under height coalescence shows a strong correlation with a language's vowel inventory structure; see Casali (1998, 2003) and Causley (1999a) for discussion.

and final /a/'s are noun class affixes. The tonal changes are due to independent processes discussed in Plunkett.)

(14)	a.	/i+a/	>	[eː]	/á-bì-á/	→	[ábêː]	'seeds'	cf. [dí-bí-lì]	'seed'
	b.	/ʊ+a/	>	[ɔː]	/á-sʊ́-á/	→	[ǎsɔ́ː]	'ears'	cf. [kʊ̀-sʊ́]	'ear'
	c.	/u+a/	>	[oː]	/á-jù-á/	→	[ájǒː]	'millet'	cf. [dú-jú-Ii]	'millet'

Sequences in which V_2 is high and V_1 is non-high do not undergo coalescence; compare the /ʊ+a/ sequence in (14b) with the /a+ʊ/ sequence in /kʊ̀-tá-ʊ́/ 'bow', which is retained in the surface form, [kʊ̀táʊ̀].

Other languages with reverse height coalescence patterns include Tem (Tchagbale 1976; de Craene 1986), Chagga (Nurse and Philippson 1977; Saloné 1980), Ewe (Westermann 1930), Bakossi (Hedinger and Hedinger 1977), and Nkengo (Hulstaert 1970). Interestingly, such patterns seem to occur predominantly at root–suffix boundaries, a restriction that partly parallels some limitations on the distribution of V_2 elision (§2.2).

A further coalescence pattern that should presumably be expected to occur is one in which front unrounded and back rounded vowels coalesce to form a front rounded vowel, e.g. /i+u/ > [y], /e+o/ > [ø], etc. Patterns of this type appear to be considerably less common than height coalescence. Two possible cases, Rotuman and Korean, are discussed in de Haas (1988) (see also Sohn 1987; Rice 1995; Causley 1999a). Coalescence of /e+o/ to [ø] is also described in Obolo (Faraclas 1982).

2.4 *Glide formation*

In Ganda (cf. (3) above) and quite a few other languages, both front and back V_1's are subject to glide formation. It is also quite common, however, to find that only back round vowels glide and that front V_1's trigger a different resolution strategy, most commonly elision. This is the case for example in Xhosa (see §2.5 below) and Chumburung (Snider 1985). Though they are seemingly less common, there are also languages (e.g. Polish; Rubach 2000) in which only front vowels glide.

A second point of variation involves the height of V_1. Generally, if a language has glide formation at all, high V_1's will undergo the process (Rosenthall 1994, 1997; Casali 1995). In some languages (e.g. Ebira; Adive 1989), *only* high V_1's glide. In quite a large number of other languages, however, mid V_1's also glide.[6] One such case, Chicano Spanish, is illustrated in the examples below (from Baković 2007, with phonemic forms substituted for orthographic ones):

(15)	a.	/mi ultima/	→	[mjultima]	'my last one (FEM)'
		/mi obra/	→	[mjoβra]	'my deed'
		/tu epoka/	→	[twepoka]	'your time'
		/tu alma/	→	[twalma]	'your soul'
	b.	/me urxe/	→	[mjurxe]	'it is urgent to me'
		/porke a beses/	→	[porkjaβeses]	'because sometimes'
		/komo eba/	→	[komweβa]	'like Eva'
		/lo abla/	→	[lwaβla]	'speaks it'

[6] In rare cases, e.g. Aghem (Hyman 1979), languages may glide the low vowel /a/ as well.

Further variation exists as well. In some languages, glide formation does not apply to sequences in which V_1 and V_2 share the same frontness and roundness. In Gichode (Casali 1998: 168–169), for example, glide formation of a round vowel occurs only before non-round vowels, e.g. /u+i/ > [wi] but /ʊ+o/ > [o] (*[wo]). (Contrast this with realization of /u+o/ as [wo] in Ganda, as in (3) above.) Glide formation is also blocked in some languages (e.g. Ganda; Clements 1986) following certain consonants.[7] Typically, both sorts of restrictions can be attributed to constraints that are effective quite generally in the language (e.g. languages that fail to glide /u/ or /o/ before a round vowel typically lack [Cw] before round vowels in general).

Finally, some languages impose less stringent restrictions on glide formation when V_1 occurs in absolute word-initial position. In Ganda, for example, only high V_1's generally glide in word-internal /CV_1-V_2/sequences. Word-initially, however, mid and even low V_1's undergo glide formation (in this case without compensatory lengthening), as in the examples below (Clements 1986: 75, n. 1):[8]

(16)	/o-a-gula/	→	[wagula]	'you (SG) bought'
	/a-a-gula/	→	[jagula]	'he/she bought'
	/e-a-laba/	→	[jalaba]	'it (cl. 9) saw'

Rather similar patterns are reported in Nyarwanda (Kimenyi 1979).

Notwithstanding the considerable variation that exists in its patterning, there is one very significant respect in which the behavior of glide formation is surprisingly regular across languages. Quite consistently, (non-word-initial) sequences in which V_1 and V_2 are identical regularly fail to undergo glide formation. We can illustrate this restriction with additional examples from Ganda (Clements 1986):

(17)	/mi-iko/	→	[miːko]	*[mjiːko]	'trowels'
	/lu-uji/	→	[luːji]	*[lwuːji]	'side'

Moreover, sequences such as /o+u/ and /e+i/, in which V_1 and V_2 are both front or both round and V_1 is *lower* than V_2, rarely if ever trigger glide formation, but are resolved instead by vowel elision or coalescence (Casali 1995, 1998: 172, n. 5).

Exceptions to these generalizations clearly arise in absolute word-initial position in some languages, as in the Ganda example in (16) above. I am not aware of any languages that consistently violate these restrictions word-internally, however.

2.5 *Multiple hiatus resolution strategies in the same language*

It is quite common to find two or more different hiatus resolution processes at work in the same language. In some such cases, different processes are operative

7 In some languages, glide formation following certain consonants triggers further changes, e.g. /siV/ and /ziV/ are realized as [ʃV] and [ʒV] respectively in Ebira (Adive 1989).

8 Intervocalic gliding of non-high vowels also occurs in some three-vowel sequences discussed by Clements, as in /te-a-a-gula/ 'he/she didn't buy', realized as [tejagula].

in different morphosyntactic contexts. In Axininca Campa (Baković 2003) and Washo (Midtlyng 2005), for example, hiatus is resolved by vowel elision at a prefix–stem boundary but by epenthesis at a stem–suffix boundary. In Lugisu (Brown 1970), a sequence /a+i/ is resolved by coalescence (to [e]) across a word boundary, but by eliding /a/ word-internally.

There are also many cases, however, in which multiple strategies apply in exactly the same morphosyntactic context, targeting different vowel sequences. Especially common are cases (see Casali 1998: 83–84) in which vowel elision occurs along with glide formation, coalescence, or both. Languages with both vowel elision and glide formation (but not coalescence) include Ganda, Etsako (Elimelech 1976), Igede (Bergman 1968), and Chicano Spanish (Baković 2007). Languages with coalescence and vowel elision (but not glide formation) include Afar (Bliese 1981) and Owon Afa (Awobuluyi 1972). Particularly intricate patterns are found in a considerable number of languages (32 cases are listed in Casali 1998: 83–84) that manifest all three processes. One such language is Xhosa (McLaren 1955; Aoki 1974), whose hiatus resolution alternations conform to the following generalizations:

(18) *Hiatus resolution in Xhosa*

a. Where V_1 is non-high and V_2 is high, the outcome is a [−high] version of V_2.
b. A round V_1 undergoes glide formation before a following non-round vowel.[9]
c. Elsewhere, V_1 elision applies.

The overall pattern corresponding to these generalizations is shown below in Table 61.1, where coalescent realizations are underlined and those involving glide formation are italicized. Note that in the case of the input /o+i/, *both* coalescence and glide formation apply.

Table 61.1 Glide formation, coalescence, and vowel elision (in Xhosa)

				V_2		
		i	e	a	o	u
	i	i	e	a	o	u
	e	$\underline{e}$	e	a	o	$\underline{o}$
V_1	a	$\underline{e}$	e	a	o	$\underline{o}$
	o	$\underline{\textit{we}}$	*we*	*wa*	o	$\underline{o}$
	u	*wi*	*we*	*wa*	o	u

[9] Aoki's description implies that glide formation should apply before round vowels as well, e.g. /u+o/ > [wo], but he gives no examples of such realizations. In contrast, McLaren's data and explicit statements (1955: 10) strongly suggest that gliding of /u/, /o/ occurs only before non-round vowels. I follow McLaren's account here.

Examples illustrating some of these realizations in Xhosa (Aoki 1974) are shown below:

(19)	/esisu-ini/[10]	→	[esiswini]	'stomach (LOC)'
	/ni-oʤa/	→	[noʤa]	'you roast'
	/ndi-akha/	→	[ndakha]	'I build'
	/ni-enza/	→	[nenza]	'you make'
	/wa-ejele/	→	[wejele]	'he fell in'
	/aɓa-oni/	→	[aɓoni]	'wrong doers'
	/aɓa-akhi/	→	[aɓakhi]	'builders'
	/wa-inkosi/	→	[wenkosi]	'of the chiefs'
	/wa-umfazi/	→	[womfazi]	'of the woman'
	/esilo-ini/	→	[esilweni]	'animal'

The descriptive summary of the Xhosa patterns in (18) illustrates something that is quite typical of languages that combine vowel elision with glide formation and/or coalescence, which is that it is generally possible to regard vowel elision as a kind of default process. That is, the simplest way of describing the relevant generalizations is often to specify the conditions under which glide formation and/or coalescence apply, with a statement that vowel elision applies elsewhere.

All three processes – vowel elision, glide formation, and coalescence – can occur either with or without compensatory lengthening, depending on the language (see CHAPTER 64: COMPENSATORY LENGTHENING). Typically, if compensatory lengthening applies with one process it will apply with the others as well. Thus, Ganda shows compensatory lengthening with both vowel elision and glide formation, while Xhosa does not show compensatory lengthening with either of these, nor with coalescence. It also appears generally true that languages (e.g. Ganda) with contrastive vowel length manifest compensatory lengthening while those with no phonemic length do not, but it remains to be seen how universal this correlation is.

3 Theoretical treatments and issues

3.1 *Early generative phonology*

Many analyses of hiatus resolution patterns in particular languages (e.g. Brown 1970; Aoki 1974; Phelps 1975, 1979; Elimelech 1976; Halle 1978; Shaw 1980; Snider 1985) were carried out within early generative phonological frameworks conforming roughly to the model proposed in Chomsky and Halle (1968) or its offshoots. In such models, hiatus resolution processes are due to the operation of language-specific phonological rules. To account for the Xhosa hiatus resolution patterns in (19), for example, Aoki (1974: 239) posits three ordered rules of Vowel Lowering, Glide Formation, and Vowel Deletion, which (with minor notational adjustments) are essentially those in (20):

[10] Aoki (1974: 238) displays the underlying forms of the first and last forms in (19) as /isisu-ini/ and /isilo-ini/, respectively, but describes the lowering of the word-initial vowel form /i/ to [e] as a morphosyntactic replacement, suggesting that initial /e/ is present underlyingly.

(20) a. *Vowel Lowering*

$$V \rightarrow [-\text{high}] \; / \begin{bmatrix} V \\ -\text{high} \end{bmatrix} \_\_$$

b. *Glide Formation*

$$\begin{bmatrix} V \\ +\text{round} \end{bmatrix} \rightarrow [-\text{vocalic}] \; / \; \_\_ \; V$$

c. *Vowel Deletion*

$$V \rightarrow \varnothing \; / \; \_\_ \; V$$

Derivations illustrating the operation of these rules are shown below (Aoki 1974: 40):[11]

(21)			
Underlying Form	/wa-umfazi/	/esisu-ini/	/esilo-ini/
Vowel Lowering	wa-omfazi	—	esilo-eni
Glide Formation	—	esisw-ini	esilw-eni
Vowel Deletion	w-omfazi	—	—
Output	[womfazi]	[esiswini]	[esilweni]

The formal apparatus of early generative phonology frequently offered multiple possibilities for analyzing a given pattern. For example, in contrast to Aoki's analysis of Xhosa coalescence using separate vowel lowering and elision rules, other researchers (e.g. Phelps 1975, 1979; Halle 1978) treated very similar patterns in other languages using a type of rule, known as a *transformational* rule, which is capable of simultaneously affecting (and, in the case of coalescence, merging the features of) two different segments. Perhaps not surprisingly, much of the literature on hiatus resolution patterns of this period focused on issues of rule formulation and the related question of when the rules for two potentially related processes might appropriately be collapsed into a single rule. Aoki's paper, which provides extensive arguments against a transformational rule analysis of coalescence (on the grounds that it is arbitrary and unrevealing and that it leads to an unnecessary increase in the complexity and power of the theory), is itself an interesting case in point. Other relevant work includes Brown (1970), Harms (1973), Hyman (1973), Shaw (1980), Snider (1985), and an extended debate (Chomsky and Halle 1968; Phelps 1975, 1979; Halle 1978) over some particularly intricate patterns in Kasem.

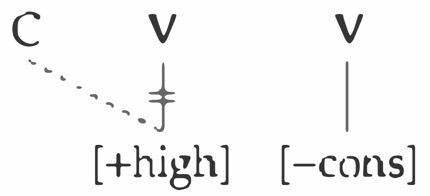
3.2 *Autosegmental and non-linear generative phonology*

The late 1970s and 1980s saw the development of alternative and greatly elaborated *autosegmental* or *non-linear* conceptions of phonological structure in which some or all phonological features are assumed to occur on separate structural tiers (see CHAPTER 14: AUTOSEGMENTS). A number of studies of hiatus resolution phenomena (e.g. Katamba 1985; Clements 1986; Pulleyblank 1986, 1988; Sohn 1987; de Haas 1988; Snider 1989) were carried out using such models. We will look at one representative (and influential) case in some detail, Clements' (1986) treatment

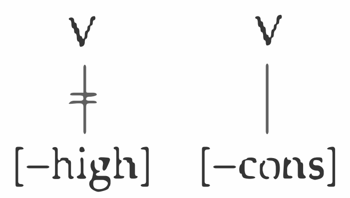
[11] The derivations in (21) differ slightly from those shown in Aoki due to an apparent typo in his derivation of [esiswini] and a minor (and irrelevant) difference in choice of underlying forms (see note 10).

of glide formation and elision in Ganda (see (3) above). Clements' analysis employs the rules in (22).[12]

(22) a. *Glide Formation*

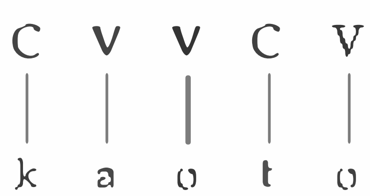


b. *Non-high Vowel Deletion*

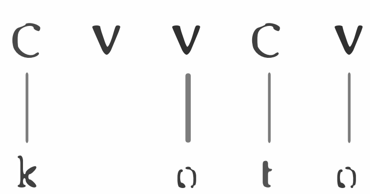


An appealing feature of Clements' analysis is that it provides a very straightforward account of the compensatory lengthening that accompanies both elision and glide formation in Ganda. Both rules in (22) have the effect of delinking a V element from its associated vowel features. This is illustrated below for the case of vowel elision. (23a) shows the underlying form corresponding to /ka-oto/ ([ko:to]) 'small fireplace' within Clements' model, while (23b) shows the representation that results when this form is subjected to rule (22b) (Non-high Vowel Deletion), which delinks /a/ from its associated V element, in conjunction with a further (universal) convention that is assumed to delete unassociated segments (in this case the delinked /a/).

(23) a.

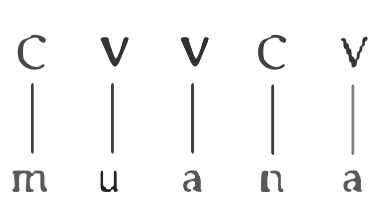


b.

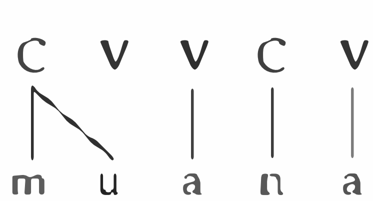


The parallel forms in (24) illustrate the application of the Glide Formation rule (22a). (24a) shows the underlying form of /mu-ana/ ([mwa:na]) 'child', and (24b) shows the result of applying Glide Formation to this form.

(24) a.

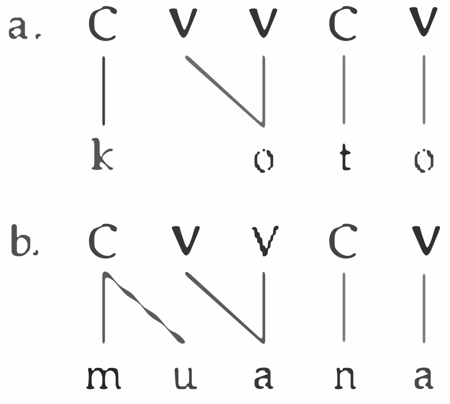


b.



[12] Glide Formation as formulated in (22a) does not account for the cases where non-high vowels glide word-initially in (16). Clements proposes an additional rule to account for these, which we will not treat here.

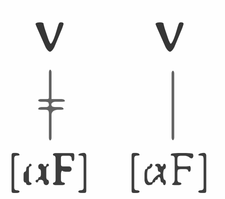
Following the application of these rules, the forms in (23b) and (24b) both contain an unassociated V element. Clements assumes that there is a universal *Linking Convention* that has the effect of automatically reassociating such an unassociated V element to an accessible vowel segment (subject to a general prohibition on crossing of association lines). Applied to the representations in (23b) and (24b), this convention yields the representations in (25a) and (b), respectively.

(25)


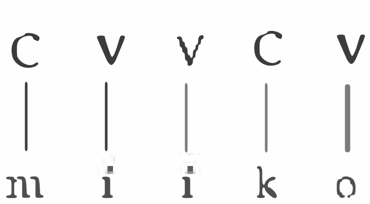
In these surface representations, V_2 emerges as a long vowel, since it is linked to two V elements (see CHAPTER 54: THE SKELETON). This account encodes quite directly the intuition that compensatory lengthening involves the transfer of duration from one segment to another.

There is a further wrinkle to the analysis. As noted in §2.4 above, glide formation does not apply in Ganda to the sequences /i+i/ and /u+u/, in which V_1 and V_2 are identical. To prevent the Glide Formation rule (22a) from applying to these sequences, Clements posits an additional rule of Twin Vowel Deletion that is ordered before Glide Formation and functions to remove sequences of identical high vowels as possible inputs to the latter:

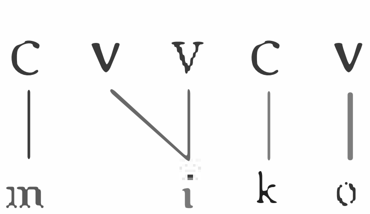
(26) *Twin Vowel Deletion*



This rule is applicable to words like /mi-iko/ [miːko] 'trowels', whose underlying form is shown in (27):

(27)


Application of Twin Vowel Deletion, along with the universal convention requiring deletion of unassociated segments and the Linking Convention that accomplishes reassociation of a free V element, will convert this to (28).

(28)


Since the form in (28) does not meet the structural description for Glide Formation to apply, the analysis correctly predicts [mi:ko] and not *[mji:ko] as the surface form. While the analysis derives the correct forms, however, the need to posit the language-specific rule (26) implies that immunity of sequences of identical vowels from glide formation is an idiosyncratic characteristic of the language. As noted in §2.4, such sequences appear to be regularly exempt from glide formation in other languages as well, suggesting that something more universal than a language-specific rule (26) is at work. (Potentially, this presents an interesting challenge not only for autosegmental models like Clements' but for other approaches as well.)

A strong interest of many autosegmental theories is the specification of phonological features. Much research has been done in particular on the possibility of accounting for certain phonological patterns based on the assumption that only one value of a feature is phonologically specified (see CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION). Underspecification models of this type have potential implications for the analysis of vowel coalescence. One of the analytical questions that arises in connection with coalescence is what determines which features of the two merged vowels are preserved in the output. An interesting general answer to this question, pursued in a study by de Haas (1988) (see also Sohn 1987 and Snider 1989), is that the underlyingly specified features from both vowels are preserved in the output. Preservation of all specified features of both vowels under coalescence would presumably be impossible in cases where the two vowels have opposite values of some feature, since this would lead to a surface vowel simultaneously specified as both [+F] and [−F] for some feature [F]. Following previous work in radical underspecification theory, de Haas assumes that only one value of each feature is underlyingly specified. Consider in this regard the symmetric coalescence of /o/ and /a/ to [ɔ] in Attic Greek, as in the relevant forms in (4) above, repeated here as (29).

(29)	/tiːma-omen/	→	[tiː.mɔ̂ː.men]	(/ao/ > [ɔː])	'honor (1PL PRES IND)'
	/ajdo-a/	→	[aj.dɔ̂ː]	(/oa/ > [ɔː])	'shame (ACC SG)'

In de Haas's underspecification analysis, /a/ is specified only as [+low] and [+back] at the point where coalescence applies, while /o/ is specified only as [+round]. Combining all three feature values yields a [+low], [+back], [+round] vowel, which in de Haas's analysis is equivalent to [ɔ].

Many autosegmental treatments of hiatus resolution processes were also concerned with the relationship between hiatus resolution and syllable structure and attempted to establish a formal connection between the two. In the model of de Haas (1988), for example, (symmetric) coalescence is contingent on prior resyllabification of two adjacent vowels into a single syllable. Other autosegmental analyses that attempted to connect hiatus resolution to syllabification include Katamba (1985), Pulleyblank (1986), Walli-Sagey (1986), Schane (1987), and Sohn (1987).

3.3 *Optimality Theory*

Analyses of hiatus resolution patterns within Optimality Theory (OT) date from the early years of the paradigm and include, among other studies, Rosenthall (1994, 1997), Casali (1995, 1997, 1998), Orie and Pulleyblank (1998), Senturia (1998), Causley (1999a, 1999b), and Baković (2003, 2007). Though they differ somewhat in detail, most such analyses share the following general components:

(i) Some constraint (which must be highly ranked) that militates against heterosyllabic adjacent vowel sequences. There is some controversy (discussed below) over the exact identity of this constraint. For now, we will simply label it "NoHiatus."

(ii) Constraints that are violated by various hiatus resolution possibilities. Generally, vowel elision is assumed to violate a constraint Max, which requires underlying segments to be represented in surface forms. Epenthesis is assumed to violate a constraint Dep against insertion of material (as well as relevant markedness constraints against the features of the inserted consonant – see below). Diphthong formation violates a constraint NoDiph against diphthongs. Glide formation violates, minimally, a markedness constraint, here labeled *CC, against consonant + glide sequences.[13] Coalescence violates a constraint Uniformity, which prohibits merger of two underlyingly distinct segments into a single segment in the output.

Given these assumptions, hiatus resolution is forced whenever NoHiatus is ranked sufficiently high. At a rough first approximation, the particular form of hiatus resolution that occurs is determined by the constraint that is ranked lowest. For example, epenthesis is predicted to occur if the constraint Dep is outranked by the remaining constraints, as illustrated in (30), using a hypothetical input /ku abo/.

(30)

/ku abo/	NoHiatus	Max	NoDiph	*CC	Uniformity	Dep
a. .ku.a.bo.	*!					
b. .kua.bo.			*!			
c. .ka.bo.		*!				
d. .kwa.bo.				*!		
e. .ko.bo.					*!	
☞ f. .ku.ʔa.bo.						*

The simplified analysis sketched above would need to be significantly elaborated to account for the intricate patterns and interactions found in many languages.[14] It does, however, illustrate one important general feature of OT analyses, which is that all phonological processes occur in response to some markedness constraint(s). In this case, the primary markedness constraint is the constraint labeled 'NoHiatus' in (30). One of the issues that has been debated is the exact nature of this constraint. In what follows, we will look briefly at this question and several other important issues that arise within OT approaches to hiatus resolution.

13 Under some analyses (e.g. Baković 2007), gliding of [−high] vowels will also incur violations of a constraint Ident[high], which prohibits changes to the feature [high], since the resulting semivowel [w]/[j] is assumed to be [+high].

14 For some proposed constraints relevant to compensatory lengthening which are not considered in this simplified analysis, see Rosenthall (1997).

3.3.1 *What drives hiatus resolution?*

Many descriptions and analyses of vowel hiatus resolution processes (e.g. Brown 1970; Mtenje 1980; Shaw 1980; Katamba 1985; Pulleyblank 1986; Walli-Sagey 1986; Sohn 1987; de Haas 1988; Wiltshire 1992; Balogné Bérces 2006) have suggested that such processes are motivated by factors related to canonical syllable structure, and in particular the need to avoid onsetless syllables (see CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE and CHAPTER 55: ONSETS). In OT, this notion has often been formalized by high ranking of a constraint ONSET that requires syllables to have onsets, thus disallowing heterosyllabic V.V sequences which would arise in contexts where hiatus is maintained.

An alternative view is that hiatus resolution derives from an avoidance of vowel sequences, and not a requirement that all syllables have onsets. Such a view is made plausible by the observation that vowel hiatus seemingly involves unique phonetic difficulties not found with word-initial onsetless syllables. At least two kinds of difficulty have been cited. First, mutual co-articulatory interaction in a sequence of adjacent vowels tends to perturb the quality of each vowel, potentially making accurate identification of vowel qualities more difficult (Borroff 2003). A different explanation is proposed by de Haas (1988), who sees the problem as a kind of "sonority clash" or "bad syllable contact" (see CHAPTER 49: SONORITY and CHAPTER 21: VOWEL HEIGHT). The adjacent heterosyllabic vowels have (roughly) equal sonority, whereas the preferred transition between syllables should involve a sonority trough. Under the widespread assumption that constraints exist in response to particular phonetic challenges, these considerations lend support to the view that there should be some phonological constraint that specifically excludes hiatus.

Several studies have raised novel arguments that hiatus resolution cannot always be attributed to ONSET. Orie and Pulleyblank (1998) argue that attributing hiatus resolution to ONSET in Yoruba misses important generalizations about the conditions that govern the distribution of different hiatus resolution strategies across different contexts. They adopt instead a constraint NOHIATUS, which is violated by vowels in hiatus but not by onsetless syllables in general. Borroff (2003, 2007) presents data from a number of languages in which the same hiatus resolution patterns found with clear /VV/ sequences apply to /VʔV/ sequences as well. In Chickasaw (Borroff 2007: 57, citing Ulrich 1993), for example, /VV/ hiatus is resolved by glide epenthesis, as shown in (31a). Interestingly, the same process applies to /VʔV/, as in (31b).

(31)	a.	/tof-to-a/	→	[toftowa]	'to spit more than once'
	b.	/boʔ-a/	→	[boʔw-a]	'to be beaten'

On the assumption that an intervocalic [ʔ] should suffice to satisfy ONSET, the fact that the same glide epenthesis process applies even when an intervocalic [ʔ] is present argues that something other than ONSET is responsible for hiatus resolution in this case. Borroff (2003) argues for a constraint VCV-COORD, motivated with reference to phonetic facts involving the sequencing of vowel gestures, which in essence requires that a consonantal target appear between two different vowels.[15]

[15] More precisely, the label VCV-COORD is a shorthand for a conjoined alignment constraint (see CHAPTER 62: CONSTRAINT CONJUNCTION) ALIGN(V_1, release, C_1, target) & ALIGN(C_1, release, V_2, target), which is described in prose as a requirement to "align the release of the first vowel in a sequence of vowels with the achievement of the target of a consonant, and align the release of that same consonant with the achievement of the target of the second vowel of a sequence" (Borroff 2003: 11).

Though the constraint is equivalent for most purposes to NoHiatus, it is (in contrast to Onset) crucially not satisfied by an intervocalic glottal stop, which lacks an (oral) gestural target. An alternative analysis (Borroff 2007) is to assume that a prevocalic glottal stop does not in fact satisfy Onset. In either case, patterns such as these raise interesting challenges for familiar assumptions about hiatus resolution and its motivations.

3.3.2 *Directionality in vowel elision*

Any analysis of vowel elision in hiatus contexts must account for the choice of vowel, V_1 or V_2, that is elided. In rule-based models, the deleted vowel is typically specified directly in the form of the elision rule. For example, both the linear deletion rule (20c) and the autosegmental deletion rule (22b) given above stipulate deletion of the *first* of two adjacent vowels. In contrast, an account within Optimality Theory must assume that elision of V_1 or V_2 in a given context will violate different constraints, whose relative ranking determines which outcome occurs. The problem then becomes to identify the relevant constraints. The possible rankings of these constraints should also suffice to generate the V_1 or V_2 elision cases that are attested cross-linguistically, without predicting patterns that are unattested. Arguably, the relevant generalizations to be accounted for are at least approximately as summarized in (12) above.

A possible account of these generalizations is outlined in Casali (1997). The explanation assumes that at a prefix–root juncture or a boundary between two content words, V_2 is protected by a constraint MaxMI or MaxWI, demanding, respectively, preservation of morpheme- and word-initial vowels. In addition, the analysis continues to assume a generic Max constraint that is violated by deletion of a segment in *any* context. The analysis also assumes a constraint MaxLex requiring preservation of segments in roots and in content words. Crucially, there are no analogous Max constraints that specifically target word- or morpheme-*final* position, or affixes or function words.

A consequence of these assumptions is that in some contexts the constraint violations incurred by elision of V_1 will be a subset of those incurred by V_2 elision. At a prefix–root boundary, for example, elision of V_2 violates MaxMI (since V_2 is the root-initial segment), MaxLex, and (ordinary) Max, while elision of V_1 violates only the latter (assuming we are dealing with a minimally CV prefix, so that V_1 is not morpheme-initial).[16] Since the constraint violations incurred by V_1 elision in this context are a subset of those arising with V_2 elision, eliding V_2 in this context should, all else being equal, be more costly than eliding V_1. Thus, only V_1 elision is ordinarily expected in this context. This is illustrated below, using a hypothetical CV prefix and VCV root. Note that there is no ranking of the constraints under which the second candidate, with V_2 elision, is optimal.

(32)

/CV$_1$-V$_2$CV/	MaxMI	MaxLex	Max
☞ a. CV$_2$CV			*
b. CV$_1$CV	*!	*	*

[16] The full analysis in Casali (1997) actually predicts that V_2 elision should be possible at prefix–root boundaries in the special case of a V prefix, since V_1 is protected by MaxWI and an additional constraint MaxMS requiring preservation of monosegmental morphemes. We will ignore these complications here.

Similarly, only V_1 elision is predicted when underlying vowels abut at the boundary between two content words. In this case, both V_1 elision and V_2 elision violate MAXLEX and general MAX; the two possibilities thus tie on these constraints. However, since V_2 elision violates MAXWI while V_1 elision does not, the former outcome is less optimal. This is illustrated below for a sequence of two hypothetical VCV content words.

(33)

/VCV₁ V₂CV/	MAXWI	MAXLEX	MAX
☞ a. VC V₂CV		*	*
b. VCV₁ CV	*!	*	*

In other contexts, elision of either vowel is predicted to be possible. For example, at a root-suffix boundary V_2 elision violates MAXMI but not MAXLEX. Thus, V_2 elision is possible if MAXLEX outranks MAXMI, as shown below using a hypothetical VCV root and VC suffix:

(34)

/VCV₁-V₂C/	MAXLEX	MAXMI	MAX
a. VCV₂C	*!		*
☞ b. VCV₁C		*	*

V_1 elision, which violates MAXLEX but not MAXMI, is predicted under the opposite ranking:

(35)

/VCV₁-V₂C/	MAXMI	MAXLEX	MAX
☞ a. VCV₂C		*	*
b. VCV₁C	*!		*

For roughly analogous reasons, both V_1 elision and V_2 elision are predicted possibilities at the boundary between a content word and a following function word; in this context V_1 elision violates MAXLEX but not MAXWI, while V_2 elision violates only the latter.

Note that this model encodes no general context-independent preference for elision of V_1; the overall statistical predominance of V_1 elision noted above arises indirectly from the fact that V_1 elision is predicted in a wider range of contexts. An alternative interpretation of the observed typology might suppose that there is a general context-independent preference for preservation of V_2, expressible as some constraint(s), and that this can be overridden in cases where V_1 occurs in a prominent position (and hence falls under the protection of some positional faithfulness constraint). The view that hiatus patterns reveal a general context-independent preference for preservation of V_2 is expressed by Lamontagne and Rosenthall (1996) (see also Alderete 2003), who refer to this effect as the *persistence of* V_2.

Finding evidence to distinguish the two accounts is not easy. There is one context, however, in which the two views potentially make different predictions: where underlying vowels come together morpheme-internally, for example due to the optional deletion of an intervening consonant. While the Casali (1997) model offers no clear predictions in such cases, a model assuming general persistence of V_2 should predict, all else being equal, that V_1 must elide. A number of languages do show vowel elision in such cases, and in at least some of them, this prediction is not borne out. Yoruba (Orie and Pulleyblank 1998; Pulleyblank 1998) and Igbo (Emenanjo 1972) both elide V_2, not V_1, in such cases. Though it might be premature to rule out the possible influence of other factors in these cases, these patterns at least appear to challenge the Persistence of V_2 view, especially since both languages normally show V_1 elision in other contexts (which could be attributed to constraints favoring preservation of initial segments).

3.3.3 *Epenthetic consonants and markedness*

As noted in §2.1 above, only certain consonants are widely observed to function epenthetically as hiatus interrupters. An adequate phonological theory should explain why this is so. Within OT, the problem of explaining the range of possible epenthetic consonants is closely tied to the question of *markedness* (see CHAPTER 4: MARKEDNESS; CHAPTER 12: CORONALS; CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION). Since epenthetic consonants, by definition, are not present underlyingly, their featural content is not affected by faithfulness constraints requiring preservation of phonological material. Consequently, the epenthetic consonant used to resolve hiatus in a given language should be the consonant that is optimal with respect to relevant markedness constraints alone, as these are ranked in the language. The predicted typological range of possible epenthetic consonants should thus follow from the set of universal markedness constraints posited, together with any restrictions (assumed in some models) on their possible rankings.

We can illustrate the basic principles at issue with reference to markedness constraints on place of articulation (POA), which have received much attention in the recent literature. OT models have generally assumed markedness constraints targeting each major POA feature, e.g. the constraints *Lab, *Cor, *Dors, and *Glot, which ban, respectively, labial, coronal, dorsal (e.g. velar), and glottal consonants. All else being equal, the particular epenthetic consonant employed in a language is predicted to have the POA of whichever POA constraint is ranked lowest, e.g. a glottal consonant is expected if *Glot is lowest-ranked.

In a theory in which the possible ranking of these POA constraints varies freely across languages, we should expect that any POA could function epenthetically in some language. However, some phonologists have assumed that certain places of articulation are universally more marked than others. For example, de Lacy (2006) assumes the fixed scale in (36), where ">" means "is more marked than."

(36) dorsal > labial > coronal > glottal

It would be straightforward enough to translate this scale into a universally fixed ranking (i.e. one which is stipulated to hold in all languages as part of Universal Grammar) of POA constraints, as in (37).

(37) *Dors >> *Lab >> *Cor >> *Glot

Fixed rankings of this sort, with some disagreement over details, have played a role in a number of OT analyses (see for example Lombardi 2002). In place of such a fixed hierarchy, however, de Lacy (2006: 2) adopts a different technical implementation of the same general idea, specifically the set of freely rankable POA markedness constraints in (38):

(38) a. *[Dors] Assign a violation for each [dorsal] feature.
b. *[Dors,Lab] Assign a violation for each [dorsal] and each [labial] feature.
c. *[Dors,Lab,Cor] Assign a violation for each [dorsal], each [labial], and each [coronal] feature.
d. *[Dors,Lab,Cor,Glot] Assign a violation for each [dorsal], each [labial], each [coronal], and each [glottal] feature.

In this system, a consonant at a POA further to the left on the scale in (36) will always incur worse violations of these POA constraints than one further to the right. This is because the violations incurred by a POA further to the left are necessarily a superset of those incurred by a POA further to the right, regardless of how these constraints are ranked, as shown below (de Lacy 2006: 50):

(39)

	*{Dors}	*{Dors,Lab}	*{Dors,Lab,Cor}	*{Dors,Lab,Cor,Glot}
k	*	*	*	*
p		*	*	*
t			*	*
ʔ				*

(Crucially, there are no further POA constraints, e.g. *{Cor} or *{Cor,Glot}, targeting other individual places or place combinations.)

If a fixed place markedness hierarchy of this sort were the whole story, we would predict that epenthetic consonants would always be glottals, since an epenthetic glottal consonant is always least costly according to this constraint system. This prediction is too restrictive, as it does not account for various other possibilities (e.g. coronals or a homorganic semivowel) that are reported to exist (see §2.1 above).

De Lacy's solution assumes that there are additional markedness scales that refer to dimensions other than place, and that these interact with the place markedness hierarchy to produce the observed range of typological possibilities. For example, the possibility of epenthesizing a coronal stop [t], as in Axininca Campa, follows from the assumption of an additional set of markedness constraints (this time related not to place but to manner of articulation) against high-sonority consonants in onsets, along with the further (and controversial – see Lombardi 2002; Uffman 2007) assumption that glottal consonants [ʔ] and [h] are higher in sonority than all non-glottal consonants (see chapter 49: sonority). These assumptions motivate a constraint *__ Δ_σ/Glot prohibiting glottals in syllable margins (onsets or codas). In languages in which *__ Δ_σ/Glot is ranked above the relevant POA markedness constraints, glottals will be excluded as epenthetic hiatus interrupters, despite their (universal) optimality with respect to POA alone. With glottals ruled out, the predicted outcome (all else being equal) should be the POA that fares second best according to the constraint system (see (39)). This is coronal.[17] The predicted outcome is illustrated in (40), using a hypothetical input /ai/.

[17] See Lombardi (2002) for a similar proposal.

(40)

/ai/	*__ Δ_σ/ GLOTTAL	*{DORS}	*{DORS,LAB}	*{DORS,LAB, COR}	*{DORS,LAB, COR,GLOT}
a. aki		*	*	*	*
b. api			*	*	*
☞ c. ati				*	*
d. aʔi	*				*

Epenthetic homorganic semivowels (e.g. [w] following /u/, [j] following /i/) are predicted in de Lacy's theory in cases where further markedness constraints requiring consonants (including epenthetic ones) to agree in their place and manner features with adjacent vowels are highly ranked. Additional markedness constraints generate a few further predicted possibilities in languages in which they are highly ranked. In all, the model predicts the following restricted range of epenthetic consonants in hiatus contexts: [ʔ t h ɹ r w j]. De Lacy claims that this set corresponds to the attested range of possibilities.

While de Lacy's theory provides a detailed, plausible, and comprehensive OT account of consonant epenthesis, it is unlikely to be the last word on the subject. The topic of markedness (both with respect to epenthesis and other areas) has been an extremely complex and controversial one. Among other things, some phonologists (see Hume 2003; Rice 2007; CHAPTER 12: CORONALS; CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION) have questioned the claim that glottal (or any other) place of articulation is universally unmarked, arguing that either dorsal or labial (as well as coronal) can also function as the unmarked place in some languages. If this is correct, it would suggest the possibility of epenthetic consonants such as [p] or [k] as well. It remains, perhaps, to be seen whether such cases exist. De Lacy discusses several reported cases, but argues that they are better analyzed in other terms (for example because putative epenthetic consonants in some such cases can be treated as present underlyingly).

At present, a clear understanding of the typology of consonant epenthesis is arguably somewhat clouded by lack of clear consensus on relevant empirical generalizations. Considerable disagreement exists over the interpretation of patterns in some individual languages, a famous example being the question of whether the "intrusive r" phenomenon found in some English dialects (e.g. the pronunciation of *saw it* as [sɔɹɪt] in some Eastern Massachusetts dialects, including my own) constitutes epenthesis (see de Lacy 2006, Lombardi 2002, and Uffman 2007 for discussion of this and other cases). Certain cross-linguistic generalizations have also been disputed. For example, while glottal stop is widely regarded as a frequent choice of hiatus interrupter, Uffmann (2007) proposes that glottal stops are not typically inserted primarily to avoid hiatus, but are generally used (German is cited as one example) to provide an onset in prosodically strong positions, e.g. word-initially or before a stressed vowel, where they function to create a maximized sonority contrast with the following vowel. (This account crucially assumes that glottal stops are the *lowest* sonority consonants, which is exactly the opposite of what de Lacy assumes.) Undoubtedly, there will be further debate over some of the relevant empirical generalizations, as well as their appropriate theoretical treatment.

3.4 *The problem of gradience*

An important distinction in most phonological theories is the distinction between *categorical* and *gradient* processes (see CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY). A categorical change involves a clear "either-or" shift in the presence of one or more segments or their features, as in a case where an underlying segment is removed completely (elision) or undergoes changes in the binary values of one or more features. Frequently, however, languages manifest gradient processes that involve changes in the *degree* of some feature, e.g. a phonemically oral vowel is slightly nasalized next to a nasal consonant but remains less nasal than phonemic nasal vowels in the same language. In hiatus contexts, a possible gradient change might involve the "near elision" of one of the adjacent vowels, e.g. a case where an underlying /V_1 V_2/ sequence is realized phonetically as V_2 (perhaps with lengthening) preceded by a short and variable remnant of V_1.

Hiatus resolution processes have most often been described and analyzed in terms that suggest categorical changes. However, two recent instrumental studies, Baltazani (2006) and Zsiga (1993, 1997), have shown that hiatus resolution patterns (glide formation and/or vowel elision) that had previously been described as categorical in two languages, Modern Greek and Igbo, respectively, actually involve gradient and highly variable timing adjustments. For reasons of space, we will consider only the Igbo case here.

Sequences of adjacent vowels arise very commonly in Igbo in cases where a word ending in a vowel precedes a word beginning in a vowel, as in the phrases shown below (from Zsiga 1997; the diacritics mark [–ATR] vowels).

(41) /a̙sa̙to̙ a̙to̙/ 'three sevens'
/o̙ti̙ o̙zo̙/ 'another grub'
/ezi a̙to̙/ 'three loans'
/ede a̙to̙/ 'three coco-yams'

Three Igbo subjects in Zsiga's (1993) study each produced six repetitions of each of these and various similar phrases in which one of the eight Igbo vowels occurs word-finally before one of the words [a̙to̙] 'three' or [o̙zo̙] 'another'. (In all, each of the eight vowels was used in two utterances.)

Vowel formant measurements of the digitized recordings showed extreme variation, even for the same utterance produced by the same speaker, in the realization of the underlying vowel sequences. These ranged from tokens showing essentially no deletion or assimilation (i.e. in which both vowels clearly surface) to those showing complete loss of V_1 (i.e. with the output consisting entirely of a lengthened version of V_2). If all the observed outcomes were of one of these two types, this might suggest a categorical but optional rule eliding V_1 with compensatory lengthening of V_2 (or a rule of total assimilation of V_1 to V_2). Importantly, however, the results show a range of intermediate realizations as well, in which formant values near the beginning of the vocalic span show a quality intermediate between V_1 and V_2. This intermediate quality varies across repetitions of the same utterance from one that is more similar to V_1 to one that is more similar to V_2. Zsiga argues that such findings are not easily reconcilable with an analysis that treats hiatus resolution as optional but categorical, and that the process is better understood as an adjustment in the relative timing of V_1 and V_2. More specifically, achievement of the target articulatory gestures for V_2 varies from relatively late (allowing for a more or less normal manifestation of a preceding V_1) to relatively early (resulting in partially assimilated tokens) to virtually at the release of the preceding consonant (in which case V_1 is essentially gone). Seen from this perspective, superficial instances of categorical deletion in some of the tokens are better regarded as simply the extreme endpoint of a process that applies along a continuum.

Though specific proposals vary, it has been widely assumed that the familiar kinds of phonological rules and/or constraints standardly used in the analysis of categorical processes are not appropriate to the treatment of gradient sound changes. Zsiga analyzes gradient hiatus resolution in Igbo using the framework of articulatory phonology (Browman and Goldstein 1990), a model that is well suited to handling variable adjustments in the relative timing of gestures.

In addition to highlighting the importance of (and need for additional) explicit theoretical treatments of gradient changes in hiatus contexts, these studies raise an important empirical issue as well. Hiatus resolution in both Igbo and Modern Greek had been described in some previous studies as categorical. This raises the possibility (see Zsiga 1997: 265) that other hiatus resolution processes that have been described as categorical in the literature might turn out to be gradient upon closer examination. Studies such as Baltazani's and Zsiga's underscore the need for careful attention to the possibility of gradience in the context of descriptive phonological fieldwork.

4 Summary

Hiatus resolution patterns are extremely varied. This chapter has provided a brief and necessarily selective look at some of the variation that occurs in the behavior of particular hiatus resolution processes and in their co-occurrence and interaction.

The range of explanatory models that have arisen in connection with hiatus resolution phenomena is also very broad. We have looked at a sample of theoretical proposals from several time periods, including early generative treatments, autosegmental analyses, and several OT models. The central research questions have varied somewhat from model to model. Whereas rule formalism and related issues were a central concern in early generative analyses, autosegmental analyses used more elaborated phonological representations to suggest new solutions to problems such as compensatory lengthening, the featural output of coalescence, and the role of syllable structure in hiatus resolution. Issues that have arisen within OT include the primary markedness constraint that triggers hiatus resolution, the constraint rankings that determine which of two adjacent vowels elides, and the problem of accounting for the range of consonants that can function epenthetically as hiatus interrupters. Finally, we have looked briefly at an issue, gradient hiatus-related processes, which poses potentially important theoretical and empirical challenges for any approach.

ACKNOWLEDGMENTS

I am grateful to Beth Hume, Marc van Oostendorp, Keith Snider, and two reviewers for their valuable comments and suggestions on an earlier version of this paper, and to Sudharsan Seshadri Nagarajan for assistance with relevant library research.

REFERENCES

Adive, John Raji. 1989. *The verbal piece in Ebira.* Arlington: Summer Institute of Linguistics & University of Texas at Arlington.

Adjekum, Grace, Mary E. Holman & Thomas W. Holman. 1993. *Phonological processes in Anufɔ.* Legon: Institute of African Studies, University of Ghana.

Ahmad, Zaharani. 2001. Onset satisfaction and violation in Malay: An optimality account. In Graham W. Thurgood (ed.) *Papers from the 9th Annual Meeting of the Southeast Asian Linguistics Society, 1999*, 135–159. Tempe: Arizona State University Program for Southeast Asian Studies.

Alderete, John. 2003. Structural disparities in Navajo word domains: A case for LexCat-Faithfulness. *The Linguistic Review* 20. 111–157.

Aoki, Paul K. 1974. An observation of vowel contraction in Xhosa. *Studies in African Linguistics* 5. 223–241.

Awobuluyi, Oladele. 1972. The morphophonemics of Owon Afa. *Research Notes, Department of Linguistics and Nigerian Languages, University of Ibadan* 5. 25–44.

Baković, Eric. 2003. Conspiracies and morphophonological (a)symmetry. Paper presented at the 1st Old World Conference in Phonology, Leiden. http://works.bepress.com/cgi/viewcontent.cgi?article=1041&context=ebakovic (April 2010).

Baković, Eric. 2007. Hiatus resolution and incomplete identity. In Fernando Martínez-Gil & Sonia Colina (eds.) *Optimality-theoretic studies in Spanish phonology*, 62–73. Amsterdam & Philadelphia: John Benjamins.

Balogné Bérces, Katalin. 2006. What's wrong with vowel-initial syllables? *SOAS Working Papers in Linguistics* 14. 15–21.

Baltazani, Mary. 2006. Focusing, prosodic phrasing, and hiatus resolution in Greek. In Louis M. Goldstein, Douglas Whalen & Catherine T. Best (eds.) *Laboratory phonology 8*, 473–494. Berlin & New York: Mouton de Gruyter.

Bergman, Richard. 1968. Vowel sandhi in Igede and other African languages. M.A. thesis, Hartford Seminary Foundation.

Bliese, Loren F. 1981. *A generative grammar of Afar.* Arlington: Summer Institute of Linguistics.

Borroff, Marianne L. 2003. Against an Onset approach to hiatus resolution. Paper presented at the 77th Annual Meeting of the Linguistic Society of America, Atlanta (ROA-586).

Borroff, Marianne L. 2007. A landmark underspecification account of the patterning of glottal stop. Ph.D. dissertation, Stony Brook University.

Browman, Catherine P. & Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341–376. Cambridge: Cambridge University Press.

Brown, Gillian. 1970. Syllables and redundancy rules in generative phonology. *Journal of Linguistics* 6. 1–17.

Casali, Roderic F. 1995. Patterns of glide formation in Niger-Congo: An optimality account. Paper presented at the 69th Annual Meeting of the Linguistic Society of America, New Orleans.

Casali, Roderic F. 1996. A typology of vowel coalescence. *UC Irvine Working Papers in Linguistics* 2. 29–42.

Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73. 493–533.

Casali, Roderic F. 1998. *Resolving hiatus*. New York & London: Garland.

Casali, Roderic F. 2003. [ATR] value asymmetries and underlying vowel inventory structure in Niger-Congo and Nilo-Saharan. *Linguistic Typology* 7. 307–382.

Causley, Trisha. 1999a. Complexity and markedness in Optimality Theory. Ph.D. dissertation, University of Toronto.

Causley, Trisha. 1999b. Faithfulness and contrast: The problem of coalescence. *Proceedings of the West Coast Conference on Formal Linguistics* 17. 117–131.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1986. Compensatory lengthening and consonant gemination in LuGanda. In Wetzels & Sezer (1986), 37–77.

Craene, Robert de. 1986. Le verbe conjugé en Tem. *Studies in African Linguistics* 17. 1–37.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.

Elimelech, Baruch. 1976. A tonal grammar of Etsako. *UCLA Working Papers in Phonetics* 35. Available at http://escholarship.org/uc/item/7qd5v492.

Emenanjo, E. N. 1972. Vowel assimilation in Igbo. *Research Notes, Department of Linguistics and Nigerian Languages, University of Ibadan* 5. 7–18.

Faraclas, Nicholas. 1982. Elision and other morpheme boundary phenomena in the western dialects of Obolo. *Journal of West African Languages* 12. 69–82.

Haas, Wim G. de. 1988. A formal theory of vowel coalescence: A case study of ancient Greek. Ph.D. dissertation, University of Nijmegen.

Halle, Morris. 1978. Further thoughts on Kasem nominals. *Linguistic Analysis* 4. 167–185.

Harms, Robert T. 1973. How abstract is Nupe? *Language* 49. 439–446.

Hedinger, Robert & Sylvia Hedinger. 1977. *Phonology of Akɔɔse (Bakossi)*. Yaoundé: Summer Institute of Linguistics.

Howard, Irwin. 1973. *A directional theory of rule application in phonology*. Bloomington: Indiana University Linguistics Club.

Hulstaert, Gustaaf. 1970. *Esquisse du parler des Nkengo*. Tervuren: Musée Royal de l'Afrique Centrale.

Hume, Elizabeth. 2003. Language specific markedness: The case of place of articulation. *Studies in Phonetics, Phonology and Morphology* 9. 295–310.

Hyman, Larry M. 1973. Nupe three years later. *Language* 49. 447–452.

Hyman, Larry M. (ed.) 1979. *Aghem grammatical structure. Southern California Occasional Papers in Linguistics* 7.

Kaisse, Ellen M. 1977. Hiatus in Modern Greek. Ph.D. dissertation, Harvard University.

Katamba, Francis. 1985. A non-linear account of the syllable in Luganda. In Didier L. Goyvaerts (ed.) *African linguistics: Essays in memory of M. W. K. Semikenke*, 267–283. Amsterdam: John Benjamins.

Kimenyi, Alexandre. 1979. *Studies in Kinyarwanda and Bantu phonology*. Carbondale, IL: Linguistic Research Inc.

Kutsch Lojenga, Constance. 1994. *Ngiti: A Central-Sudanic language of Zaire*. Cologne: Rüdiger Köppe Verlag.

Lamontagne, Greg & Sam Rosenthall. 1996. Contiguity constraints and persistent vowel parsing. Unpublished ms., University of British Columbia & Ohio State University.

Lombardi, Linda. 2002. Coronal epenthesis and markedness. *Phonology* 19. 219–251.

McLaren, James. 1955. *A Xhosa grammar*. 2nd edn. Cape Town: Longmans, Green & Co.

Midtlyng, Patrick J. 2005. Washo morphophonology: Hiatus resolution at the edges or let them be vowels. *Santa Barbara Papers in Linguistics* 16. 50–63.

Mtenje, Al. 1980. Aspects of Chichewa derivational phonology and syllable structure. M.A. thesis, Southern Illinois University, Carbondale.

Nurse, Derek & Gérard Philippson. 1977. Tones in Old Moshi (Chaga). *Studies in African Linguistics* 8. 49–80.

Orie, Olanike Ola & Douglas Pulleyblank. 1998. Vowel elision is not always onset-driven. Unpublished ms., Tulane University & University of British Columbia.

Parker, Kirk. 1985. Baka phonology. *Occasional Papers in the Study of Sudanese Languages* 4. 63–85.

Parker, Steve. 1989. The sonority grid in Chamicuro phonology. *Linguistic Analysis* 19. 3–58.

Parkinson, Frederick B. 1996. The representation of vowel height in phonology. Ph.D. dissertation, Ohio State University.

Payne, David L. 1981. *The phonology and morphology of Axininca Campa.* Austin, TX: Summer Institute of Linguistics.

Phelps, Elaine. 1975. Simplicity criteria in generative phonology: Kasem nominals. *Linguistic Analysis* 4. 297–332.

Phelps, Elaine. 1979. Abstractness and rule ordering in Kasem: A refutation of Halle's maximizing principle. *Linguistic Analysis* 5. 29–69.

Picard, Marc. 2003. On the emergence and resolution of hiatus. *Folia Linguistica Historica* 24. 44–57.

Plunkett, Gray C. 1991. The tone system of Foodo nouns. M.A. thesis, University of North Dakota.

Pulleyblank, Douglas. 1986. Underspecification and low vowel harmony in Okpe. *Studies in African Linguistics* 17. 119–153.

Pulleyblank, Douglas. 1988. Vowel deletion in Yoruba. *Journal of African Languages and Linguistics* 10. 117–136.

Pulleyblank, Douglas. 1998. Yoruba vowel patterns: Deriving asymmetries by the tension between opposing constraints. Unpublished ms., University of British Columbia.

Rice, Keren. 1995. On vowel place features. *Toronto Working Papers in Linguistics* 14. 73–116.

Rice, Keren. 2007. Markedness in phonology. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 79–97. Cambridge: Cambridge University Press.

Rosenthall, Sam. 1994. Vowel/glide alternation in a theory of constraint interaction. Ph.D. dissertation, University of Massachusetts, Amherst.

Rosenthall, Sam. 1997. The distribution of prevocalic vowels. *Natural Language and Linguistic Theory* 15. 139–180.

Rubach, Jerzy. 2000. Glide and glottal stop insertion in Slavic languages: A DOT analysis. *Linguistic Inquiry* 31. 271–317.

Saloné, Sukari. 1980. Vowel coalescence and tonal merger in Chagga (Old Moshi): A natural generative approach. *Studies in African Linguistics* 11. 75–100.

Schane, Sanford A. 1987. The resolution of hiatus. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 279–290.

Senturia, Martha B. 1998. A prosodic theory of hiatus resolution. Ph.D. dissertation, University of California, San Diego.

Shaw, Patricia A. 1980. *Theoretical issues in Dakota phonology and morphology.* New York: Garland.

Snider, Keith L. 1985. Vowel coalescence across word boundaries in Chumburung. *Journal of West African Languages* 15. 3–13.

Snider, Keith L. 1989. Vowel coalescence in Chumburung: An autosegmental analysis. *Lingua* 78. 217–232.

Sohn, Hyang-Sook. 1987. On the representation of vowels and diphthongs and their merger in Korean. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23. 307–323.

Tchagbale, Zakari. 1976. Phonologie et tonologie du tem. Ph.D. dissertation, Université de la Sorbonne Nouvelle, Paris.

Tranel, Bernard. 1992–1994. Tone sandhi and vowel deletion in Margi. *Studies in African Linguistics* 23. 111–183.

Tucker, Archibald N. 1962. The syllable in Luganda: A prosodic approach. *Journal of African Languages* 1. 122–166.

Uffmann, Christian. 2007. Intrusive [ɾ] and optimal epenthetic consonants. *Language Sciences* 29. 451–476.

Ulrich, Charles H. 1993. The glottal stop in Western Muskogean. *International Journal of American Linguistics* 59. 430–441.

Walli-Sagey, Elisabeth. 1986. On the representation of complex segments and their formation in Kinyarwanda. In Wetzels & Sezer (1986), 251–295.

Westermann, Diedrich. 1930. *A study of the Ewe language*. London: Oxford University Press.

Wetzels, W. Leo & Engin Sezer (eds.) 1986. *Studies in compensatory lengthening*. Dordrecht: Foris.

Wiltshire, Caroline. 1992. Syllabification and rule application in harmonic phonology. Ph.D. dissertation, University of Chicago.

Zsiga, Elizabeth C. 1993. Features, gestures, and the temporal aspects of phonological organization. Ph.D. dissertation, Yale University.

Zsiga, Elizabeth C. 1997. Features, gestures, and Igbo vowels: An approach to the phonology–phonetics interface. *Language* 73. 227–274.

62 Constraint Conjunction

Megan J. Crowhurst

> . . . conjunctive interaction is from a formal point of view entirely natural in OT; indeed, in an important sense, its *absence* would be *unnatural*. By this I mean simply that without conjunction, basic OT typologies are not strongly harmonically complete, but with conjunction, they are. (Smolensky 2006: 139)

Interactions among constraints in early Optimality Theory (Prince and Smolensky 1993) were limited to a mode of strict domination, represented by the connective ">>." However, researchers soon noted that a theory of constraint interaction relying on strict domination did not adequately account for certain well-known sound behaviors, prompting arguments in favor of additional modes of interaction. Three main proposals for combining simple constraints into more complex ones using connectives other than ">>" were advanced: local conjunction, an analog of Boolean conjunction, and material implication.

1 Classical Optimality Theory

Optimality Theory (OT) made its debut in linguistics through the work of Prince and Smolensky (1993), who adapted a constraint-based model with a long history in other fields (including evolutionary biology, the information sciences, and economics) for the formal analysis of sound phenomena in language. In Prince and Smolensky's model, input states are mapped to output states through a procedure in which a set of potential analyses of the input (output candidates) is passed through a filter consisting of an evaluator, Eval, and a hierarchy of constraints restricting properties of the output. The constraints are members of Universal Grammar, and the priorities assigned to them vary by language. OT constraints are soft constraints – any constraint can be violated, but only under pressure from a constraint with higher priority. The predicted output is the candidate that survives the filter. This candidate is *optimal* in that it represents the input–output mapping that best satisfies the hierarchy by minimizing violations of higher-ranking constraints at the expense of lower-ranking ones.

According to the doctrine of strict domination, if two constraints A and B interact, then constraint A outranks B (or vice versa) in the hierarchy, notated as

A >> B (or B >> A). An interaction A >> B is exposed when the requirements of A and B are incompatible, and there is an input whose properties are such that no viable candidate satisfies both constraints. In such cases, the higher-ranking constraint A is easy to identify because its effect is evident, while B's effect is obscured. (Constraint B may, however, take effect when A is not at stake.) The following tableau illustrates the evaluation of a mini-set of two candidates, each violating either A or B and not the other. Being higher-ranked, A acts on the candidate set first: the candidate which best satisfies A is kept, and the less successful candidate (in regard to constraint A) is rejected. In this simple example, the candidate that fares best on A is the optimal candidate. Were the rankings of A and B reversed, candidate 2 would be the winner instead.

(1) *Constraint A >> Constraint B*

Input	Constraint A	Constraint B
☞ a. output candidate 1		*
b. output candidate 2	*!	

Given that OT is a theory of constraint interaction, an important issue naturally concerns the ways in which constraints might interact. As noted, for Prince and Smolensky and much subsequent work, only strict domination was sanctioned as a mode of constraint interaction.[1] Over time, various researchers have challenged this doctrine in considering what other relationships might hold among constraints. Some researchers have proposed a relation of *non-dominance*, in which neither constraint A nor B dominates the other (Crowhurst 2001). Under non-dominance, violations of A and B together are evaluated cumulatively. Others have argued that constraints can exist in a free ranking relationship to account for optional sound phenomena (e.g. Reynolds 1994; Prince 2001; Itô and Mester 2003a; Jacobs 2004). This chapter reviews proposals that have addressed the question "Can simple constraints combine with one other, and if so, how?" Or, "What connectives other than '>>,' if any, define relationships that can hold among constraints?"

2 Local conjunction: Rejecting the "worst of the worst"

The first proposal for combining constraints beyond the standard mode of strict domination, introduced in Smolensky (1993, 1995, 1997) and subsequently worked out in a series of presentations culminating in Smolensky (2006), was that elemental OT constraints can be *locally conjoined* to form a more complex constraint that is violated only if both of its members are violated in a specified domain. Other influential discussions of the details and applications of local conjunction appeared in a series of papers by Itô and Mester (e.g. 1996, 1998, 2003a), culminating in Itô and Mester (2003b). Smolensky's definition of local conjunction appears in (2) (Smolensky 2006: 43).

[1] Discussion here is limited to proposals advanced in the core OT phonological literature. Combinatorial devices have been used uncontroversially in probabilistic models of OT.

(2) *Local conjunction within a domain D*

*A $\&_D$ *B is violated if and only if a violation of *A and a (distinct) violation of *B both occur within a single domain of type D.

In evaluating local conjunctions, EVAL returns a mark "*" only when both conjuncts are violated, (3a). Some authors have observed (e.g. Hewitt and Crowhurst 1996) that local conjunction is in fact analogous to logical disjunction: "*" is equivalent to False (F), and the absence of a mark is equivalent to True (T). This can be seen by comparing (3a) with the truth table for logical disjunction.

(3) a. *Evaluation of a local conjunction*

	C_1	$C_1\&_DC_2$	C_2
Cand1			
Cand2	*		
Cand3			*
Cand4	*	*!	*

b. *Logical disjunction*

	C_1	$C_1 \vee C_2$	C_2
Cand1	T	T	T
Cand2	F	T	T
Cand3	T	T	F
Cand4	F	F	F

Itô and Mester (2003b: 24) and Smolensky (2006) assume that "&" is a combinatorial operation made available by UG to individual languages, which may *activate* "&" to derive complex constraints on a language-specific basis. On this view, all of the constraints specified in Con_{UG}, the universal set, plus any language-specific local conjunctions, are mapped onto a larger, language-specific constraint set, Con_G. So, ">>" determines strictly hierarchical rankings, while "&" combines constraints into "superconstraints," which can then be inserted into hierarchies defined by ">>."[2] The role of the "&" operator in grammars is formally defined by Itô and Mester (2003b: 25) as in (4).

(4) *Role of local conjunction in grammars*

A grammar G can expand the basic constraint set Con inherited from Universal Grammar to a superset $\text{Con}_G = \text{Con} \cup \{C_1\&C_2\}$, for $C_1, C_2 \in \text{Con}$. Expansion is potentially recursive, so that Con_G can in turn be extended to a superset Con_G' by adding $C_3\&C_4$ to Con_G, for $C_3, C_4 \in \text{Con}_G$, and so on.

2 But see Itô and Mester (1998) and Baković (2000) for a slightly different view.

The main goal of local conjunction is to derive empirical generalizations about markedness from irreducible principles. Constraint conjunction and other proposals for combining constraints have been criticized on the grounds that they greatly increase the expressive power of the formal architecture of OT by exponentially expanding the constraint set. However, the insights, the improvements in precision, and to some extent the economies achieved with local conjunction have often been impressive.

2.1 Coda conditions

The earliest works to employ local conjunction (e.g. Itô and Mester 1996; Spaelti 1997; Smolensky 2006) noted that the mechanism could be used to advantage in analyzing coda conditions, constraints which impose strict conditions on syllable codas rather than penalizing them outright (CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE; CHAPTER 53: SYLLABLE CONTACT). Coda conditions have often been captured by the schema CODACOND[α], where α denotes restricted features. A common requirement is for obstruents in coda positions to be voiceless, other factors being equal (e.g. German, Turkish, Russian; see CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). The coda condition for voicing appears in (5).

(5) CODACOND[voi]

$$\left. \begin{array}{c} *\text{C} \\ | \\ \begin{bmatrix} -\text{son} \\ +\text{voice} \end{bmatrix} \end{array} \right]_\sigma$$

(One * for any voiced obstruent syllabified exclusively as a syllable coda.)

Early work promoting local conjunction noted that coda conditions can be treated as local conjunctions of two well-established constraints, NOCODA in (6a) on the one hand, and a markedness constraint such as (6b) on the other, to yield (6c), with the segment as the local domain of evaluation (Smolensky 1993, 2006; Itô and Mester 2003b; see also Morris 2002).

(6) a. NOCODA
Syllables do not have codas. (One * per syllable with a coda.)

b. *VOIOBST
*[−sonorant, +voice]. (One * for any voiced obstruent.)

c. NOCODA $\&_{SEG}$ *VOIOBST
(One * for any segment which is in a syllable coda *and* which is a voiced obstruent.)

Local conjunctions such as (6c) capture the intuition that constraints expressing coda conditions are really more restrictive versions of NOCODA. A standard assumption has been that a local conjunction is ranked above its conjuncts, and this is consistent with the understanding that the effects of special constraints are visible only when some constraint intervenes between the special constraint and a related, less highly ranked general constraint. In the case under discussion, restrictions on codas are visible in a grammar only when the following are true: (i) Max$_{IO}$(Seg) outranks both instantiations of NoCoda – the unconjoined version and the local conjunction in (6c) – so that the effects of the conjunction are not obscured by deletion; and (ii) Ident$_{IO}$[voi] ranks above the unconjoined constraint *VoiObst, allowing a surface obstruent voicing contrast, but below (6c), so that only voiceless obstruents will occur in coda position, unless higher-ranking constraints demand otherwise (as in cases of regressive voice assimilation). A tableau making this point is given in (7) (adapted from Itô and Mester 2003b: 27–28).

(7)

liːb	Max$_{IO}$(Seg)	NoCoda &$_{SEG}$ *VoiObst	Ident$_{IO}$[voi]	*VoiObst	NoCoda
☞ a. liːp			*		*
b. liːb		*!		b	*
c. liː	*!				

liːbə	Max$_{IO}$(Seg)	NoCoda &$_{SEG}$ *VoiObst	Ident$_{IO}$[voi]	*VoiObst	NoCoda
☞ a. liːbə				b	
b. liːpə			*!		

Three advantages to the reformulation of (5) as (6c) are immediately apparent. First, locally conjoining *VoiObst with NoCoda avoids the redundancy that occurs when *VoiObst is stated twice, once as the general feature co-occurrence constraint in (6b), and once again, embedded in the coda condition, (5). Second, restating a given CodaCond constraint as the local conjunction of NoCoda with a standard markedness constraint explains why we have restrictions on coda consonants, but no constraints that express similar restrictions on onsets (Itô and Mester 2003b: 29). In a grammar that allows conjunction, nothing prevents the local conjunction of Onset and a markedness constraint like *VoiObst, but as tableau (8) shows, a conjunction like Onset &$_{SEG}$ *VoiObst could never be violated: if Onset is violated, then there is no onset consonant to check against *VoiObst. Conversely, if *VoiObst is violated by a consonant in onset position, then Onset is clearly not violated.

(8)

bit	Onset &$_{SEG}$ *VoiObst	*VoiObst	Onset
a. bit	*!	b	
☞ b. pit			
c. it			*!

This latter result of local conjunction weighs strongly in its favor: given that *VoiObst encodes a standard generalization about markedness, and given that coda conditions stated in traditional terms following the model of (5) are common, it is notable that no evidence has been found for mirror image constraints imposing comparable restrictions on segments in onset position. The traditional approach has been to assume that a formal asymmetry matches the empirical one – that is, that there are no Onset conditions, or constraints imposing CodaCond-like requirements on segments in onset position. As we have just seen, the local conjunction approach requires no formal asymmetry: local conjunctions like Onset $\&_{SEG}$ *VoiObst might exist, but as they can never be active in selecting surface candidates, they will never be rankable, and their existence is moot.

2.2 *Universal markedness hierarchies*

Local conjunction has also been used successfully in explaining the fact that universal markedness hierarchies are preserved in multiple domains. Much work establishes the place harmony scale in (9a) (chapter 22: consonantal place of articulation). In OT grammars, this scale is expressed by the hierarchy in (9b). The subscript "UG" indicates that this ranking is "fixed," or specified in universal grammar, and does not vary across languages.

(9) a. Coronal > Labial, Dorsal
    b. *Lab, *Dors $>>_{UG}$ *Cor

The place markedness hierarchy in (9) is known to be preserved in different domains, and Smolensky (1993, 2006) advances a detailed argument as to why this might be so: given that (9b) is fixed in UG, any hierarchy in which the constraints in (9b) are locally conjoined with other constraints will be similarly fixed. To continue with the example of coda conditions, various authors, including Smolensky (1993, 2006), Zoll (1998), and Itô and Mester (2003b), have observed that locally conjoining the markedness constraints in (9b) with NoCoda produces the hierarchy in (10), which favors coronals over labials and dorsals in syllable codas.

(10) NoCoda $\&_{SEG}$ *Lab, NoCoda $\&_{SEG}$ *Dors $>>_{UG}$ NoCoda $\&_{SEG}$ *Cor

Smolensky (2006) uses the same reasoning to explain the commonly observed proliferation of segmental contrasts among coronal consonants in many segmental inventories, relative to the labial and dorsal classes (chapter 12: coronals). As an example, consider the consonant inventory of Tohono O'odham (Uto-Aztecan) in (11).

(11) *Tohono O'odham consonants*

	coronal				
labial	dental	retroflex	palato-alveolar	velar	glottal
p b	t̪ d̪	ɖ	ʧ ʤ	k g	ʔ
	s̪	ʂ			h
m	n̪		ɲ	ŋ	
w		ɭ	j		

The coronal class dominates the inventory: 11 of the 19 consonants are coronals. Note also that the coronals are represented by three places and three manners of articulation (stops, fricative, and affricates) in the obstruent class, whereas the labials and velars are represented only by stops. To abbreviate Smolensky's point, if a single-feature markedness constraint like *[+cont] (specified for fricatives and affricates; see CHAPTER 28: THE REPRESENTATION OF FRICATIVES; CHAPTER 16: AFFRICATES) can be locally conjoined with the fixed hierarchy in (9b), yielding (12), then a grammar that interposes a constraint IDENT$_{IO}$[Place] above *COR &$_{SEG}$ *[+cont] produces an inventory that has [+continuant] coronals, but no continuant obstruents at other places of articulation.

(12) *LAB &$_{SEG}$ *[+cont], *DORS &$_{SEG}$ *[+cont] >>$_{UG}$ *COR &$_{SEG}$ *[+cont]

2.3 *Feature harmonies*

Local conjunction has also been used to account for patterns of vowel and consonant harmony by locally conjoining markedness constraints to form *feature domains* (see also CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS; CHAPTER 72: CONSONANT HARMONY IN CHILD LANGUAGE; CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS; CHAPTER 118: TURKISH VOWEL HARMONY; CHAPTER 123: HUNGARIAN VOWEL HARMONY). In Smolensky (2006), any contiguous sequence of segments which share one or more phonological features forms a feature domain, and any feature domain is a possible instantiation of the domain of a conjunction. An abbreviated version of Smolensky's (2006: 64) definition of a feature domain is given in (13).

(13) *Definition of feature domain* ("φ" stands for any feature)

A maximal contiguous span of φ-bearers with a common value [±φ] is a [±φ] feature domain D[±φ]. (Thus, by definition, contiguous domains of the same φ value are impossible.)

Smolensky's use of the feature domain in OT builds on earlier work (e.g. Kirchner 1993; Smolensky 1993; Cole and Kisseberth 1994; Cassimjee and Kisseberth 1998), but the insight that feature domains can be analyzed in terms of local conjunctions of constraints is due to Smolensky. Examples of Smolensky's (2006) use of feature domains as restrictors on local conjunction are seen in his treatments of vowel harmony and restrictions on consonant clusters.

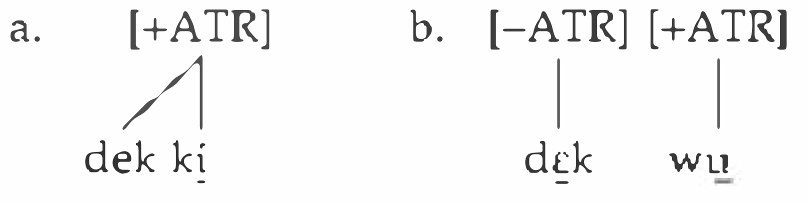
Smolensky (2006) provides an extended discussion of vowel harmony, with special attention given to a pattern of source-conditioned [ATR] harmony found in Lango (Nilotic: Okello 1975; Bavin Woock and Noonan 1979; Noonan 1992; Archangeli and Pulleyblank 1994). The Lango facts are complex; for our purposes, a fragment of Smolensky's account will serve to make the point. Lango has the [+ATR] vowels [i e ə o u] and the [−ATR] vowels [ɪ ɛ a ɔ ʊ] (Noonan 1992). The examples in (14) show a root with a [−ATR] mid vowel /ɛ/ combining with a suffix containing a [+ATR] high vowel, /u/ or /i/. (Tone is omitted in these examples.) In (14a), we see that /ɛ/ assimilates [+ATR] when the suffix vowel is /i/, but not when it is /u/, as in (14b).

(14) *Lango* [ATR] *harmony*

a. *Regressive* [+ATR] *harmony*
/dɛk +Ci/ dek.ki 'your (SG) stew'

b. *No harmony*
/dɛk + wu/ dɛk.wu 'your (PL) stew'

In autosegmental terms, [+ATR] spreads regressively from /i/ to /ɛ/ in [dek.ki], forming a [+ATR] feature domain whose head is the source vowel [i̲], as shown in (15a). (Smolensky marks heads of feature domains with a superscript "o"; we will adopt the convention of underlining heads of feature domains.) The form [dɛk.wu], where no assimilation has applied, has both a [−ATR] and a [+ATR] feature domain.

(15)


According to Archangeli and Pulleyblank's (1994) autosegmental analysis of Lango, [+ATR] spreads regressively from a [+high] source vowel to a target vowel if one of the following is true: (i) the source vowel is [+front] and the target is any vowel in either an open or closed syllable; (ii) the source vowel is not [+front], source and target vowels are both [+high], and the target is in either an open or closed syllable; or (iii) the source vowel is not [+front], and the target is not [+high], and the target is in an open syllable. [+ATR] spread in (14a) meets condition (i), but (14b) meets none of the conditions for harmony.

Following Archangeli and Pulleyblank (1994), Smolensky's analysis (2006) of Lango draws on the insight that ATR harmony is conditioned by the markedness of segments that combine [−ATR] and [+ATR] with other features. Well-established markedness constraints are shown in (16).[3]

(16) a. [ATR] *and backness*

*[+ATR, −front]	(Avoid [+ATR] back vowels.)
*[−ATR, +front]	(Avoid [−ATR] front vowels.)

b. [ATR] *and height*

*[+ATR, −high]	(Avoid [+ATR] mid and low vowels.)
*[−ATR, +high]	(Avoid [−ATR] high vowels.)

c.

$*V_{[+A]}C]_{\sigma}$	(No [+ATR] vowels in closed syllables.)

Returning to the examples in (14), note that in [dek.ki], in which underlying /ɛ/ harmonizes with /i/, the outcome optimizes the constraint *[−ATR, +front], which prefers [e], but at the cost of violating *[+ATR, −high], which prefers [ɛ]. In [dɛk.wu], however, we see the opposite pattern of constraint satisfaction: [ɛ] in the output optimizes *[+ATR, −high] but violates *[−ATR, +front]. The critical difference is in the source vowel. Archangeli and Pulleyblank's intuition, which Smolensky seeks to capture, is that less marked segments make better domain heads, and that when possible, better domain heads propagate their features through harmony. To the constraints in (16), then, Smolensky's account adds *Hd[+ATR] and *Hd[−ATR], which penalize a segment for being the head of a [+ATR] or [−ATR] domain, respectively.

[3] The constraints in (16) and similar constraints, were proposed as elements of the "grounded phonology" framework developed in Archangeli and Pulleyblank (1994).

(17) a. Hd-L[−ATR]
A [−ATR] domain must be left-headed.
(No regressive [−ATR] spread.)

b. (*[−ATR, +front] & *Hd[ATR]) $\&_{D[ATR]}$ F[ATR]
No [+front] head of an unfaithful [−ATR] domain.
(No [−ATR] spread from a [+front] vowel.)

c. *[−ATR, +high]
Vowels do not combine the features [−ATR] and [+high].
(One * for either of [ɪ ʊ].)

d. (*[+ATR, −front] & Hd-L[ATR]) $\&_{D[ATR]}$ (*[+ATR, −high] & $^*V_{[+ATR]}C]_\sigma$ & F[ATR])
A [+ATR] domain with a [−front] head that is not leftmost must be faithful at a [−high] vowel at a closed syllable.
(No regressive [+ATR] spread from a [−front] source onto a [−high] vowel in a closed syllable.)

The constraints in (17b) and (17d) are complex and require further explanation. The embedded local conjunction (*[−ATR, +front] & *Hd[ATR]) penalizes a vowel which is both the head of an ATR domain and has the features *[−ATR, +front]. According to this restriction, the lax vowels [ɪ ɛ] cannot head an ATR domain. The syllable [dɛk] in [dɛk.wu] violates this requirement. Smolensky uses the expression F[ATR] in the "macro" local conjunction more or less as the more standard Ident[ATR] would be used, to require the segments in an ATR domain to be faithful to the value for [ATR] they came with. Conjoining (*[−ATR, +front] & *Hd[ATR]) and F[ATR], taking an ATR domain as the locus of violation, has the effect of minimizing ATR domains headed by [−ATR, +front] vowels. That is, the macro-conjunction prevents [−ATR] spread from a front vowel, and this is why /dɛk + wu/ surfaces as [dɛk.wu] and not [dɛk.wʊ].

By itself, the first embedded local conjunction in (17d), (*[+ATR, −front] & Hd-L[ATR]), would penalize a [+ATR] domain whose head, a [−front] vowel [u], [ə] or [o], is not leftmost within the domain.[4] The second embedded local conjunction, (*[+ATR, −high] & $^*V_{[+ATR]}C]_\sigma$ & F[ATR]), penalizes an unfaithful [+ATR] domain which contains a closed syllable whose vowel is [−high]. The macro-conjunction formed by locally conjoining the two smaller local conjunctions with an ATR domain as the locus of violation means just this: for a [+ATR] domain whose head is on the right and is one of the vowels [u ə o] to pass the macro-conjunction, there cannot be further to the left a closed syllable with a [−high] vowel, which is unfaithful through assimilation to the head. This is why regressive [+ATR] spread does not apply in /dɛk + wu/, so that on the surface we find [dɛk.wu] and not [dek.wu].

Tableau (18) shows why /dɛk + wu/ surfaces as [dɛk.wu] and not [dɛk.wʊ] or [dek.wu] (i.e. why we have neither regressive [+ATR] harmony nor progressive [−ATR] harmony in this case, under Smolensky's analysis).

[4] For example, a [+ATR] domains such as *e . . . u* would be penalized. Three types of ATR domain are permitted by the local conjunction (*[+ATR, −front] & Hd-L[ATR]). If the head is one of the set [i e ɪ ɛ a ɔ ʊ] (i.e. anything other than [u ə o]), then the head is rightmost or leftmost in the ATR domain. The head can be one of the set [u ə o] if it is domain-initial.

(18) *Lango: No regressive* [+ATR] *harmony*

/dɛk+wu/	(17d)	HD-L[−ATR]	(17b)	*[−ATR,hi]	AGREE[ATR]
a. dek.wu̲	*!				
☞ b. dɛ̲k.wu̲					*
c. dɛ̲k.wʊ			*!	*	

Tableau (19) shows how the analysis works for [dek.ki̲] (from /dɛk + Ci/). In this case, progressive [−ATR] is blocked just as for [dɛ̲k.wu̲], and for the same reason. However, in [dek.ki̲], regressive [+ATR] harmony is not blocked, because in this case, neither the head (a front vowel this time) nor the target of assimilation (the vowel /ɛ/, in an open, not closed syllable) violates the macro-conjunction in (17d).

(19) *Lango: Regressive* [+ATR] *harmony*

/dɛk+Ci/	(17d)	HD-L[−ATR]	(17b)	*[−ATR,hi]	AGREE[ATR]
☞ a. dek.ki̲					
b. dɛ̲k.ki̲					*!
c. dɛ̲k.kɪ			*!	*	

The brief discussion offered here falls short of representing the very complicated pattern of [ATR] harmony found in Lango, or of Smolensky's (2006) minutely detailed account of the same. The goal here has been to illustrate a proposal for combining constraints that are already complex into a macro-constraint. More detailed discussions of the Lango pattern can be found in Noonan (1992), Archangeli and Pulleyblank (1994), and Smolensky (2006).

Another very common type of restriction can be seen in conditions on consonant clusters: in languages that permit consonant clusters at all, coronals tend to combine with other consonants much more permissively than do labials and dorsals (CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS). For example, English allows coronals to cluster with other coronals, with labials, and with dorsal consonants, as shown in (20a). However, with the exception of a few loans (e.g. *Akbar*, *Afghan*) and clusters formed across compound boundaries (e.g. [[*cup*][*cake*]], [[*black*][*bird*]]), labials and dorsals do not cluster with either labials or dorsals, so that (with the regular exception of homorganic nasal + obstruent sequences), clusters such as those in (20b) are generally disallowed.

(20) a.

Cor with Cor	Cor with Lab	Cor with Dors
state	*apt*	*acne*
holder	*Abner*	*task*
chortle	*abduct*	*silk*
adze	*help*	*argue*
parlour	*spry*	*disclose*
Atlantic	*almond*	*alcohol*
bolster	*arm*	*agree*
trap	*atmosphere*	*Atkins*

b.	Lab with Lab	Dors with Dors	Dors with Lab
	*pb, *bp	*kg, *gk	*kp, *pk, *gb, *bg
	*pm, *bm		*km, *gm
	*pf, *bv, *fp, etc.		*kf, *fk, *gv, *vg, etc.

The asymmetric distribution of place in consonant clusters is another effect of the place hierarchy in (9). Smolensky (2006) shows that these effects can be accounted for by locally conjoining the constraints in (9b) with one another to produce a cluster place markedness hierarchy like that in (21) ((21) expands Smolensky's hierarchy to include dorsals)

(21) *Place markedness hierarchy for two-consonant clusters*

$$\left\{\begin{matrix} {}^{*}[\text{Lab}]\&_{CL}{}^{*}[\text{Lab}] \\ {}^{*}[\text{Dors}]\&_{CL}{}^{*}[\text{Lab}] \\ {}^{*}[\text{Dors}]\&_{CL}{}^{*}[\text{Dor}] \end{matrix}\right\} \gg_{UG} \left\{\begin{matrix} {}^{*}[\text{Lab}]\&_{CL}{}^{*}[\text{Cor}] \\ {}^{*}[\text{Dors}]\&_{CL}{}^{*}[\text{Cor}] \end{matrix}\right\} \gg_{UG} {}^{*}[\text{Cor}]\&_{CL}{}^{*}[\text{Cor}]$$

In his analysis of phonotactic conditions on consonant sequences, Smolensky (2006) identifies the domain of local conjunction as a consonant cluster, CL. However, a consonant cluster per se is not a unit of phonological structure. Although Smolensky doesn't discuss consonant clusters in this light, any sequence of segments whose members share features can be thought of as instantiating feature domains. An obstruent cluster, for example, would instantiate the D[±φ] formed by the feature [–sonorant], and a nasal-stop cluster would correspond to the feature domain [–continuant].

2.4 *Counterfeeding effects*

One of the more striking advantages claimed for local conjunction is its usefulness in accounting for counterfeeding effects. Moreton and Smolensky (2002) show, refining a proposal in Kirchner (1996), that local conjunction can account for synchronic chain shifts of the type A → B, B → C.[5] In Western Basque, for example, when two identical vowels are juxtaposed, the low vowel /a/ raises to [e] before [a], and /e/ raises to [i] before [e], as shown in (22) (Kirchner 1996; Moreton and Smolensky 2002; Kawahara 2003).

(22)	a.	/a/ → [e] / __ V (violates Ident[low])			
		alaba bat	'daughter (INDEF)'	alabea	'daughter (DEF)'
		neska bat	'girl (INDEF)'	neskea	'girl (DEF)'
	b.	/e/ → [i] / __ V (violates Ident[high])			
		seme bat	'son (INDEF)'	semie	'son (DEF)'
		ate bat	'door (INDEF)'	atie	'door (DEF)'

The Kirchner and the Moreton and Smolensky account goes as follows: if the unmarked vowel is [i], then we'd expect both /a/ and /e/ to raise to [i] in the shifting environment, other factors being equal. However, each vowel moves only one step up, so that /a/ changes its value for the feature [low], and /e/ changes

5 See also Beckman (2003) and Smolensky (2006: §5).

its value for [high]. Neither adjustment results in a change of both features. This effect can be accounted for if a conjunction of the constraints Ident[low] and Ident[high] is ranked above a set of markedness constraints (presumably including at least one OCP constraint) that promote raising in a hiatus; call this set HR. Thus, either of the Ident constraints can be violated on its own, when they are ranked below HR, but no shift violating both constraints within a segment (the relevant domain) can occur. The tableaux in (23), adapted from Moreton and Smolensky (2002), illustrate the analysis for two identical sequences of vowels, /aa/ in /alaba-a/ and /ee/ in /seme-e/.

(23)

/alaba-a/	Ident[low] $\&_{SEG}$ Ident[high]	HR	Ident[low]	Ident[high]
☞ a. alabea			*	
b. alabaa		*		
c. alabia	*!			*

/seme-e/	Ident[low] $\&_{SEG}$ Ident[high]	HR	Ident[low]	Ident[high]
☞ a. semie				*
b. semee		*!		

In the tableaux for [alabea] and [semie], we see that the sequences /aa/ and /ee/ are ruled out by the markedness constraints (the HR set). For /alabaa/, an output [alabia], with raising to [i], is rejected by the local conjunction – the output is the candidate that violates only one of the Ident constraints (in this case Ident[high]). In the case of /semee/, [semie] with a high vowel is possible because raising in this case violates only Ident[high] and not Ident[low]. Violating only one of the Ident constraints, [semie], does not violate the local conjunction.

2.5 *The conjunction of markedness and faithfulness constraints*

So far, we have discussed the local conjunction of markedness constraints, and the conjunction of faithfulness constraints (chapter 63: markedness and faithfulness constraints). Baković (2000) and Łubowicz (2002, 2005) argue for a special role for the conjunction of markedness with faithfulness constraints (M&F conjunction). Baković (2000) shows that M&F conjunction can be used to resolve what he calls the "majority rules" problem in the analysis of voicing assimilation in obstruent clusters. In cross-syllabic obstruent clusters, voicing assimilation is generally controlled by the obstruent in onset position. In Dutch, for example, the medial clusters /kd/ *zakdoek* 'handkerchief' and /dk/ in *bloedkoraal* 'red coral' surface as [gd] and [tk], respectively (Kager 1999).[6] However, patterns of assimilation

[6] The pattern reported here for Dutch is very common. A seemingly odd characteristic of the dialect Kager describes that voiced fricatives in onset position are always devoiced (e.g. /zv/ in *kaasvorm* 'cheese mould' surfaces as [sf]), and this requirement compels the surface voicelessness of any preceding obstruent.

in word-final obstruent clusters are not influenced by requirements on syllable onsets. Consider a language in which an obstruent voicing contrast is preserved in word-final position, which has word-final clusters of two obstruents such as /bt#/ and /pd#/, and in which the markedness constraint AGREE[voice] (obstruent clusters must agree in voicing) dominates the faithfulness constraint IDENT[voice]. In such a case, where all candidates passing AGREE[voice] fail the faithfulness constraint, the candidate containing the voiceless cluster is selected by *VOIOBST. Where the input provides a two-consonant cluster which already satisfies the agreement constraint, for example /pt/ or /bd/, input voicing is generally preserved, showing that IDENT[voice] dominates *VOIOBST (as long as no "spoiler" constraint intervenes).

(24)

/pd/	AGREE[voice]	IDENT[voice]	*VOIOBST
a. pd	*!		*
b. bd		*	*!*
☞ c. pt		*	

/bt/	AGREE[voice]	IDENT[voice]	*VOIOBST
a. bt	*!		*
b. bd		*	*!*
☞ c. pt		*	

This analysis predicts a different outcome for clusters of more than two obstruents. Tableau (25) shows that in such cases the ranking IDENT[voice] >> *VOIOBST predicts that cluster voicing will be determined by the majority. This yields the typologically supported outcome for clusters like /skd/, in which two of three obstruents are voiceless at input (e.g. English *basked*). Unfortunately, it also predicts rare (if at all attested) outcomes like /zgt/ → [zgd] when two of three obstruents are voiced at input.

(25)

/zgt/	AGREE[voice]	IDENT[voice]	*VOIOBST
a. zgt	*!		**
☞ b. zgd		*	***
c. skt		**!	

/skd/	AGREE[voice]	IDENT[voice]	*VOIOBST
a. skd	*!		*
b. zgd		**!	***
☞ c. skt		*	

Baković shows that the "majority rules" problem is avoided if the constraints IDENT[voice] and *VOIOBST can be locally conjoined, assuming the segment as the domain of the conjunction. In (26) we see that on this analysis, of the two clusters that survive AGREE[voice], [zgd] and [skt], the fully voiced cluster [zkd] is rejected even though it is more faithful, because it violates the conjunction IDENT[voice] $\&_{SEG}$ *VOIOBST (which is ranked above both of its conjuncts).

(26)

/zgt/	AGREE[voice]	IDENT[voice] $\&_{SEG}$*VOIOBST	IDENT[voice]	*VOIOBST
a. zgd		*!	*	***
b. zgt	*!			**
c. zkt	*!	*	*	*
☞ d. skt			**	

Baković's larger point is that the presence of marked segments (in this case voiced obstruents) on the surface is one thing, but the presence of *unfaithful* marked segments is another. In cases of assimilation in mixed clusters, the presence of unfaithful marked segments can be compelled by high-ranking faithfulness constraints (e.g. assimilation to a voiced onset in cross-syllabic clusters). However, when such pressures are absent, *assimilation to the unmarked* is the pattern attested across languages. In Baković's example, an analysis that admits the local conjunction IDENT[voice] $\&_{SEG}$ *VOIOBST guarantees the typologically supported outcome, whereas an analysis that relies exclusively on the unconjoined constraints admits the (unsupported) *assimiliation to the marked* alternative.

Łubowicz (2002, 2005) argues that M&F conjunctions are necessary to account for derived environment effects (CHAPTER 88: DERIVED ENVIRONMENT EFFECTS), using palatalization in Polish (CHAPTER 121: SLAVIC PALATALIZATION) as an example. Other works using local conjunction to account for derived environment effects include Downing (2001) and Itô and Mester (2003a).

2.6 *Local self-conjunction*

As a special case of local conjunction, some researchers have argued that markedness constraints can be "self-conjoined," and that self-conjunction offers new insights into cases of dissimilation that have often been attributed to the OCP (e.g. Fukazawa 1999, 2001; Itô and Mester 1998, 2003b; Alderete 2004; Smolensky 2006). As Itô and Mester (2003b: 29) put it:

> the "Obligatory Contour Principle" is by itself neither a constraint nor a formal universal in phonological theory. The culprit in OCP-type dissimilations is not the adjacency of identical feature specifications on a tier, but the multiple presence of a marked type of structure within some domain.

Self-conjunction is intended to capture the intuition that in dissimilation, one violation of a markedness constraint is tolerated, but a double violation crosses what Itô and Mester aptly call a *markedness threshold*, and is categorically rejected.

As an example, let us consider the role of obstruent voicing in Japanese. In the lexical classes comprised by native Japanese and Sino-Japanese forms, voiced obstruents cannot co-occur within a morpheme (CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS), but neither voiceless obstruents nor voiced sonorants are subject to voicing restrictions (Itô and Mester 1998, 2003b). Thus, morphemes such as those in (27a) and (27b) are typical of Japanese (Itô and Mester 2003b: 34–35). However, native Japanese and Sino-Japanese morphemes containing more than one voiced obstruent (e.g. *[kabazi]) are not.

(27)	a.	kusa	'grass'	kataki	'enemy'
		tako	'octopus'	hotoke	'Buddha'
		sato	'village'	tatami	'straw mat'
	b.	kaze	'wind'	kasegi	'earning'
		geta	'clogs'	kagami	'mirror'
		hada	'skin'	-bakari	'only'

The obstruent voicing contrast in Japanese requires the constraint ranking IDENT_{IO} >> *VOIOBST. The fact that multiple occurrences of voiced obstruents within a morpheme is prohibited is accounted for by self-conjoining the constraint *VOIOBST, with the morpheme as the domain. The conjunction $\text{*VOIOBST}^2_{\text{MORPH}}$ (=*VOIOBST $\&_{\text{MORPH}}$ *VOIOBST) is ranked above IDENT_{IO}. Tableau (28) shows how a set of candidates based on the form *kaze* and the hypothetical input /gaze/ are evaluated under these rankings.

(28)

/kaze/	$\text{*VOIOBST}^2_{\text{MORPH}}$	IDENT_{IO}(Seg)	*VOIOBST
☞ a. kaze			z
b. kase		*!	
c. gase		*!*	g
d. gaze	*!	*	g z

/gaze/	$\text{*VOIOBST}^2_{\text{MORPH}}$	IDENT_{IO}(Seg)	*VOIOBST
☞ a. kaze		*	z
b. kase		**!	
☞ c. gase		*	g
d. gaze	*!		g z

Self-conjunction can account not only for passive restrictions on the occurrence of similar structures within a domain, but also cases in which the existence of alternations makes the dissimilation more obvious. A well-known example is Lyman's Law, which blocks the effect of Rendaku, a phenomenon seen in Japanese lexical compounds. Rendaku assigns the feature [+voice] to a voiceless obstruent appearing as the initial consonant of the second member of a compound. The feature [+voice] is treated as a special morpheme in this case, and as such, its presence in the output is required by the constraint RealizeMorph, defined in (29) (Walker 2000; chapter 103: phonological sensitivity to morphological structure).

(29) RealizeMorph

Every morpheme has a phonological exponent in the output. (One * for any null morpheme.)

The Rendaku effect requires RealizeMorph to be ranked more highly than Ident_{IO}(Seg). Tableau (30) is adapted from Itô and Mester (2003b: 37) (constraint names have been changed to match the names used here). The examples used in tableaux (31) and (30) are /ori + kami/ → [origami] 'paper folding' and /kami + kaze/ → [kamikaze] 'divine wind'.

(30)

/ori+R+kami/	RealizeMorph	Ident_{IO}(Seg)	*VoiObst
☞ a. origami		*	g
b. orikami	*!		

The Lyman's Law effect occurs because the self-conjunction $\text{*VoiObst}^2_{\text{Morph}}$ dominates RealizeMorph, neutralizing its effect.

(31)

/kami+R+kaze/	$\text{*VoiObst}^2_{\text{Morph}}$	Realize Morph	Ident_{IO}(Seg)	*VoiObst
☞ a. kamikaze		*		z
b. kamigaze	*(gaze)!		*	g z

Beyond the examples discussed in §2 local conjunction and its special variant, self-conjunction have been used in the analysis of a diverse set of phenomena. Local conjunction has been used in analyses of sonority distance restrictions on syllable structures (Baertsch 1998; Baertsch and Davis 2003; Smolensky 2006), the sonority hierarchy (Smolensky 2006), tone sandhi in Mandarin Chinese (Lin 2000), glide formation in German (Hall 2004, 2007), and accentual phenomena (Alderete 1999). Finally, Levelt *et al.* (1999) and Levelt and van de Vijver (1998) propose that local conjunction plays a role in children's acquisition of the constraint hierarchy of the language they are learning.

3 Other modes of constraint combination

Other proposals for deriving complex constraints from simpler ones using connectives other than "&" have included constraint analogs of Boolean conjunction and two versions of material implication.

3.1 *The "best of the best"*

Hewitt and Crowhurst (1996) and Crowhurst and Hewitt (1997) argue that in addition to local conjunction and self-conjunction, a proper account of some phonological patterns is best achieved using complex constraints derived using a connective, "$\wedge$," with the semantics of Boolean conjunction. Crowhurst and Hewitt call this mode of interaction simply "constraint conjunction." Here, we will call it b-conjunction, to distinguish it from local conjunction. In contrast to local conjunction, which bans the "worst of the worst," b-conjunction insists on the "best of the best." See the definitions and tables in (32) and (33).

(32) a. *Boolean conjunction*
The conjunction $A \wedge B$ is true iff proposition A is true *and* proposition B is true.

b. *B-conjunction*
A candidate Cand passes a b-conjunction $A \wedge B$ iff Cand passes constraint A *and* Cand passes constraint B.

(33) a. *Boolean conjunction*

A	$A \wedge B$	B
T	T	T
F	T	T
T	T	F
F	F	F

b. *Evaluation of a B-conjunction*

	C_1	$C_1 \wedge C_2$	C_2
Cand1			
Cand2	*	*!	
Cand3		*!	*
Cand4	*	*!	*

Crowhurst and Hewitt restrict b-conjunction to constraints that share what they call a *focus* (or *fulcrum* in Hewitt and Crowhurst 1996) which identifies the locus of violation for each of the conjuncts.

(34) *Focus (of a constraint)$_{def}$*

a. Every constraint has a unique focus.
b. A constraint's focus is identified by the universally quantified argument.

Crowhurst and Hewitt's use of b-conjunction is exemplified by their analysis of alignment effects in the stress system of Diyari (Pama-Nyungan; Austin 1981). In Diyari, any morpheme of two or more syllables has initial stress, and secondary stress falls on the penult in morphemes longer than three syllables, as shown in

(35a). The cases of special interest are polymorphemic words containing monosyllabic suffixes, exemplified in (35b). Monosyllabic suffixes are never stressed. Moreover, no syllable preceding *any* morpheme is ever stressed (cf. [ˈpulʲudu-ɳi-ˌmaʈa], [ˈmaɖa-ɭa-ɳi]). Thus, although they have the same number of syllables, the forms [ˈmaɖa-ɭa-ɳi] and [ˈkaɲa-ˌwaɽa] are assigned stress quite differently: the bisyllabic suffix [-ˈwaɽa-] has stress, but the sequence of suffixes [-ɭa-ɳi-] does not.[7]

(35)	a.	ˈkaɲa	(ˈka.ɲa)	'man'
		ˈpinadu	(ˈpi.na)du	'old man'
		ˈŋandaˌwalka	(ˈŋan.da)(ˌwal.ka)	'to close'
		ˈwintaraˌnaja	(ˈwinta)ra(ˌnaja)	'how long'
		ˈkaɲa-ˌwaɽa	(ˈka.ɲa)-(ˌwa.ɽa)	'man-PL'
		ˈtaji-ˌjatiˌmaji	(ˈta.ji)-(ˌja.ti)(ˌma.ji)	'to eat-opt'
	b.	ˈmaɖa-ɭa-ɳi	(ˈma.ɖa)-ɭa-ɳi	'hill-CHAR-LOC'
		ˈpulʲudu-ɳi-ˌmaʈa	(ˈpu.lʲu)du-ɳi-(ˌma.ʈa)	'mud-LOC-IDENT'
		ˈpinadu-ˌwaɽa	(ˈpi.na)du-(ˌwa.ɽa)	'old man-PL'
		ˈɲanda-na-ˌmaʈa	(ˈɲan.da)-na-(ˌma.ʈa)	'hit-PART-IDENT'

The lynchpin of Crowhurst and Hewitt's account of Diyari is the conjunction in (36c) of the two alignment constraints in (36a) and (36b). The individual constraints require an edge (left or right) of a morpheme to be aligned with the same edge of a foot. In conjunction, they require each edge of a morpheme to be aligned with the same edge of a foot.

(36) a. MORPHEMEFT-LEFT (MFL)
ALIGN-L(Morpheme, Ft)
(One * for any morpheme whose left edge does not coincide with the left edge of some foot.)

b. MORPHEMEFT-RIGHT (MFR)
ALIGN-R(Morpheme, Ft)
(One * for any morpheme whose right edge does not coincide with the right edge of some foot.)

c. MFL $\wedge^{\text{MORPH}}$ MFR
(One * per morpheme which fails MFL, MFR, or both.)

Crowhurst and Hewitt treat Diyari as a language that avoids foot structure when possible. They rank the standard constraint PARSE-σ below *STRUC(Ft), which assigns a penalty for each foot in the output. The conjunction in (36c) outranks *STRUC(Ft), resulting in a foot at each edge of morphemes whose syllable count makes this possible. FTMIN(σ) prevents the assignment of monosyllabic feet. Tableau (37) shows how the correct outputs (ˈwinta)ra(ˌnaja) and (ˈta.ji)-(ˌja.ti)(ˌma.ji) are selected under the rankings FTMIN >> MFL $\wedge^{\text{MORPH}}$ MFR >> *STRUC(Ft). (In the following tableaux, violations of the individual conjuncts that are promoted to

[7] Note that the underived pentasyllabic form [ˈwintaraˌnaja] has initial and penultimate (not antepenultimate) stress. Hewitt and Crowhurst (1996) use this example (from Austin 1981) to show that stress in Diyari is not assigned from left to right, as many theoretical accounts of Diyari have assumed (e.g. Poser 1989; Crowhurst 1994; Kager 1994). Rather, Diyari assigns a trochaic foot at both edges of a morpheme when it can.

violations of the conjunction are shown in parentheses. Violations of the alignment constraints are shown with a superscript corresponding to the initial phoneme of the morpheme that incurred the penalty.)

(37)

/taji-jatimaji/	FtMin	MFL	$\wedge^{MORPH}$	MFR	*Struc(Ft)
☞ a. ('ta.ji)-(ˌja.ti)(ˌma.ji)					***
b. ('ta.ji)-ja.ti(ˌma.ji)		(*^{i})	*^{i}!		**
c. ('ta.ji)-(ˌja.ti).ma.ji			*^{i}!	(*^{i})	**
d. ('ta.ji)-ja.ti.ma.ji		(*^{i})	*^{i}!	(*^{i})	*

/wintaranaja/	FtMin	MFL	$\wedge^{MORPH}$	MFR	*Struc(Ft)
☞ a. ('winta)ra(ˌnaja)					**
b. ('win.ta)ra.na.ja		(*^{w})	*^{w}!		**
c. win.ta.ra(ˌna.ja)			*^{w}!	(*^{w})	*

The crucial cases are those that contain monosyllabic affixes. In ['ɲan.ɖa-na-ˌma.ʈa], the monosyllabic suffix [-na-] is not footed, but the two bisyllabic morphemes are aligned at each edge with a foot. The especially interesting case is ['ma.ɖa-ɭa-ɳi], which has adjacent monosyllabic suffixes. The candidate ('ma.ɖa)-(ˌɭa)-(ˌɳi), which satisfies the alignment constraints, is ruled out by FtMin. If the constraints MFL and MFR were not conjoined but were evaluated independently, then of the remaining two, the optimal candidate should be ('ma.ɖa)-(ˌɭa-ɳi), which minimizes violations of MFL and MFR. However, the optimal candidate is in fact ('ma.ɖa)-ɭa-ɳi, in which neither of the suffixes is aligned with foot structure. When MFL and MFR are b-conjoined, however, a morpheme aligned at only one edge is no better than a morpheme aligned at both edges, so that ('ma.ɖa)-(ˌɭa-ɳi) and ('ma.ɖa)-ɭa-ɳi both fail the conjunction MFL $\wedge^{MORPH}$ MFR. In this case, *Struc(Ft) decides for the candidate ('ma.ɖa)-ɭa-ɳi, in which the monosyllabic suffixes are unfooted. The analysis of ['ma.ɖa-ɭa-ɳi] under b-conjunction is shown in (38).

(38)

/maɖa-ɭa-ɳi/	FtMin	MFL	$\wedge^{MORPH}$	MFR	*Struc(Ft)
☞ a. ('ma.ɖa)-ɭa-ɳi		(*^{ɭ} *^{ɳ})	*^{ɭ} *^{ɳ}	(*^{ɭ} *^{ɳ})	*
b. ('ma.ɖa)-(ˌɭa-ɳi)		(*^{ɳ})	*^{ɭ} *^{ɳ}	(*^{ɭ})	*
c. ('ma.ɖa)-(ˌɭa)-(ˌɳi)	*!*				***

The analysis just presented for ['ma.ɖa-ɭa-ɳi] predicts that trisyllabic roots should not be aligned with foot structure, yet, like all other root morphemes, they have initial stress. Crowhurst and Hewitt's analysis attributes the presence of initial stress to the constraint MainStress-L in (39), which insists that the head of the prosodic word stand exactly at the left edge of the prosodic word. Ranking MainStress-L above the conjunction MFL $\wedge^{MORPH}$ MFR compels the presence of an initial foot, even though that foot is poorly aligned with the root morpheme in the output.

(39) MainStress-L (MSL)
Align(PrWd, L; Head(PrWd), L)
(Every PrWd is left aligned with the stressed syllable in its head foot; one * per misaligned PrWd.)

B-conjunction is also useful in stating templatic constraints on the size of reduplicants (chapter 100: reduplication). A constraint like Red=Ft, for example, can be stated as in (40a). However, note that this statement informally assumes the interpretation provided by b-conjunction: any reduplicant must be both left aligned and right aligned with the same foot. Under b-conjunction, the templatic effect is achieved by b-conjoining the constraints in (40b) and (40c), with the foot as the focus of the conjunction. Under this analysis, any foot will be evaluated against the conjunction. Any foot not co-extensive with Red will fail the conjunction, although it might be required by higher-ranking constraints on metrical structure. However, the constraint will be satisfied by reduplicated forms in which Red and a foot are co-extensive.

(40) a. Red=Ft
The reduplicant string is co-extensive with a foot (e.g. Downing 2000: 3).
b. FootRed-L
Align-L(Ft, $Affix_{RED}$)
One * for any reduplicant that is not left aligned with a foot.
c. FootRed-R
Align-R (Foot, $Affix_{RED}$)
One * for any reduplicant that is not right aligned with a foot.
d. FootRed-L $\wedge^{FOOT}$ FootRed-R

The connective "$\wedge$" is also used by Downing (1998, 2000) to account for exceptional alignment behavior found with onsetless syllables (e.g. as stress and tone bearers) and prosodic restrictions on reduplication in Kinande.

3.2 *Implication*

A different form of constraint combination, most closely related to material implication in logic, was proposed by Crowhurst and Hewitt (1997) as a way of analyzing effects of conflicting directionality (Kiparsky 1973; Zoll 1997) in the assignment of stress. As an example, Dongolese Nubian (Armbruster 1960) assigns a single stress, which surfaces on the rightmost of heavy syllable, if there are any (where only vowel length is relevant for syllable weight). Examples are shown in (41a). However, when only light syllables are present, as in (41b), stress falls on the initial syllable.

(41) a. *Heavy syllables*

'bɛɛ.kat.tɪ	'to be killed'
do.'goo.gɪr	'raise it'
tɛ.lɛ.'graaf.kɪ	'a telegram'
tɪn.tɪ.'nɛɛŋ.kɛ.gɪd	'their maternal aunt'
maa.'suu.ra	'tube, pipe'
maa.'lɛɛʃ	'it doesn't matter'
sɛ.rɛɛ.gɪr.ʃug.lɛɛ.rɛ.'daag	'be in the situation of having worked well'

b. *Light syllables*

'bu.run	'it is a girl'
'ta.ra.ga	'page, leaf'
'mu.go.san	'tell to leave'
'ʤɪ.ɲɪ.ran	'tell him (her) to go and wait'

Crowhurst and Hewitt's analysis of Dongolese uses the constraints in (42). The pattern found in words containing heavy syllables seems to require that the constraint HEAVYHEAD in (42a) be ranked above HEADS-R in (42b), which must in turn be ranked above HEADS-L, (42c).

(42) a. HEAVYHEAD
The head syllable of a foot is bimoraic.
(One * for any stressed light syllable.)

b. HEADS-R
ALIGN-R(Head(Ft), PrWd)
(One * per syllable coming between any stressed syllable and the right edge of the dominating PrWd.)

c. HEADS-L
ALIGN-L(Head(Ft), PrWd)
(One * per syllable coming between any stressed syllable and the left edge of the dominating PrWd.)

The problem is that the ranking required for the heavy syllable cases doesn't account for stress in forms containing only light syllables: these seem to require that the alignment constraints be ranked in the reverse order, HEADS-L >> HEADS-R.

The insight Crowhurst and Hewitt (1997) seek to capture is that in a language like Dongolese, heavy syllable stress and any conditions imposed on heavy syllables under stress (in this case, HEADS-R) take priority. HEADS-L becomes relevant only when no heavy syllables are present. Crowhurst and Hewitt propose that HEAVYHEAD and HEADS-R combine in a complex constraint that takes the form of an implication, in which the satisfaction of one requirement is unilaterally dependent on the satisfaction of another. Under their interpretation of material implication, whether a candidate passes A > B depends primarily on the candidate's success on constraint A, and secondarily on its success on constraint B. Crowhurst and Hewitt's definition of implication appears in (43).[8]

(43) *Implication* (Crowhurst and Hewitt 1997)

If Cand passes A, then Cand is evaluated with respect to B:
 If Cand passes B, then Cand passes A > B;
 If Cand fails B, then Cand fails A > B.
If Cand fails A, then Cand fails A > B and Cand's success on B is irrelevant.

As in the case of conjunction, Crowhurst and Hewitt propose that only constraints which share an argument may combine to form implications. Crowhurst

[8] Note that Crowhurst and Hewitt's use of implication does not have the semantics of classical material implication in Boolean logic.

and Hewitt's account of Dongolese ranks the implication HEAVYHEAD $>^{HEAD(FT)}$ HEADS-R above HEADS-L. Tableau (44) shows how their analysis works in forms like ['mu.go.san], which contain only light syllables. All viable candidates violate HEAVYHEAD (constraint A), and this translates to one violation each against the implication HEAVYHEAD $>^{HEAD(FT)}$ HEADS-R. (HEADS-R cells have been shaded to indicate that, per Crowhurst and Hewitt's definition, this constraint has no effect on the outcome.) So it is that in forms containing only light syllables, HEADS-L selects the initially stressed candidate.

(44)

/mugosan/	HEAVYHEAD $>^{HEAD(FT)}$		HEADS-R	HEADS-L
☞ a. ('mu.go)san	(*)	*	(**)	
b. mu('go.san)	(*)	*	(*)	*!
c. mu.go('san)	(*)	*		*!*

Tableau (45) shows how the analysis selects the correct result in forms containing long vowels, [maa.'suu.ra].

(45)

/maasuura/	HEAVYHEAD $>^{HEAD(FT)}$		HEADS-R	HEADS-L
☞ a. maa('suu)ra			(*)	*
b. ('maa)suu.ra		*!	(**)	
c. (ˌmaa)('suu)ra		*!	(***)	*
d. maa.suu('ra)	(*)	*!		

In response to Crowhurst and Hewitt, Balari *et al.* (2000) argue essentially that Crowhurst and Hewitt's interpretation of implication should not be allowed. These authors maintain that if complex constraints are to be derived using logical connectives, they should be connectives with classical Boolean semantics. Like Crowhurst and Hewitt, Balari *et al.*'s analysis of Dongolese uses HEADS-L and HEADS-R, but they replace HEAVYHEAD with a constraint LIGHTHEAD in (46a), which requires stress to fall on a *monomoraic* syllable, and they propose the (classically evaluated) implication in (46b).[9] The tableaux for ['mu.go.san] and [maa.'suu.ra] under BVM's analysis appear in (47).[10]

(46) a. LIGHTHEAD
The head σ of a foot is monomoraic.

b. LIGHTHEAD → HEADS-L
The implication is violated only if LIGHTHEAD is satisfied and HEADS-L is not.

[9] Balari *et al.* use the symbol "→" in the same way as ">" in Boolean logic. Balari *et al.*'s take on conflicting directionality is very close to that of Zoll (1997). Zoll proposes the constraint below to account for stress in Selkup. Note that Zoll's constraint is an unacknowledged complex constraint, which, when unpacked, has the semantics of Balari *et al.*'s implicational constraint.

(i) ALIGN-L('σμ, PWd): A light stressed syllable should be word-initial.

[10] Balari *et al.* differ from CH in assuming that HEADS-L and HEADS-R restricts the foot, not just the stressed syllable. (47) is taken from their work and therefore reflects this difference.

(47)

/mugosan/	LightHead	→	Heads-L	Heads-R
☞ a. ('mu.go)san				*
b. mu('go.san)		*!	(*)	
c. mu.go('san)		*!	(**)	

/maasuura/	LightHead	→	Heads-L	Heads-R
☞ a. maa('suu)ra	(*)		(*)	*
b. ('maa)suu.ra	(*)			**!
c. (ˌmaa)('suu)ra	(*)		(*)	**!*
d. maa.suu('ra)		*!	(**)	

In the end, it turns out that neither version of implication is strictly necessary to account for standard effects of conflicting directionality in stress. Conflicting directionality in stress, as in Dongolese, can be analyzed using the local conjunction of the constraints *Pk$_{\text{FT}}$/σμ and InitialStress in (48a) and (48b) (taking the domain of the conjunction to be the prosodic word).

(48) a. *Pk$_{\text{FT}}$/σμ
Avoid stress on monomoraic syllables.

b. InitialStress
Align-L(Head(PrWd), PrWd)
The head of the PrWd, syllable with main stress, occurs initially in its prosodic word.

c. *Pk$_{\text{FT}}$/σμ &$_{\text{PRWD}}$ InitialStress

Tableau (49) shows that the hierarchy *Pk$_{\text{FT}}$/σμ &$_{\text{PRWD}}$ InitialStress >> *Pk$_{\text{FT}}$/σμ >> Heads-R correctly accounts for Dongolese.

(49)

/mugosan/	*Pk$_{\text{FT}}$/σμ&$_{\text{PRWD}}$InitialStress	*Pk$_{\text{FT}}$/σμ	Heads-R
☞ a. ('mu.go)san		*	**
b. mu('go.san)	*!	*	*
c. mu.go('san)	*!	*	

/maasuura/	*Pk$_{\text{FT}}$/σμ&$_{\text{PRWD}}$InitialStress	*Pk$_{\text{FT}}$/σμ	Heads-R
☞ a. maa('suu)ra			*
b. ('maa)suu.ra			**!
c. maa.suu('ra)	*!	*	

In addition to Balari *et al.*, other arguments in favor of constraints derived using a connective with the semantics of Boolean material implication are presented in Archangeli *et al.* (1998) and Łubowicz (2005).

4 Different perspectives on the domain of conjunction

One distinguishing feature of work using various forms of conjunction has been researchers' different assumptions concerning what can constitute D, the domain of a conjunction. As the definition in (3) makes clear, Itô and Mester's (2003b) view is that D must be instantiated by a member of the set of phonological or morphological constituents. Smolensky's use of D is similar, but perhaps more inclusive in his use of the feature domain – a "maximal contiguous span of φ-bearers with a common value [±φ]" (see again (13)). We can say that for both Itô and Mester as well as Smolensky, the domain of a local conjunction is identified with a designated node in the phonological or morphological structure, and all material associated with this node (the association possibly being mediated by other intervening nodes). One difference between these authors seems to be that for Smolensky, but perhaps not Itô and Mester (2003b), any phonological feature can be the node that determines a domain.

A different perspective on D is found in Hewitt and Crowhurst (1996), Crowhurst and Hewitt (1997), Downing (1998, 2000), Baković (2000), and Łubowicz (2002, 2005), all of whom argue in one way or another that constraints can be conjoined only when they have the same locus of violation. As noted earlier, Hewitt and Crowhurst (1996), Crowhurst and Hewitt (1997), and Downing (1998, 2000) argue that b-conjunction should be limited to constraints that share an argument, which serves as the locus of violation of the conjoined constraint. In Crowhurst and Hewitt's analysis of Diyari reviewed earlier, for example, the locus of violation of the constraints MorphemeFt-L (MFL) and MorphemeFt-R (MFR) is the morpheme, and the conjunction of these constraints produces a complex constraint, MFL $\wedge^{MORPH}$ MFR, which has the same locus of violation.

An extended formal discussion of the *locus of violation* as a principled restriction on the domain of local conjunctions of markedness and faithfulness constraints is developed in Łubowicz (2005). Her definitions of restricted local conjunction and of the locus for local conjunction are shown in (50).

(50) a. *Restricted local conjunction*
$C = C_1 \& C_2$ is violated iff $LOC_{C1} \cap LOC_{C2} \neq Ø$.
b. *Locus for local conjunction*
The locus for local conjunction is the intersection of the sets LOC_{C1} and LOC_{C2}.

Formally, (50a) states that the loci of violation for the conjuncts of a local conjunction must intersect and that this intersection may not be null; (50b) states that the locus of violation for the local conjunction is the intersection of sets defined in (50a).

In a similar vein, Baković (2000) proposes that two constraints can be conjoined only if they are *co-relevant*, meaning that the definition of each conjunct specifies a particular feature also mentioned by the other conjunct. The co-relevance restriction plays an important role in Baković's use of M&F (markedness and faithfulness) conjunctions to account for "assimilation to the unmarked" phenomena, as discussed earlier. In his words, "the net effect of a co-relevant local conjunction of markedness and faithfulness is to specifically prohibit the unfaithful introduction of a marked segment" (Baković 2000: 7). Baković notes that the co-relevance restriction (or Łubowicz's more precisely defined restriction based on locus of violation) provides an answer to Itô and Mester (1998), who take the position that M&F conjunctions should not be allowed because they lead to undesirable results. Positioning the conjunction NoCoda & Ident[voice] above the fragment *[+voice] >> Ident[voice], for example, would falsely predict a language in which obstruents are voiced only in syllable codas. Baković notes that conjuncts used to illustrate Itô and Mester's point are not co-relevant, and in fact uses the example to reinforce his claim that *only* co-relevant constraints are conjoinable.

5 Taming the beast: Curbing the expressive power of constraint conjunction

Proposals for combining simpler constraints into more complex ones using connectives other than ">>" have not been universally welcomed (see for example Orgun and Sprouse 1999; Parker 2001; Idsardi 2006; Iverson and Salmons 2007; Zhang 2007). The issue of restricting the power of a formal theory that permits the derivation of complex constraints from simpler ones has generated concern, and inspired various proposals aimed at restrictiveness. The position of some scholars that restrictions on conjunction should be encoded in restrictions on the domain of conjunction was discussed in the last section. Other ways in which the issue of restrictiveness has played into arguments for and against constraint conjunction are reviewed below.

5.1 *Chain shifts: An argument against material implication*

Wolf (2007) calls for restricting the set of connectives that can be used to combine constraints. He asserts that local conjunction and strict domination exhaust the ways in which constraints may interact. In particular, he argues that complex constraints with the semantics of material implication must *not* be allowed, because admitting such constraints would have the consequence of allowing OT to model synchronic circular chain shifts. A circular chain shift takes the form A → B → A (where B can be expanded to accommodate additional links in the chain), in which a phoneme /A/ shifts to sound [B], and phoneme /B/, also present in the language, shifts to [A] (see also CHAPTER 73: CHAIN SHIFTS). The existence of a circular chain shift in which all links occur synchronically would present a problem for the OT doctrine of *harmonic ascent* (Prince and Smolensky 1993; McCarthy 2002). This is because a change A → B could only occur if B is less marked than A (and this would require a particular set of markedness constraints to outrank a particular set of faithfulness ones). But then, if B is less marked than A, there would be no reason for the change B → A (and the ranking that accounted for A → B could not account for the second change). Moreton (1999) provides a formal proof showing that an OT grammar that admits only faithfulness and markedness constraints is incapable of modeling circular chain shifts. As it turns out, there are no convincing synchronic examples of a circular chain shift conditioned by purely phonological factors, although there do appear to be examples triggered by morphological conditions (Anderson and Brown 1973). Diachronic examples of phonologically conditioned circular chain shifts are quite common (e.g. the flip-flop /i/ → /i/ → /ɨ/ in Siriono: Crowhurst 2000; the Germanic *Kreislauf*: Iverson and Salmons 2008), but these might be best handled by an analysis that assumes changes in constraint rankings at various stages of a language's development, along the lines proposed in Holt (2003).

5.2 *Other proposals for promoting restrictiveness*

Several other proposals aimed at promoting restrictiveness should be mentioned.

Some researchers have proposed restrictions (beyond the locus of violation) on the kinds of constraints that can combine. As noted earlier, Itô and Mester (1998) argued against the conjunction of markedness with faithfulness constraints. Fukazawa and Miglio (1998) proposed that conjunction should be limited to constraints within the same family, which would have much the same effect. These proposals have been countered by persuasive demonstrations of the benefits of conjoining faithfulness and markedness constraints (see §2.5).

A commonly accepted restriction has been to allow only complex constraints whose parts are independently necessary constraints. To do otherwise would permit unnatural results. Smolensky (2006) provides the example of a rule spreading [+ATR] from a [−high] vowel (e.g. [o . . . e] → [ɔ . . . ɛ]). This pattern is exactly what we don't find in ATR harmony, because [−ATR] non-high vowels are marked, and only unmarked feature bearers propagate their features (Archangeli and Pulleyblank 1994). Smolensky (2006) notes that ranking the conjunction *[+ATR, +hi] & HD-L[+ATR] above F[ATR] (the constraint requiring a feature domain to be faithful) could produce this result. However, the conjunction *[+ATR, +hi] & HD-L[+ATR] could simply never exist; one of its conjuncts *[+ATR, +hi] is not a member of Con, since from a markedness perspective, [+ATR] favors high vowels.

However pernicious, conjunctive relations of this type have in fact been proposed in the literature, sometimes to account for very real phenomena. An example would be the implication in (46b), used by Balari *et al.* (2000) to account for directional stress effects. Their constraint LightHead, which *demands* that stressed syllables be light, has no independent motivation – it is in fact an *anti*-harmonic constraint (since stress on light syllables is less harmonic than stress on heavy syllables). If LightHead is not a plausible candidate for membership in Con, then it would seem reasonable to conclude that Balari *et al.*'s conjunction LightHead → Heads-L is not an admissible complex constraint. Another example would be the constraint *Lapse in (51), employed by Elenbaas and Kager (1999: 282), who are proponents of the view that only independently motivated constraints may be combined.

(51) *Lapse

Every weak beat must be adjacent to a strong beat or the word edge.

Elenbaas and Kager note the disjunction in the requirement imposed by (51), but they point out that one of the disjuncts has no justification as a constraint in its own right. The first requirement, "every weak beat must be adjacent to a strong beat," is a cross-linguistically common restriction. However, Elenbaas and Kager note that there is no clear and independent motivation for the second requirement, that "every weak beat must be adjacent to the word edge." For this reason, they conclude that *Lapse should not be treated as a complex constraint. However, whether acknowledged or not, *Lapse is a form of disjunction (local conjunction), and a proper definition of the constraint would have to take account (more on this below). *Lapse is more problematic than the previous example, because the need for *Lapse is well established: as we showed in the last section, the Balari *et al.* implication is unnecessary; alternatives are readily available.

6 Concluding remarks

Proponents of allowing devices for combining constraints have argued that they further the goal of deriving empirical generalizations about markedness from irreducible principles (Itô and Mester 2003b; Smolensky 2006) and (a related point) that they promote greater precision and formal explicitness. Resistance to accepting complex constraints has been due largely to the observation that despite the advantages claimed, devices for combining constraints greatly increase the expressive power of the formal architecture, especially in regard to the constraint inventory Con. On the other hand, Itô and Mester (2003b: 22) note that in fact, the use of connectives keeps in check the proliferation of inexplicit constraints that duplicate aspects of others without a formal account of their components and the relations among them. (That is, since the need for the constraints in (6a) and (6b) is uncontroversial, and if we can combine them as in (4c), then we don't need the coda condition in (5).)

In closing, I will put forward a consideration that may seem provocative, but which ought to be taken seriously. Smolensky (2006) points out that the statement normally given for a constraint such as (51) (of which there are many examples in the literature) may pass as a translation of the definition of a constraint into English (or pseudo-formal language blended with English), but it is not a proper definition of that constraint (i.e. stated in the precise language of formal logic). A precise definition of (51) in formal language must necessarily take account of the disjunction it contains; it cannot do otherwise. Whether complex constraints *should* be allowed is a non-issue. In fact, they *have been* used without controversy from the inception of OT, camouflaged as informal descriptions that have been accepted as definitions of constraints. The issue that deserves consideration is whether the position that complex constraints should be excluded can be maintained in light of its implications – most seriously, the implication that an enormous number of generally accepted constraints ought to be expelled from Con, or at least rethought. On the issue of overgeneration, we may perhaps leave the last word, for now, to Itô and Mester:

> The broadly defined outline of local conjunction theory . . . admits a huge number of conjoined constraints, only a small subset of which will turn out to play a role in grammar, and many of which are unwanted. In our view, the task of distinguishing between "reasonable," "plausible," "expected" conjunctions and "unreasonable," "implausible," "unexpected" conjunctions cannot be relegated to the syntax of conjunction, which simply provides a system for expressing derived constraints. The distinction is an issue of phonological substance and phonetic groundedness, not one of formalization (Itô and Mester 2003b: 24).

ACKNOWLEDGMENTS

The author appreciates useful comments received from the editors, Marc van Oostendorp and Keren Rice, and from Scott Myers, Joe Salmons, and two anonyomous reviewers. Their input has helped to make this a better chapter.

REFERENCES

Alderete, John. 1999. Faithfulness to prosodic heads. In Ben Hermans & Marc van Oostendorp (eds.) *The derivational residue in phonology*, 29–50. Amsterdam & Philadelphia: John Benjamins.

Alderete, John. 2004. Dissimilation as local conjunction. In John J. McCarthy (ed.) *Optimality Theory in phonology: A reader*, 394–406. Malden, MA & Oxford: Blackwell.

Anderson, Stephen R. & Wayles Browne. 1973. On keeping exchange rules in Czech. *Papers in Linguistics* 6. 445–482.

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Archangeli, Diana, Laura Moll & Kazutoshi Ohno. 1998. Why not *NÇ? *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 34. 1–26.

Armbruster, Charles Hubert. 1960. *Dongolese Nubian: A grammar*. Cambridge: Cambridge University Press.

Austin, Peter. 1981. *A grammar of Diyari, South Australia*. Cambridge: Cambridge University Press.

Baertsch, Karen. 1998. Onset sonority distance constraints through local conjunction. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 34. 1–16.

Baertsch, Karen & Stuart Davis. 2003. The split margin approach to syllable structure. *ZAS Papers in Linguistics* 32. 1–14.

Baković, Eric. 2000. Harmony, dominance, and control. Ph.D. dissertation, Rutgers University.

Balari, Sergio, Rafael Marín & Teresa Vallverdu. 2000. Implicational constraints, defaults, and markedness. Unpublished ms., Universitat Autònoma de Barcelona (ROA-396).

Bavin Woock, Edith & Michael Noonan. 1979. Vowel harmony in Lango. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 15. 20–29.

Beckman, Jill N. 2003. The case for local conjunction: Evidence from Fyem. *Proceedings of the West Coast Conference on Formal Linguistics* 22. 56–69.

Cassimjee, Farida & Charles W. Kisseberth. 1998. Optimal Domains Theory and Bantu tonology: A case study from Isixhosa and Shingazidja. In Larry M. Hyman & Charles W. Kisseberth (eds.) *Theoretical aspects of Bantu tone*, 33–132. Stanford: CSLI.

Cole, Jennifer & Charles W. Kisseberth. 1994. Nasal harmony in Optimal Domains Theory. Unpublished ms., University of Illinois.

Crowhurst, Megan J. 1994. Prosodic alignment and misalignment in Diyari, Dyirbal, and Gooniyandi: An optimizing approach. *Proceedings of the West Coast Conference on Formal Linguistics* 13. 16–31.

Crowhurst, Megan J. 2000. A flip-flop in Sirionó (Tupian): The mutual exchange of /i ɨ/. *International Journal of American Linguistics* 66. 57–75.

Crowhurst, Megan J. 2001. Coda conditions and *um* infixation in Toba Batak. *Lingua* 111. 561–590.

Crowhurst, Megan J. & Mark Hewitt. 1997. Boolean operations and constraint interactions in Optimality Theory. Unpublished ms., University of North Carolina & Brandeis University (ROA-229).

Downing, Laura J. 1998. On the prosodic misalignment of onsetless syllables. *Natural Language and Linguistic Theory* 16. 1–52.

Downing, Laura J. 2000. Morphological and prosodic constraints on Kinande verbal reduplication. *Phonology* 17. 1–38.

Downing, Laura J. 2001. Liquid spirantisation in Jita. *Malilime: Malawian Journal of Linguistics* 2. 1–27.

Elenbaas, Nine & René Kager. 1999. Ternary rhythm and the lapse constraint. *Phonology* 16. 273–329.

Fukazawa, Haruka. 1999. Theoretical implications of OCP effects on features in Optimality Theory. Ph.D. dissertation, University of Maryland at College Park (ROA-307).

Fukazawa, Haruka. 2001. Local conjunction and extending sympathy theory: OCP effects in Yucatec Maya. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 231–260. Cambridge: Cambridge University Press.

Fukazawa, Haruka & Viola Miglio. 1998. Restricting conjunction to constraint families. *Proceedings of the Western Conference on Linguistics* 9. 102–117.

Hall, T. A. 2004. German glide formation and constraint conjunction. Unpublished ms., University of Indiana.

Hall, T. A. 2007. German glide formation and its theoretical consequences. *The Linguistic Review* 24. 1–31.

Hewitt, Mark & Megan J. Crowhurst. 1996. Conjunctive constraints and templates in Optimality Theory. *Papers from the Annual Meeting of the North East Linguistic Society* 26. 101–116.

Holt, D. Eric (ed.) 2003. *Optimality Theory and language change*. Dordrecht: Kluwer.

Idsardi, William J. 2006. A simple proof that Optimality Theory is computationally intractable. *Linguistic Inquiry* 37. 271–275.

Itô, Junko & Armin Mester. 1996. Rendaku 1: Constraint conjunction and the OCP. Paper presented at the Kobe Phonology Forum.

Itô, Junko & Armin Mester. 1998. Markedness and word structure: OCP effects in Japanese. Unpublished ms., University of California, Santa Cruz (ROA-255).

Itô, Junko & Armin Mester. 2003a. On the sources of opacity in OT: Coda processes in German. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*, 271–303. Cambridge: Cambridge University Press.

Itô, Junko & Armin Mester. 2003b. *Japanese morphophonemics: Markedness and word structure*. Cambridge, MA: MIT Press.

Iverson, Gregory K. & Joseph C. Salmons. 2007. Domains and directionality in the evolution of German final fortition. *Phonology* 24. 121–145.

Iverson, Gregory K. & Joseph Salmons. 2008. Germanic aspiration: Phonetic enhancement and language contact. *Sprachwissenschaft* 33. 257–278.

Jacobs, Haike. 2004. Rhythmic vowel deletion in OT: Syncope in Latin. *Probus* 16. 63–89.

Kager, René. 1994. Generalized alignment and morphological parsing. Unpublished ms., University of Utrecht.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kawahara, Shigeto. 2003. On a certain type of hiatus resolution in Japanese. *Phonological Studies* 6. 11–20.

Kiparsky, Paul. 1973. "Elsewhere" in phonology. In Stephen R. Anderson & Paul Kiparsky (eds.) *A Festschrift for Morris Halle*, 93–106. New York: Holt, Rinehart & Winston.

Kirchner, Robert. 1993. Turkish vowel harmony and disharmony: An optimality theoretic analysis. Unpublished ms., University of California, Los Angeles (ROA-4).

Kirchner, Robert. 1996. Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry* 27. 341–350.

Levelt, Clara C. & Ruben van de Vijver. 1998. Syllable types in cross-linguistic and developmental grammars. Paper presented at the 3rd Biannual Utrecht Phonology Workshop (ROA-265).

Levelt, Clara C., Niels O. Schiller & Willem J. Levelt. 1999. A developmental grammar for syllable structure in the production of child language. *Brain and Language* 68. 291–299.

Lin, Hui-Shan. 2000. Layered OCP, unparsed preposition, and local constraint conjunction in Mandarin tone sandhi. Paper presented at the 33rd International Conference on Sino-Tibetan Languages and Linguistics, Ramkhamhaeng University, Thailand.

Łubowicz, Anna. 2002. Derived environment effects in Optimality Theory. *Lingua* 112. 243–280.

Łubowicz, Anna. 2005. Locality of conjunction. *Proceedings of the West Coast Conference on Formal Linguistics* 24. 254–262.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

Moreton, Elliott. 1999. Non-computable functions in Optimality Theory. Unpublished ms., University of Massachusetts, Amherst (ROA-364).

Moreton, Elliott & Paul Smolensky. 2002. Typological consequences of local constraint conjunction. *Proceedings of the West Coast Conference on Formal Linguistics* 21. 306–319.

Morris, Richard. 2002. Coda obstruents and local constraint conjunction in north-central Peninsular Spanish. In Teresa Satterfield, Christina Tortora & Diana Cresti (eds.) *Current issues in Romance languages: Selected papers from the 29th Linguistic Symposium on Romance Languages*, 207–224. Amsterdam & Philadelphia: John Benjamins.

Noonan, Michael. 1992. *A grammar of Lango*. Berlin & New York: Mouton de Gruyter.

Okello, Jenny. 1975. Some phonological and morphological processes in Lango. Ph.D. dissertation, Indiana University.

Orgun, Cemil Orhan & Ronald L. Sprouse. 1999. From MParse to Control: Deriving ungrammaticality. *Phonology* 16. 191–224.

Parker, Steve. 2001. Non-optimal onsets in Chamicuro: An inventory maximized in coda position. *Phonology* 18. 361–386.

Poser, William J. 1989. The metrical foot in Diyari. *Phonology* 6. 117–148.

Prince, Alan. 2001. Invariance under re-ranking. Paper presented at the 20th West Coast Conference on Formal Linguistics, University of Southern California.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Reynolds, William. 1994. Variation and phonological theory. Ph.D. dissertation, University of Pennsylvania.

Smolensky, Paul. 1993. Harmony, markedness, and phonological activity. Paper presented at the Rutgers Optimality Workshop 1, Rutgers University (ROA-87).

Smolensky, Paul. 1995. On the structure of the constraint component *Con* of UG (ROA-86).

Smolensky, Paul. 1997. Constraint interaction in generative grammar II: Local conjunction, or random rules in universal grammar. Paper presented at the Hopkins Optimality Theory Workshop/University of Maryland Mayfest 1997.

Smolensky, Paul. 2006. Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature domain markedness. In Paul Smolensky & Géraldine Legendre (eds.) *The harmonic mind: From neural computation to optimality-theoretic grammar*, vol. 2: *Linguistic and philosophical implications*, 27–160. Cambridge, MA: MIT Press.

Spaelti, Philip. 1997. Dimensions of variation in multi-pattern reduplication. Ph.D. dissertation, University of California, Santa Cruz.

Walker, Rachel. 2000. Nasal reduplication in Mbe affixation. *Phonology* 17. 65–115.

Wolf, Matthew. 2007. What constraint connectives should be permitted in OT? *University of Massachusetts Occasional Papers in Linguistics* 36. 151–179.

Zhang, Jie. 2007. Constraint weighting and constraint domination: A formal comparison. *Phonology* 24. 433–459.

Zoll, Cheryl. 1997. Conflicting directionality. *Phonology* 14. 263–286.

Zoll, Cheryl. 1998. Positional asymmetries and licensing. Unpublished ms., MIT (ROA-282).

63 Markedness and Faithfulness Constraints

PAUL DE LACY

1 Introduction

Objects and mechanisms called "constraints" have featured in many theories of the phonological and syntactic modules. However, the explicit bifurcation into "markedness and faithfulness" constraints is specifically found in Optimality Theory (OT; Prince and Smolensky 1993) and its developments (especially McCarthy and Prince 1999), as well as in theories based on OT (Stochastic OT: Boersma and Hayes 2001; Targeted Constraint Theory: Wilson 2001; OT with Candidate Chains: McCarthy 2007; Stratal OT: Bermúdez-Otero, forthcoming; Kiparsky, forthcoming). So this chapter focuses on the things called "constraints" in OT (specifically the "classical OT" of Prince and Smolensky 1993 and McCarthy and Prince 1999). In particular, it focuses on OT constraints in the phonological module; there are also OT theories of the syntactic module and OT theories of morphology – they will not be discussed here.

This chapter's aim is to examine the basic syntax and semantics of constraints. On the syntax side: What is the form of constraints? What is the "constraint construction language"? On the semantics side: How are constraints "interpreted" – i.e. how are constraints used to assess a candidate's violation marks?

This chapter focuses on the basics of constraints, so it does not aspire to identify every constraint theory or list every constraint and constraint generator (§4.2) that has been proposed; for that, see the ongoing *ConCat* project.[1]

An OT constraint is commonly treated as a function that takes a candidate and returns "violation marks" (see Prince and Smolensky 1993). Violation marks are discrete elements; they are usually written as a string of asterisks with one asterisk per unique element (but violation marks in their formal implementation are not necessarily a string). For example, a constraint *DORSAL returns one violation mark for each instance of the representational element [dorsal] in an output representation (constraint names are usually written in small capitals). So for a candidate that includes an output representation [pakak], *DORSAL returns **, because there are two [dorsal] features in the form (one for each [k]). "Candidates" are sets of

[1] Available at http://concat.wiki.xs4all.nl.

forms including at least an output and input representation, but probably many other related representations as well as relations between them (§3.3; Prince and Smolensky 1993).

Constraints fit into the overall phonological system thus (following Prince and Smolensky 2004). An input is drawn from the lexicon; the input consists of phonological material with morphological and syntactic annotation/structure. A generation mechanism (GEN) produces many (perhaps an infinite number of) candidates. One or more candidates is selected from the array; the selection process (EVAL) involves constraints generating violation marks for candidates and an algorithm that uses the violation marks and other factors to determine the winning candidate(s). EVAL refers to "ranking" – a total order on constraints. Ranking does not influence how violation marks are calculated; however, ranking is crucial in discovering the winning candidate. The phonetic module then takes (one of) the winner(s) and realizes it (i.e. converts it into articulatory movements that produce speech sound).

In short, constraints are just one part of many mechanisms that work together to determine the winning output representation. Constraints do not determine winners on their own.

The term "markedness and faithfulness" as applied to constraints was coined in Prince and Smolensky (1993: §1.4). "Markedness constraints" return violation marks based solely on the form of the output representation. *Dorsal above is a markedness constraint. Unfortunately, the term "markedness" can cause confusion because it seems to imply an inherent connection to theories of markedness (see CHAPTER 4: MARKEDNESS). However, theories of markedness are expressed in OT via both markedness and faithfulness constraints (e.g. de Lacy 2006). The term "output constraint" is therefore less confusing than "markedness constraint," and I will use it here. However, the phrase "markedness constraint" is in such widespread use that I fear "output constraint" will never catch on, in spite of my efforts in this chapter (to add to the confusion there are constraints called "output–output constraints," which are actually faithfulness constraints; see §3.3).

As originally used, "faithfulness constraints" are those that return violation marks based on comparison of the output representation with the input (Prince and Smolensky 1993: §1.2; though strictly speaking the output *includes* the input, as discussed in §3.1). Later work, especially McCarthy and Prince (1999), broadens the term to include any constraint that assigned violations by comparing any pair of inter- or intra-representational forms (e.g. the base of reduplication and the reduplicant (McCarthy and Prince 1999); the derivational base and the output (Benua 1997); a designated form and the output (McCarthy 1999)). The majority of work in OT now uses McCarthy and Prince's Correspondence Theory, so in these cases it is accurate to refer to "correspondence constraints" – i.e. those that use correspondence relations in their calculation of violation marks. However, non-correspondence faithfulness constraints exist in some versions of OT (e.g. containment theories –§3.1), so "faithfulness constraints" is still a usefully broad term.

This chapter focuses on a few important issues about constraints. §2 discusses constraint form in output constraints: What are constraints made of, and how do they return violation marks? §3 deals with faithfulness and correspondence constraints. §4 discusses the source of constraints – whether they are innate and how/whether they relate to external sources.

1.1 "Constraint" in other theories

The term "constraint" is used in many different ways in many different theories. In some rule-based theories there are objects called "constraints" or "filters" that – if their conditions are met – doom the derivation or block rules from applying. If an input I undergoes a series of rules to create a representation φ and there is a filter *φ, the derivation is doomed (i.e. input I has no corresponding output). See Chomsky and Lasnik (1977) for early examples in syntax, and Itô (1986) for conditions on syllable structure. The filter/condition concept does not have a direct analogue in OT – OT constraints assign violations; EVAL is the source of (relative) doom for candidates.

Occasionally, "constraint" is used to refer to the side-effects of conditions on representational primitives and to restrictions on the algorithm that generates output candidates (GEN). Obviously, output representations can only be constructed out of objects and relations that are available (i.e. prosodic nodes, features, planes, tiers; precedence, dominance). For example, there is no candidate in which a node can both precede and follow another node, because the phonological precedence relation is asymmetric (i.e. if *a* and *b* are on the same tier and *a* < *b* (i.e. *a* precedes *b*) and *b* < *a* then *a* = *b*) (see CHAPTER 34: PRECEDENCE RELATIONS IN PHONOLOGY for a discussion of precedence). One could informally call the asymmetry of phonological precedence a "constraint," but it is not an OT constraint.

2 Output constraints

A constraint takes a candidate and returns violation marks.[2] For example, the constraint *DORSAL returns one violation mark for [ka], two for [kax], and so on. A constraint's violation assignment can be described in informal terms: e.g. "*k returns a violation mark for each [dorsal] segment." This informal description is useful, but far from being a formal definition. A formal definition of a constraint must be couched in a Constraint Definition Language (CDL). A comprehensive CDL specifies representational primitives and relations and restrictions on their combination in constraints.

The same distinction can be made for rule-based theories like Chomsky and Halle (1968; *SPE*). Suppose we observe a rule R that takes an input representation /ak/ and converts it to the representation [aʔ]. R could be described as "change /k/ into a [ʔ] word-finally." However, R must be *defined* in terms of a Rule Definition Language (RDL); an RDL is the elements and relations that can be used to construct a rule and limits on their combination. Most of Chomsky and Halle (1968) is devoted to developing such an RDL; R is defined as /k/ → [ʔ] / __ [−seg, −FB, +WB] (the rightmost cluster of features is a word boundary).

[2] It is common to see comments like "constraints impose a partial order on the candidate set" (Samek-Lodovici and Prince 1999: 9), "this constraint dooms the candidate," and so on. These comments are meant as a quick way of describing the complex process of determining order among candidates; the process involves constraints, VR (§2.2), ranking, and EVAL's mechanisms; constraints are merely one part of the process of establishing the winning candidate.

Some work in OT uses informal descriptions to talk about constraints. Formal objects (representational elements like prosodic nodes, features, etc.) are often mentioned in the informal descriptions, but the constraints are nevertheless not defined in terms of an overarching CDL. There have been attempts to develop a comprehensive CDL (Eisner 1997; Potts and Pullum 2002), but most work has either focused on particular groups of constraints, or treated constraints as "black boxes."

To explain, it is possible to fruitfully investigate some (perhaps many) aspects of OT theories without knowing the precise definition of constraints, but only knowing which violation marks constraints assign in which situations. After all, the winning candidate is not directly determined by constraints, but by their violation marks. So, if the violation marks are known then the winner can be determined – the exact means by which the violation marks came to be assigned is often not crucial. For this reason I believe it is fair to say that there has been less focus on developing a CDL in OT theories than on developing RDLs in rule-based theories like Chomsky and Halle (1968).

Even so, there have been detailed proposals of CDLs for groups of OT constraints (see §4 below), and some proposals about aspects of the general CDL. In my own (joint) work, for example, Bye and de Lacy (2000) propose restrictions on how constraints can refer to constituent edges; de Lacy (2006) proposes that constraints cannot include both prosodic nodes and segmental features in their definitions.

An explicit CDL is both useful and ultimately essential to a complete optimality theory. A CDL can tell us which constraint formulations are valid, and thus set a bound on which constraints can and cannot exist.

The following subsections will discuss a CDL. There is a strong uniformity in constraint descriptions and definitions that suggests broad agreement about certain aspects of the CDL. For expository reasons, I will start with the CDL for output constraints. As a word of warning, due to space limitations I will discuss only a few CDLs, and focus on the basic components of just one. The CDL discussed below deals with a broad set of output constraints that I believe every phonologist would accept as possible constraints. I will not attempt to comprehensively discuss all extant CDLs or aspects of CDLs, but instead focus on basic CDL properties.

2.1 *Output constraints: Representation*

There is not a lot of explicit discussion about how constraints work in OT. It seems to me that the majority of work in OT treats constraints as functions from candidates to violation marks. Output constraints inspect the output representation in a candidate, and return a string of violation marks. So *Dorsal returns one violation for candidates with an output representation [ka], two for [kag], three for [gaxikan], and so on, leading to the description: "Assign one violation for each dorsal segment."

We are seeking a CDL in which *Dorsal can be formulated. One issue to address is the CDL's representational primitives. For example, *Dorsal might be cast in terms of a representational theory in which velar consonants are [+back, +high] (Chomsky and Halle 1968: 303), or one in which a Place node dominates a dorsal node which dominates [+back] and [+high] terminal nodes, or one in which an oral cavity node dominates a C-place node which dominates a dorsal terminal node (see Hall 2007 for an overview of feature theories). There are many extant representational theories, but for the purposes of this chapter I will adopt

Clements and Hume's model (there is no widespread consensus on which representational theory is correct, though; even less now than in 1995, I suspect; see also CHAPTER 27: THE ORGANIZATION OF FEATURES).

The CDL also must specify how representational elements can be combined in constraints. For example, a relatively lax CDL could allow several different versions of *DORSAL, as in (1). I use • for "root node" – the lowest node that dominates all segmental features. The symbol ↓ stands for the immediate dominance relation: a↓b means "a immediately dominates b" (i.e. a dominates b and there is no c such that a dominates c and c dominates b); a↓b↓c means "a immediately dominates b and b immediately dominates c." Dominance is an asymmetric transitive relation that holds between nodes on different autosegmental tiers. The descriptions are given in (1); the violation marks the constraint assigns are shown for [k], [kː] (assuming a one-root geminate theory), and [ŋk] (assuming obligatory feature sharing for adjacent elements; Schein and Steriade 1986).

(1) *DORSAL *versions*

Return a violation for ...	[k]	[kː]	[ŋk]
a. each distinct root node • s.t. •↓CPlace↓[dorsal]	*	*	**
b. each distinct [dorsal] feature	*	*	*
c. each distinct prosodic node *p* s.t. p↓•↓CPlace↓[dorsal]	*	**	**[3]

I do not know which of the constraints in (1) exist. Suppose it turns out that we need only (1a). The non-existence of (1b) and (1c) could be achieved by placing restrictions on the CDL such that all output constraints must refer to a root node in their definition and no constraint may mention both prosodic nodes and segmental features. Every extant theory of representation provides a CDL with a great deal of potentially expressive power. So, it is highly likely that any CDL theory will have to incorporate extensive limitations on permissible representations in constraints; it is probably too hopeful that all limits on constraints will be a side-effect of inherent limitations on representations (see §4).

The CDL must also specify how the representation is used to assess violation marks. For example, suppose a constraint mentions the structure [•↓CPlace↓[dorsal]]. How is this structure used to assess violation marks relative to some candidate? In the constraint description (1a), I assumed that the constraint searches the candidate's output representation and one violation mark is returned for each distinct structure that has the form [•↓CPlace↓[dorsal]]. However, could there be a constraint which returns one violation regardless of how many [•↓CPlace↓[dorsal]] structures there were? Such a constraint would return * for [ka], [kax], and [kaxga]. Let's turn to this issue now.

2.2 *Function or representation?*

I asserted without comment above that a constraint is a function: i.e. it takes a candidate as an input and returns violation marks. Conceiving of constraints as

[3] (1c) could return one violation if [ŋ] and [k] were both dominated by the same μ or σ node (e.g. as in [oiŋk]$_σ$).

independent functions opens up the possibility that different constraints could assign violation marks in very different ways.

For example, the ALIGN schema from McCarthy and Prince (1993) takes four arguments and assesses violation marks with respect to designated prosodic constituents. The violation marks from ALIGN(Ft, R; PrWd, R) are the sum of the number of syllables between the right edge of each foot and the right edge of the PrWd. So, the constraint returns nine violations for [(σσ)(σσ)(σσ)σ] (see McCarthy and Prince 1993: 15–16 for details; to understand this constraint see especially 1993: 10 and definitions 14–16). It is clear that the way in which this ALIGN constraint assesses violation marks is quite different from the way in which violations of *DORSAL are assessed (cf. McCarthy 2003).

If constraints are self-contained algorithms that return violation marks and the CDL is sufficiently powerful, we might see pairs of constraints that refer to the same representational structure but differ in how they calculate violation marks. For example, there could be a pair of constraints *∀DORSAL and *∃DORSAL, where *∀DORSAL returned as many violation marks as there are [dorsal] features in an output representation, but *∃DORSAL returned only one violation mark regardless of how many [dorsal] features there are in a form, as long as there is at least one (see Wolf 2007a for relevant discussion). The two constraints refer to the same structure – [•↓CPlace↓[dorsal]] – and differ only in terms of how that structure is used to assess violation marks from a candidate. A pair of constraints like this – i.e. that refer to the same representational structure but differ only in their quantification – would be strong evidence that each constraint is an independent algorithm that assigns violations (or at least that there are several groups of constraints that differ in how they assign violations).

However, my impression is that the constraints-as-functions approach is too powerful. The output constraints that have been proposed in phonological literature are often very similar: they essentially have the form *R, where R is a representation; one violation mark is assigned for each distinct occurrence of R in a candidate's output representation. *DORSAL is an example of such a constraint. The apparent uniformity in how constraints assess violations suggests that it is worthwhile considering an alternative theory of constraints in which constraints are not functions but solely representations.

In such an approach, there would be a single algorithm, Violation Assigner (VR). VR takes as its input an output constraint and a candidate and returns violation marks. VR works the same way for all constraints, thus imposing uniformity in how violation marks are assigned. So the constraint *DORSAL is really a representation [•↓CPlace↓[dorsal]]; *DORSAL itself does not assess violation marks.

There are many ways to formulate a VR algorithm that does the job described above. For example, one could take the set of all sub-representations of a candidate's output representation and compare each member of the set to a constraint representation; the number of violation marks returned for a particular constraint would be the number of subsets that were equivalent to the constraint's representation. I will instead discuss a somewhat more efficient algorithm that does a similar job.[4]

[4] See www.pauldelacy.net/VR for software which allows the user to see the VR below in action and try out various constraints and representations.

(2) *Violation Assigner (VR): Outline*

Inputs:
- a constraint C;
- a candidate that includes an output representation R;

Output:
- a set of violation marks (a set of unique identifiers) indexed to C and the candidate.

a. Take a node c in C.
b. For each node r in R:
If r is the same type and value as c,
then check whether r is connected to a structure equivalent to C.
– if it is, return a violation mark, and continue to the next r.
c. Add no other violation marks to the result.

For example, take a constraint *[+voice], which consists of one [voice] node with a value "+." Each node in the candidate's output representation is checked. If a node is a [voice] node and has the value "+," a violation mark is added to the result.

The VR might seem straightforward, but it has interesting complexities, particularly in the procedure that checks whether a node is "connected to a structure equivalent to [the constraint] C."

Take a more complex constraint – one that involves two nodes: e.g. $*\sigma_{\mu\mu}$, "Don't have bimoraic syllables." $*\sigma_{\mu\mu}$ has three nodes (σ, μ_1, μ_2) and three relations ($\sigma^{\downarrow}\mu_1$, $\sigma^{\downarrow}\mu_2$, $\mu_1 < \mu_2$). A node is selected from the constraint (it doesn't matter which one) – let's say σ in this case. The output representation is searched for σ nodes. When one is found, the next step is to check whether σ is connected to a structure equivalent to the constraint. The implementation of this checking procedure is that σ is checked to see if it is in any of the *relations* mentioned by the constraint: i.e. does the particular σ in the representation dominate two different μ nodes? If it does, then the μ nodes that are dominated by σ are checked to see whether their relations have equivalents in the constraint. After nodes and relations are found in the output representation that are equivalent to those in the constraint, a violation mark is returned.

The procedure that checks whether *n* is connected to a structure equivalent to C means that constraints cannot be unconnected. For example, suppose that there is a constraint $*\mu_1, \mu_2$ which is violated if a word contains two (not necessarily adjacent) moras; these moras are unconnected in this constraint – there is no precedence relation between them, nor is there a node that dominates them both. VR can evaluate such a constraint, but the constraint will never return any violation marks. VR checks *relations* between nodes: i.e. VR will search for a μ_1 node, and then check its relations. Since μ_1 has no connection to μ_2 via either precedence or dominance, VR will never find any structure in any R that is equivalent to the structure described by the constraint.

So the VR algorithm itself, through how it compares the constraint's structure to structures in the output representation, imposes a weak connectedness requirement on constraints. For a constraint C ever to return a violation mark, every node in C must be connected to every other node. Nodes *x* and *y* are "connected" here if it is possible to trace a direct route through precedence and dominance relations from *x* to *y*.[5]

The connectedness requirement that results from VR is *weak*, because a much stronger requirement could be imagined and implemented: i.e. in a constraint, every pair of nodes on the same tier must be in a precedence relation, and every pair of nodes on different tiers must be in a dominance relation. The difference between weak and strong connectedness can be seen in a less connected version of $*\sigma_{\mu\mu}$: one where $\sigma^{\downarrow}\mu_1$ and $\sigma^{\downarrow}\mu_2$, but there is no precedence relation between μ_1 and μ_2. Such a constraint would not be evaluated by a VR that imposes strong connectedness because it has two nodes on the same tier (μ_1, μ_2) that are not in a precedence relation. However, it is perfectly acceptable in the weak-connectedness VR described above, because every node is connected to every other node.

Weak connectedness allows constraints of the form $*[x \ldots x]_D$, where there are two nodes of type x within a particular domain D (e.g. Suzuki 1998; see also the discussion of local conjunction in §4.2).

The connectedness side-effect of VR is desirable – as far as I am aware, no one has proposed constraints that have completely unconnected elements.

The larger point here is that the nature of the algorithm that assigns violation marks is crucial in any theory of constraints. The algorithm not only determines how violation marks are assigned, but whether particular constraints will ever assign violation marks (i.e. it effectively puts restrictions on constraint form just as VR means that constraints must contain connected representations).

If constraints are representations and there is a single VR, there should be broad uniformity in the way that violation marks are assigned. For example, the constraint [$\bullet^{\downarrow}$CPlace$^{\downarrow}$[dorsal]] will assign violations for each occurrence of its representation in the candidate's output representation. In contrast, there is no way to formulate a constraint like *∃dorsal: if the constraint consists of the representation [$\bullet^{\downarrow}$CPlace$^{\downarrow}$[dorsal]], then the VR will return a violation for *each* occurrence of [$\bullet^{\downarrow}$CPlace$^{\downarrow}$[dorsal]]; it cannot be limited to assigning one violation regardless of the number of occurrences of [$\bullet^{\downarrow}$CPlace$^{\downarrow}$[dorsal]].

So the VR theory means that output constraints should all assign violations in fundamentally the same way, while the constraints-as-functions theory allows for significant differences. It is even possible that there is a middle ground: there could be several violation assignment algorithms, with VR being just one of them. With several VRs, we would expect to see uniformity in how violations are assigned, but only within particular groups of constraints.

Which view is correct? How much uniformity in violation mark assignment is there?

2.3 *Regularities and irregularities in violation assignment*

There is a great deal of regularity in the way that violation marks are assessed in output constraints. In fact, there is so much regularity that I am sure there would be no confusion among phonologists about how a constraint like *ʃr works (even though I concocted it just now): it would return a violation mark for each

[5] Let us define a transitive symmetric relation ⊕: $x \oplus y$ if $x < y$ or $x^{\downarrow}y$. Nodes x and y are connected if $x \oplus y$. For example, in the constraint $*\{\sigma^{\downarrow}\mu_1, \sigma^{\downarrow}\mu_2\}$, μ_1 and μ_2 are connected because $\sigma \oplus \mu_1$ and $\sigma \oplus \mu_2$, so $\mu_1 \oplus \sigma$ (by symmetry) and $\mu_1 \oplus \mu_2$. There is no need for a direct implementation of ⊕ in VR; connectedness follows as a side-effect of the procedure that matches the constraint to the representation.

[ʃ] immediately followed by [ɾ] in the output representation. There is an implicit "quantificational" regularity here: one violation mark is returned for *each* distinct [ʃɾ] (i.e. [ʃɾaʃɾi] returns two violations), not for just having some [ʃɾ] (where [ʃɾaʃɾi] would return just one violation).

As mentioned above, such "quantification" regularity is expressed straightforwardly in VR. The process of iterating through every node in the representation means that a violation will be returned *each time* a node is found that is part of a representation that matches the constraint's.

There are other widespread similarities in constraints. Many output constraints refer to tier-contiguous representations. For example, a constraint like *ŋk refers to *adjacent* root nodes; *ŋk returns one violation for [aŋko], but [aŋok] gets none (i.e. *ŋk is ŋ < k, where "<" is used here to mean "immediately precedes"). I have not yet seen a constraint *ŋ . . . % . . . k proposed which is violated whenever [ŋ] and [k] appear in the same output in any order (i.e. one violation for each of [aŋko], [aŋok], [a kŋo], [akoŋ], etc.) (the "%" notation is from Bach 1968).

The particular implementation of VR discussed above has intra-tier contiguity as a side-effect of the way that the algorithm works. When a node is encountered in the output representation, the node's relations (i.e. precedence and dominance) are checked against the constraint's. If the node is in a precedence or dominance relation with some other node, that other node's relations are then checked, and so on. This method of checking means that only nodes that are connected in the representation will be examined. So, for a constraint *ŋ . . . % . . . k, a node [ŋ] might be found in the representation, but [ŋ] does not precede any other node, so [k] will never be found by the VR procedure, and the constraint's representation will never be matched. (See CHAPTER 34: PRECEDENCE RELATIONS IN PHONOLOGY for more discussion of precedence.)

Some authors have proposed constraints with weakly contiguous representations to deal with processes like long-distance assimilation and dissimilation (e.g. Suzuki 1998; see also CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS). These constraints refer to segmental nodes that are not in a precedence relation, but are related by dominance (i.e. with the form $*[x \ldots \% \ldots y]_D$, where x and y are nodes dominated by constituent D. Such constraints always require or ban identical elements (not non-identical ones, as above), leading some recent work to recast such processes as involving local correspondence relations (e.g. Rose and Walker 2004) or to restrict the domain of such constraints so that the representations they refer to are mostly always tier-contiguous (e.g. Łubowicz 2005). Such a move might mean that VR should implement a strong contiguity requirement, where within a constraint every node on a particular tier is in a precedence relation with every other node on that tier.

There is a widely accepted constraint type that does not always easily fit with VR: "positive" constraints. A positive constraint penalizes a representation for *lacking* a particular property, while negative constraints like *DORSAL return violations for *having* a particular structure. A well-known constraint that is often positively formulated is ONSET: "Syllables must have onsets."[6] Another example is PLACE→DORSAL: "Every Place node must dominate a [dorsal] node." Because positive constraints do not mention a representation but rather the *lack* of a

[6] See other definitions of this constraint, including negative ones, at *Concat* http://concat.wiki.xs4all.nl/index.php?title=Onset.

representation, they cannot be fed into the VR. That is to say, PLACE→DORSAL does not mention the representation [•↓Cplace↓[dorsal]]. Instead, it requires that for every root node • there must be some member of the set of all sub-representations of a candidate that includes • and has the structure [•↓Cplace↓[dorsal]]. It is not hard to formulate a VR algorithm that can handle positive constraints (VR^{POS}), but such an algorithm is different from the one that handles negative constraints (VR^{NEG}).

Are both VR^{NEG} and VR^{POS} necessary? Many positively formulated constraints can be reformulated negatively, and many negative ones can be reformulated positively. However, the success of reformulation depends entirely on the particular CDL used. For example, within fairly standard representational theories it is not possible to reformulate *DORSAL as a positive constraint or even a set of positive constraints without resulting in the "pile-up problem" (de Lacy 2002: 90ff., 2006: chapter 2). Similarly, the success of reformulating positive constraints as negative ones depends on the CDL's representational primitives and restrictions. For example, can ONSET be reformulated as *[$_\sigma$V "Return a violation mark for each vowel at the left edge of a syllable" (Kager 1999)? Yes, but *only* if the CDL allows reference to syllable edges in this way and the VR can process it. Whether such boundary reference is possible depends entirely on the CDL and VR.

There are a variety of popular positive constraints that seem to defy easy negative reformulation; e.g. FTBIN: "Assign a violation for each Foot node that does not dominate exactly two moras or two syllables" (after McCarthy and Prince 1986; Prince and Smolensky 1993; see also CHAPTER 40: THE FOOT). Elías Ulloa (2006) proposes breaking the constraint down into constraints that put upper and lower bounds on moraic content. A constraint against more than two sub-foot elements is straightforward using VR: *{$Ft_i \Downarrow x_a$, $Ft_i \Downarrow x_b$, $Ft_i \Downarrow x_c$}, where x_a, x_b, x_c are all distinct prosodic nodes (i.e. μ or σ) and ⇓ here is *dominance*, not *immediate dominance* (↓). A constraint against having *fewer* than two sub-foot daughter nodes is more difficult: How do we penalize a monomoraic/monosyllabic foot structure using a representation ($Ft \Downarrow x$) and the VR, without also penalizing larger structures? Perhaps the VR could be tweaked, making it seek out complete constituents in evaluating constraints. Or it might be possible to capitalize on boundaries, banning a sub-foot prosodic node that is both a left and right foot boundary (e.g. *$[_{Ft}x]_{Ft}$), depending on how boundary reference is permitted in the CDL.

The literature is full of examples of positively and negatively formulated constraints. If there is a single VR, a big challenge is to figure out how to recast the positive constraints into negative ones (or vice versa, depending on one's VR), or alternatively identify sets of negative constraints that take over the effect of the positive ones.

The VR has other interesting side-effects, such as its inability to deal with ALIGN-type constraints and "gradient" constraints generally (McCarthy 2003).

Which option is right? Is there a single VR algorithm and output constraints are really representations? Or is every constraint its own violation mark calculation algorithm, opening the possibility that constraints could assign violation marks in wildly different fashions? The fact that there is such widespread similarity in how constraints work makes me hope for the former option. However, there is no doubt that recasting every extant output constraint in VR-friendly terms is extremely challenging. It is also quite possible that there are several VR-like algorithms, and that constraints are indexed for which violation calculation algorithm they undergo.

The issues become more complicated when we consider the other major set of constraints – correspondence constraints.

3 Faithfulness constraints

Output constraints refer to the output representation in candidates when violation marks are calculated. However, if only output constraints exist, each grammar would produce the same output, regardless of what the input is. So, constraints are needed to regulate inter- and intra-representational relations. For example, there must be constraints that keep outputs from being altered relative to the input form (see McCarthy 2007 for discussion). There is after all a significant difference between [ʔaki] and [ki] in terms of how they relate to the input /aki/. How are constraints used to decide which one of these competitors wins? In more common parlance, how do constraints regulate the *faithfulness* of the output to the input?

3.1 *Containment and correspondence*

Prince and Smolensky (1993) proposed a theory – Containment – that expresses differences in the output relative to the input in terms of output symbols. In the process of generating output representations from an input, a segment can be marked as unparsed (e.g. <k>), which means that it will not be phonetically realized. A segment can also be epenthesized – added to the output – and marked as such (e.g. $\underline{k}$).[7]

Containment Theory provides a straightforward way to ban deletion and epenthesis. Constraints against deletion are simply output constraints that ban unparsed segments: *<•> bans deletion of root nodes, and *$\underline{\bullet}$ bans epenthesis. Such constraints fit straightforwardly into the VR algorithm.

However, McCarthy and Prince (1999) argue for a very different approach. Their proposal – Correspondence Theory (CT) – is now the most widely used theory of inter- and intra-representational identity.

CT proposes a relation – "correspondence" – that holds between root nodes (at least; see §3.2). For example, the root node /k/ in the input /aki/ can be in a correspondence relation with the root node [k] in the output [ʔaki]. However, in the process of making output candidates from an input, GEN has a free hand in generating correspondence relations. So, there will also be a candidate consisting of the input /aki/ and an output [ʔaki], where /k/ corresponds to [ʔ]. Correspondence relations are often written with subscript numerals when it is not obvious which elements correspond (e.g. /$a_1k_2i_3$/: [ʔ$a_1k_2i_3$] *vs.* [ʔ$_2a_1k_2i_3$]).

With a free assignment of correspondence relations, there are many candidates consisting of the input /aki/ and output [ʔaki], where the only difference is correspondence. However, due to the way correspondence constraints work, the vast majority of such candidates will never win under any ranking.

Constraints that regulate the correspondence relations between input and output root nodes are responsible for keeping outputs looking like inputs. For

[7] Prince and Smolensky adopted the theory that epenthetic segments are empty prosodic positions. However, this proposal is no longer widely accepted; I have updated the description of Containment Theory accordingly.

example, IO-MAX returns a violation for each input root node that is not in a correspondence relation with some output root node. So, candidates consist of (at least) an input representation, an output representation, and the correspondence relations that hold between them.

Correspondence was a new concept in phonological theory. In *SPE* and its rule-based successors, input–output faithfulness is an epiphenomenon of rule non-application; an output form is perfectly faithful to its input if no rules apply to it. The more rules that apply, the more likely the output is to become less faithful (though not necessarily: later rules could undo the effect of preceding rules). However, at any point in the derivation of a rule-based theory, except for the first rule application, the original input is *not accessible;* the only accessible representation at any point for a rule theory is the current one (i.e. the input to the rule). In contrast, for every candidate in CT, the input representation is always directly accessible via correspondence: constraints regulate how the output candidate fares relative to the input via correspondence relations.

McCarthy and Prince (1993) proposed a framework of correspondence constraints that has remained substantially unchanged in most subsequent work (see §3.3); see their Appendix A for detailed formulations. I discuss a select few here.

IO-MAX returns a violation mark for every input root node that has no output correspondent. So, IO-MAX returns a violation for /p$_1$a$_2$n$_3$/ → [p$_1$a$_2$], because /n/ has no output correspondent; it returns no violation for /p$_1$a$_2$n$_3$/ → [p$_1$ã$_{2,3}$], because /n/ corresponds to [ã]. Notice that correspondence is not a function; an output segment can correspond to more than one input segment, and vice versa.

IO-DEP returns a violation for every output root node that has no input correspondent; IO-DEP is violated by epenthesis: one violation for /i$_1$/ → [ʔi$_1$], but none for /i$_1$/ → [j$_1$i$_1$].

IO-IDENT[F] regulates feature change; it returns a violation for each input root node whose output correspondent does not have the same value for feature F. There are probably individual IO-IDENT[F] constraints for every feature, and possibly for groups of features too. For example, IO-IDENT[continuant] is violated once in /k$_1$a/ → [x$_1$a].

Other correspondence constraints regulate multiple correspondence: UNIFORMITY (preventing coalescence), INTEGRITY (preventing breaking/diphthongization/split), CONTIGUITY, LINEARITY (preservation of precedence and adjacency), and ANCHOR (preservation of edge proximity).

A caveat is in order here. It is common in informal discussion to say that a constraint *does* some particular task: "MAX bans deletion"; "DEP militates against epenthesis," "UNIFORMITY eschews coalescence," and so on. It is dangerous to take such statements too seriously. MAX does not ban deletion; it simply returns violations if an input root node has no output correspondent. In fact, MAX has effects that do not fit easily into the traditional concept of "deletion." For example, in the mapping /p$_1$a$_2$i$_3$/ → [p$_1$e$_{2,3}$], there has not been any violation of MAX, though a standard *SPE* account would say that there has been deletion relative to the input (/pai/ → [pei] → [pe]). In the mapping /k$_1$a$_2$/ → [ʔa$_2$], it looks as if /k/ has changed into [ʔ], rather than being deleted, yet there is a violation of MAX (in an *SPE*-type analysis there *is* a step with deletion: /ka/ → [a] → [ʔa]). See Gouskova (2007) for detailed discussion of this point in relation to DEP.

Correspondence constraints essentially encourage identity – i.e. preservation of a particular input property in the output, or vice versa. IO-Max and IO-Dep require identity of the number of root nodes; IO-Ident[F] requires identity of feature values, and the other constraints preserve precedence and position.

McCarthy and Prince's correspondence constraints have a simplicity and symmetry that is surprising compared to the complexities of output constraints. For example, IO-Dep is actually IO-Max with an output-oriented focus (i.e. IO-Dep is just "OI-Max"). There also may be constraints of the OI-Ident[F] form; they differ from IO-Ident[F] constraints in terms of coalescence and breaking (see Pater 1999).

3.2 *Constraint Definition Languages for correspondence*

There are many ways to conceive of a CDL for correspondence constraints. Regardless of the CDL, though, it seems challenging to conceive of how the VR algorithm discussed in §2.2 could process correspondence constraints straightforwardly; correspondence constraints demand identity in a way that is not inherent to VR. It is therefore possible that each faithfulness constraint is its own function. However, there is such an overwhelming regularity in terms of how they behave that it is very tempting to assume that there is one algorithm that calculates violations for correspondence constraints, even if it is different from VR.

There is therefore a temptation to set up a separate violation mark calculation algorithm just for correspondence constraints (C-VR). However, it is also possible to combine containment and correspondence so that correspondence constraints are expressed in terms that the VR can process, and regulation of identity follows from the VR. In fact, this is essentially how Prince and Smolensky's Containment Theory works. Here I will discuss an updated version as an example.

For IO-Max, suppose that there is a computable function K over correspondence that is included in each candidate. For every root node in the input, K returns the output root node it corresponds to. If the root node does not correspond to any output root node, K returns a unique phonetically uninterpreted element ψ that is designated as belonging to the output (ψ is unconnected to the output representation, though).[8] Much like the other containment theories, then, IO-Max is $*\psi^{\bullet}$, effectively incurring a violation for each output ψ returned by K. IO-Dep is $*\psi^{I}$.[9] More complex constraints like IO-Max[vowel], which preserve segments that are vowels in the input, can be expressed as $*\{\bullet_i K \psi^{O}, \bullet_i^{I}[+vocalic]\}$.

IO-Ident[F] poses a greater challenge in such a quasi-containment implementation. However, the containment approach essentially replaces identity requirements with a ban on dissimilarity, and this method can be capitalized on for Ident. Let us define a relation G, such that if a root node α dominates feature F, then any α-corresponding β (including α itself, assuming that correspondence is reflexive) is in the relation βGF. The structure in (3) shows two corresponding root nodes α and β. It provides their G relations.

[8] There is a strong similarity in this approach to McCarthy's and Wolf's (2005) "string-based" correspondence, where the input string /x/ is deleted if /x/ corresponds to an empty string ε (correspondence is between strings not root nodes in this view). Here, correspondence is still between root nodes, with an additional uninterpretable root node ψ thrown in.

[9] See www.pauldelacy.net/VR for a partial implementation of this approach.

(3) *Input* *Output*

α_1	β_1
|	|
[−voice]	[+voice]
αG[−voice], αG[+voice]	βG[−voice], βG[+voice]

The constraint IO-IDENT[voice] can then be expressed as $*\{\bullet^{i}G[-\text{voice}], \bullet^{i}G[+\text{voice}]\}$ – i.e. a ban on an input segment having different instances of the same feature with different values. The virtue of this approach is that interrepresentational identity here is recast as avoidance of dissimilarity, which is what many output constraints strive to do (e.g. assimilation constraints; de Lacy 2002: chapter 7; Baković 2007).

However, there have been proposals that support the idea that correspondence constraints are independent functions (or at least that there are several FVRs). Alderete (2001) and Struijke (2000) propose that there are correspondence constraints that differ only in terms of how they assess violations in relation to quantification.

Struijke (2000) argues that there are counterparts to correspondence constraints that differ solely in terms of "quantification." So, while IO-IDENT[nasal] returns a violation for each pair of corresponding input–output segments that have different values for [nasal], IO-∃IDENT[nasal] returns a violation for each input segment for which there is *some* output that has a different value for [nasal]. The difference is seen when an input segment corresponds to more than one output segment. For example, if /ã$_1$/ splits to become [a$_1$n$_1$], IO-IDENT[nasal] returns one violation, because there is an IO pair that disagrees in [nasal]: </ã/, [a]>. However, IO-∃IDENT[nasal] does not return any violations, because for every input segment (i.e. /ã/) there is *some* pair that preserves the [nasal] value (i.e. </ã/, [n]>).

Alderete (2001) argues that at least some faithfulness constraints have "anti-faithfulness" counterparts. For example, OO-IDENT[voice] returns a violation for each corresponding segment that has different values of [voice] (for the OO-part, see §3.3). However, ¬OO-IDENT[voice] returns a violation if there is no pair of correspondents that disagrees on [voice] values. The constraints differ in terms of how they assess violations rather than the representations they refer to.

If Struijke's (2000) and Alderete's (2001) proposals are correct, they pose a serious challenge to an approach that seeks to find a single F-VR algorithm. Their proposals mean that there are sets of constraints that differ only in terms of the procedure of violation mark assignment, not in the representation and relations they refer to.

3.3 *Developments in correspondence constraints*

The theory of correspondence constraints has been reduced, altered, and extended since McCarthy and Prince (1993).

For example, Keer (1999) argues that UNIFORMITY does not exist, with the effect that coalescence is obligatory in certain situations. Similarly, a number of authors have argued that DEP does not exist (e.g. Bernhardt and Stemberger 1998). They observe that output constraints do a similar job to IO-DEP; output constraints (usually) prefer less representational structure over more, and so does IO-Dep. However, Gouskova (2007) argues that the effects of IO-Dep can be distinguished from structural constraints.

Most work has focused on extending correspondence to new nodes (§3.3.1) and dimensions (§3.3.2).

3.3.1 Loci of correspondence

McCarthy and Prince (1993: 14) proposed that correspondence holds between segments. I have adopted a particular theory of representation (autosegmental theory) in this chapter that does not provide an easy way to define "segment"; in this theory, the most natural understanding of McCarthy and Prince's proposal is to say that correspondence holds between "root nodes." McCarthy and Prince also suggest that correspondence could hold between other nodes: tonal nodes, prosodic nodes, and terminal and non-terminal feature nodes.

Myers (1997) develops this idea for tonal nodes. For example, IO-Max-T requires that every input tone node correspond to some output tone node. The most significant effect of the proposal is that tones can survive even when their segmental sponsors are deleted. See Yip (2007) for an introduction to tone constraints; also CHAPTER 45: THE REPRESENTATION OF TONE.

McCarthy (2000) argues for a variety of constraints that require (at least some) prosodic nodes to be in correspondence. Since (most) prosodic structure is apparently absent in inputs, evidence for correspondence between syllables and feet comes from identity across other forms (e.g. base–reduplicant, base–derivative; see §3.3.2).

A widely discussed extension has been feature correspondence. Lombardi (2001) argues that feature-to-feature correspondence is essential in explaining differences between how place features and voice features differ in their behavior.

However, coalescence can be achieved without feature correspondence (e.g. Pater 1999; de Lacy 2002: chapter 8), and a general concern with Max-feature approaches is the lack of observed feature autonomy. In several theories, features do not seem to have the same kind of independence as tones: while tones can survive if their sponsors are deleted, there may not be similar effects for features (featural morphemes are special cases, however; see Wolf 2007b: §2.2 and CHAPTER 82: FEATURAL AFFIXES for discussion).

3.3.2 Dimensions of correspondence

The discussion above has focused on correspondence between inputs and outputs. However, there have been many proposals that extend the reach of correspondence. The proposals fall into two categories: intra-representational correspondence and inter-representational correspondence.

McCarthy and Prince (1999) propose that intra-representational correspondence is found in reduplication (see also CHAPTER 100: REDUPLICATION). A reduplicant morpheme has no input content, but its output segments can correspond to certain other output segments (the reduplicant's "base"). For example, one of the reduplicated forms of Māori [paɾau] 'baffled' is [paɾapaɾau]. The reduplicated segments correspond to other output segments thus: [$p_1a_2ɾ_3a_4p_1a_2ɾ_3a_4u_5$].

McCarthy and Prince (1993) argue that constraints on Base–Reduplicant (BR) correspondence have the same form as constraints on IO correspondence. BR-Max requires every base element to have some correspondent in the reduplicant (violated once in [p$_1$a$_2$ɾ$_3$a$_4$p$_1$a$_2$ɾ$_3$a$_4$u$_5$]), BR-Dep requires every reduplicant segment to have a correspondent in the base, and BR-Ident[F] regulates featural identity between base and reduplicant.

What is surprising about the extension of correspondence to the Base–Reduplicant dimension is that there is essentially one formal mechanism that accounts for both the input–output relation and reduplication. Other theories of reduplication conceive of the phenomenon as involving templates or a type of long-distance assimilation, perhaps through autosegmental spreading (see e.g. McCarthy and Prince 1986). Urbanczyk (2007) gives an overview of BR correspondence and reduplication.

Other intra-representational correspondence relations have been proposed. Kitto and de Lacy (1999) argue that epenthetic segments can correspond to other output segments, resulting in "copy epenthesis": e.g. Winnebago [boːpũnũs] 'hit at random' (Miner 1992). The reason for proposing correspondence here is "overapplication": nasal vowels only occur after nasal consonants in Winnebago, except when epenthetic vowels copy a post-nasal vowel, as above. Such "overapplication" is expected with correspondence, since featural identity of corresponding elements can trump phonotactic restrictions; and is also found in reduplication and other types of correspondence (see Urbanczyk 2007 for discussion of under- and overapplication in reduplication; and Benua 1997 for output–output correspondence).

Hansson (2001) and Rose and Walker (2004) go further in arguing that any output segment can correspond to another output segment. The effect is seen in long-distance agreement. For example, in Chumash sibilants agree in anteriority within a word: /s-ilakʃ/ → [ʃilakʃ] 'it is soft'; cf. [s-ixut] 'it burns'.

There have been many proposals for inter-representational correspondence, too. Benua (1997) proposes that segments in the output representation can correspond to segments in the "trans-derivational base" of that output. The trans-derivational base of a word is basically the word minus its structurally outermost affix. So, the base of *original* [[origin]al] is *origin*. *Original* itself is the base of *originality*, and *origin* is also the base of *originate*.

OO-correspondence can be used to explain why some morphologically complex words do not follow expected phonological patterns, but instead remain similar to their base. For example, in my idiolect (and in many other English-based idiolects) the head foot avoids final syllables in nouns, but otherwise is drawn to the right edge of a PrWd: [əd('məsə)bu] *admissible*, [ədməsə('bələ)ri] *admissibility* ([ə] can be stressed in my dialect, and /l/ → [u] outside onsets). However, with some affixes the foot does not get drawn rightward as expected: [əd('məsə)bunəs] *admissibleness*, *[ədmə('səbu)nəs]. When *ness* appears in a word, it subjects the candidate to an OO-faithfulness requirement that has the effect of forcing the correspondent of the base's head syllable to also be a head. So *[ədmə('səbu)nəs] loses to [əd('məsə)bunəs] because the corresponding head syllable in the base [əd('məsə)bu] is [mə], not [sə].

Further work on interword relationships has argued that candidates should consist of entire output paradigms of related word forms. See McCarthy (2005) for references and discussion.

Yet other work has proposed correspondence relations between the output representation and another output representation that is identified by a special selection mechanism, with the aim being to account for phonological opacity.

4 Possible and impossible constraints

Even after defining the CDL's representational elements, relations, and algorithm(s) assigning violation marks, there remains the question of which constraints actually exist.

I wish I could list all the phonological constraints that exist in the human brain here. Unfortunately, there is no agreed-upon list. Many constraints have been proposed, and many algorithms too. Given the rapid changes in phonological theories and variety of constraint proposals, it is more useful to discuss general intrinsic and extrinsic restrictions on theories of constraints.

Every CDL imposes intrinsic limits on possible constraints. The nature of the elements and relations by which constraints are defined means that some imaginable constraints could not occur. For example, suppose the CDL has no disjunction operator. A constraint that assigns a violation to a segment if it is either [+voice] or [labial] is then not possible – it is impossible to formulate using the CDL's syntax. Similarly, the VR itself may impose "restrictions" on constraints in the sense that certain constraints might be well formed in the CDL, but not assign violation marks. In the VR discussed above, constraints that had unconnected representational elements would not assign violation marks; so, while such unconnected constraints could exist, they are effectively inert and will never be observed to have an effect on selecting a winning candidate. Other cases were discussed in §1.1 and §2.2.

It is also possible (even likely) that there are extrinsic limits on constraints. An extrinsic limit is a restriction on particular types of constraint even though the constraint would have a well-formed syntactic structure in the CDL. For example, suppose there was a CDL that made it possible to define a constraint that banned syllable onsets, yet such a constraint did not exist: an extrinsic limit would have to be responsible. The alternative is to suppose that there are almost no significant extrinsic limits: the set of constraints includes every constraint definable using the CDL (up to a certain level of complexity). The issue of extrinsic limits is a very difficult one. The first issue addressed below is methodological: how can we tell whether there are extrinsic limits on constraints (§4.1)? §4.2 discusses where those extrinsic limits come from.

4.1 Evidence for extrinsic restrictions

The majority of OT theories and subtheories do propose many extrinsic restrictions on constraints, specified by the CDL (cf. discussion in Blevins 2004). The evidence comes from restrictions and requirements that cannot be attributed to non-cognitive mechanisms.

To explain, suppose we never observe a particular phonological property in any human language, like an epenthetic [k] (e.g. /iti/ never surfaces as [<u>k</u>iti] in any grammar). It is possible that the lack of [k]-epenthesis is due to constraints. For [k] to be epenthetic, there has to be a (set of) output constraint(s) that returns violation marks for every segment except [k]. Without such a constraint, epenthetic [k] won't occur (see e.g. de Lacy 2006: ch. 3).

However, there are potentially non-CDL reasons why epenthetic [k] is never observed. Some other part of the phonological component could be responsible (see §4.2). There is also luck – e.g. war, pestilence, or plague – which may have accidentally wiped out all speakers of languages with epenthetic [k]. After all, every theory of phonology predicts many tens of thousands of distinct phonological systems, and only a few thousand have existed and will ever exist.

For epenthetic [k] to be observed, it must also be learnable. Actuation of a phonological change comes about through learner misperception or misarticulation. So if epenthetic [k] cannot come about through such a situation, it won't be observed. Even if a sound change can be actuated easily, it will quickly disappear if it cannot be transmitted effectively. In this particular case, though, there is evidence that learners could misperceive vowel hiatus as involving a [k] (Kingston and de Lacy 2006).

So, suppose a phonological situation P never occurs. Suppose further that the lack of P cannot be ascribed to non-constraint grammatical processes or extra-phonological mechanisms. In such a case, phonological extrinsic restrictions on constraints are responsible for the lack of P. For epenthetic [k], it is easy to come up with a set of constraints that penalize everything except [k] (e.g. *Labial, *Coronal, *Glottal), so the CDL must not permit this set of constraints (or at least, this set of constraints with free ranking).

There are several other methods of determining that a particular phonological situation is due to constraints. See Kingston and de Lacy (2006: §3.3) and references cited for discussion.

4.2 Origins and universality of constraints

If there are restrictions on possible constraints, where do those restrictions come from? There are fundamentally two different proposals: (a) innateness and (b) constraint-construction mechanisms that refer to phonology-external structures.

The innateness view is that constraints are hard-wired into the brain (i.e. part of our genetic make-up). The "hard-wired" view comes in two versions. One is that each constraint is specified independently. In this version, only those constraints that are hard-wired into the brain exist, so extrinsic limits on constraints boil down to genetics. The other version is that there are hard-wired algorithms that automatically generate constraints – "constraint generators" (sometimes called "schemas"). For example, there would be an "Ident[F]" constraint generator that produces constraints with the form D-Ident[F] where D is a pair of dimensions (input–output, base–reduplicant, etc.), and F is a subsegmental node. The generator is "complete" in that it would generate constraints for every D and every F (Green 1993). The constraint-construction algorithms determine which constraints exist.

The alternative is to propose mechanisms that are derived from phonology-external mechanisms, or at least can take phonology-external factors into account. A growing body of work argues that there are many algorithms that take phonetic factors like ease of articulation and perceptual distinctiveness into account in evaluating which phonological constraints to generate. In this view, limits on constraints are a combination of the inherent limits of the constraint-construction algorithm *and* the restrictions imposed by the phonology-external factors that those algorithms refer to.

For example, Hayes (1995) discusses phonological constraints on voiced stops. Phonetic voicing in stops is hard to maintain; the further back the stop is, the more difficult it is to maintain voicing during the closure phase: it is harder to maintain voicing during the closure phase of [g] than for [d], and it is harder for [d] than for [b]. Suppose there is a constraint generator that produces constraints on voicing in stops. It could imaginably generate many constraints, e.g. *g, *d, *b, *g/d, *g/b, *d/b, *g/d/b (where *x/y means "Return a violation for any segment that is x or y"). However, if the mechanism referred to articulation in a way that reflected voicing difficulty, the constraints would be winnowed down to *g, *g/d, *g/d/b. Hayes (1995) further observes that the CDL's intrinsic representational restrictions could impose further limits on the possible constraints: *g/d, for example, is not definable in some feature theories, as there is no feature that [g] and [d] share to the exclusion of [b]; with such representational theories, the only constraints generated by the mechanism would be *g and *g/d/b (i.e. *[+voice, −continuant, −nasal]).

To summarize, the majority of work in OT adopts the idea that there are constraint generators. However, there is ongoing disagreement over whether constraint generators can refer to phonology-external factors like ease of articulation and perceptual difficulty. Gordon (2007) provides discussion and references; see also CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY.

A related issue is constraint universality. A constraint is "universal" if it exists in every grammar. A constraint can exist in every grammar because it is hard-wired into CON (the set of constraints), or because it is produced by a constraint generator (see §4.2) that produces the same constraints in the same way for every grammar. A "language-specific" constraint is one that exists in only some languages; it must be learned. For specific discussions of constraint universality, see Green (1993), Prince and Smolensky (1993), and McCarthy (2002: §1.2.1, §3.1.5.2).

There is an important nuance to constraint universality/language specificity. It is possible that constraints are not universal, but rather *constraint generators* are. For example, ALIGN is a constraint generator that exists in every grammar. However, if ALIGN is allowed to take individual morphemes (or morphs) as arguments, it could produce language-specific constraints like ALIGN($[um]_{Af}$, L; Stem, L), "The affix *um* occurs stem-initially, is a prefix" for Tagalog, and ALIGN($[ka]_{Af}$, L; Ft′, R), "The affix *ka* follows, is a suffix to, the head foot" for Ulwa (McCarthy and Prince 1993). So, while ALIGN($[um]_{Af}$, L; Stem, L) does not exist in every language, the constraint generator that created it does. The point could be extended to other constraint generators, and even those that refer to phonology-external factors. If a constraint generator refers to an articulatory or acoustic factor that varies among speakers, it could be that the same constraint generator will produce speaker-specific constraints.

5 Summary

This chapter has left a vast number of issues about constraints untouched and only barely skimmed over a few others. However, a few points about constraints emerge. A formal theory of constraint form – a "Constraint Definition Language" – provides valuable insight into which constraints can and cannot exist. There is fairly widespread (if tacit) agreement on many aspects of such a CDL, but also many disagreements about both fundamental issues and the details. Constraints are only one part of a complex system that determines phonological winners;

constraints themselves do not determine winners, and no constraint or set of constraints has any predictive power on its own. Only when the entire collection of constraints, GEN, EVAL, and the phonetic module interface are examined together can anything be asserted about the predictive power or restrictive nature of the theory.

ACKNOWLEDGMENT

My thanks to Marc van Oostendorp, Keren Rice, Matt Wolf, and an anonymous reviewer for their extensive comments.

REFERENCES

Alderete, John. 2001. Dominance effects as transderivational anti-faithfulness. *Phonology* 18. 201–253.

Bach, Emmon. 1968. Two proposals concerning the simplicity metric in phonology. *Glossa* 2. 128–149.

Baković, Eric. 2007. Local assimilation and constraint interaction. In de Lacy (2007), 335–351.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Ph.D. dissertation, University of Massachusetts, Amherst.

Bermúdez-Otero, Ricardo. Forthcoming. *Stratal Optimality Theory*. Oxford: Oxford University Press.

Bernhardt, Barbara H. & Joseph P. Stemberger. 1998. *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45–86.

Bye, Patrik & Paul de Lacy. 2000. Edge asymmetries in phonology and morphology. *Papers from the Annual Meeting of the North East Linguistic Society* 30. 121–135.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Chomsky, Noam & Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8. 425–504.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 245–306. Cambridge, MA & Oxford: Blackwell.

de Lacy, Paul. 2002. The formal expression of markedness. Ph.D. dissertation, University of Massachusetts, Amherst.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.

de Lacy, Paul (ed.) 2007. *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press.

Eisner, Jason. 1997. What constraints should OT allow? Handout from paper presented at the 71st Annual Meeting of the Linguistic Society of America, Chicago (ROA-204).

Elías Ulloa, José. 2006. Theoretical aspects of Panoan metrical phonology: Disyllabic footing and contextual syllable weight. Ph.D. dissertation, Rutgers University (ROA-804).

Gordon, Matthew. 2007. Functionalism in phonology. In de Lacy (2007), 61–78.

Gouskova, Maria. 2007. Dep: Beyond epenthesis. *Linguistic Inquiry* 38. 759–770.

Green, Thomas. 1993. The conspiracy of completeness. Unpublished ms., MIT (ROA-8).

Hall, T. A. 2007. Segmental features. In de Lacy (2007), 311–333.
Hansson, Gunnar Ólafur. 2001. Theoretical and typological issues in consonant harmony. Ph.D. dissertation, University of California, Berkeley.
Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.
Itô, Junko. 1986. Syllable theory in prosodic phonology. Ph.D. dissertation, University of Massachusetts, Amherst.
Kager, René, 1999. *Optimality Theory.* Cambridge: Cambridge University Press.
Kager René, Harry van der Hulst & Wim Zonneveld (eds.) 1999. *The prosody–morphology interface.* Cambridge: Cambridge University Press.
Keer, Edward. 1999. Geminates, the OCP and the nature of CON. Ph.D. dissertation, Rutgers University.
Kingston, John & Paul de Lacy. 2006. Synchronic explanation. Unpublished ms., University of Massachusetts, Amherst & Rutgers University. Available (July 2010) at www.pauldelacy.net/webpage/publications.html.
Kiparsky, Paul. Forthcoming. *Paradigms and opacity.* Stanford: CSLI.
Kitto, Catherine & Paul de Lacy. 1999. Correspondence and epenthetic quality. In Catherine Kitto & Carolyn Smallwood (eds.) *Proceedings of AFLA VI: The Sixth Meeting of the Austronesian Formal Linguistics Association*, 181–200. Toronto: Department of Linguistics, University of Toronto.
Lombardi, Linda. 2001. Why Place and Voice are different: Constraint-specific alternations in Optimality Theory. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 13–45. Cambridge: Cambridge University Press.
Łubowicz, Anna. 2005. Locality of conjunction. *Proceedings of the West Coast Conference on Formal Linguistics* 24. 254–262.
McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16. 331–399.
McCarthy, John J. 2000. The prosody of phase in Rotuman. *Natural Language and Linguistic Theory* 18. 147–197.
McCarthy, John J. 2002. *A thematic guide to Optimality Theory.* Cambridge: Cambridge University Press.
McCarthy, John J. 2003. OT constraints are categorical. *Phonology* 20. 75–138.
McCarthy, John J. 2005. Optimal paradigms. In Laura J. Downing, T. A. Hall & Renate Raffelsiefen (eds.) *Paradigms in phonological theory*, 170–210. Oxford: Oxford University Press.
McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory.* London: Equinox.
McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.
McCarthy, John J. & Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology 1993*. 79–153.
McCarthy, John J. & Alan Prince. 1999. Faithfulness and identity in prosodic morphology. In Kager *et al.* (1999), 218–309.
McCarthy, John J. & Matthew Wolf. 2005. Less than zero: Correspondence and the null output. Unpublished ms., University of Massachusetts, Amherst (ROA-722).
Miner, Kenneth L. 1992. Winnebago accent: The rest of the data. *Indiana University Linguistics Club Twenty-Fifth Anniversary Volume*, 28–53. Bloomington: Indiana University Linguistics Club.
Myers, Scott. 1997. OCP effects in Optimality Theory. *Natural Language and Linguistic Theory* 15. 847–892.
Pater, Joe. 1999. Austronesian nasal substitution and other NÇ effects. In Kager *et al.* (1999), 310–343.
Potts, Christopher & Geoffrey K. Pullum. 2002. Model theory and the content of OT constraints. *Phonology* 19. 361–393.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Rose, Sharon & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80. 475–531.

Samek-Lodovici, Vieri & Alan Prince. 1999. Optima. Unpublished ms., University College London & Rutgers University (ROA-363).

Schein, Barry & Donca Steriade. 1986. On geminates. *Linguistic Inquiry* 17. 691–744.

Struijke, Caro. 2000. Existential faithfulness: A study of reduplicative TETU, feature movement and dissimilation. Revision of Ph.D. dissertation, University of Maryland, College Park.

Suzuki, Keiichiro. 1998. A typological investigation of dissimilation. Ph.D. dissertation, University of Arizona.

Urbanczyk, Suzanne. 2007. Reduplication. In de Lacy (2007), 473–493.

Wilson, Colin. 2001. Consonant cluster neutralisation and targeted constraints. *Phonology* 18. 147–197.

Wolf, Matthew. 2007a. What constraint connectives should be permitted in OT? *University of Massachusetts Occasional Papers in Linguistics* 36. 151–179.

Wolf, Matthew. 2007b. For an autosegmental theory of mutation. In Leah Bateman, Michael O'Keefe, Ehren Reilly & Adam Werle (eds.) *Papers in Optimality Theory III*, 315–404. Amherst: GLSA.

Yip, Moira. 2007. Tone. In de Lacy (2007), 195–227.

64 Compensatory Lengthening

Randall Gess

1 Types of compensatory lengthening

Compensatory lengthening is the lengthening of one segment, referred to here as the "target," in compensation for the loss or reduction of another, referred to here as the "trigger" (see also CHAPTER 37: GEMINATES and CHAPTER 20: THE REPRESENTATION OF VOWEL LENGTH). The segments are usually in close proximity to one another – either adjacent or in adjacent syllables. Theoretically, a consonant can be lengthened in compensation for the loss or reduction of another consonant or vowel, and a vowel can be lengthened in compensation for the loss or reduction of another vowel or consonant. In fact, an argument can be made that all types of compensatory lengthening exist, although as Table 64.1 indicates, some types are far more common than others. There is also a problem in the classification of some types as compensatory lengthening proper rather than as instantiations of other processes, such as total assimilation (2A, B) or rhythmic lengthening (1D).

In Table 64.1, Row 1 lists cases in which the target for lengthening is a vowel, and Row 2 lists cases in which the target for lengthening is a consonant. Column A lists cases in which the trigger for lengthening (the reduced or deleted segment) is a reduced consonant following the target, Column B those in which the trigger is a reduced consonant preceding the target, Column C those in which the trigger is

Table 64.1 Types of compensatory lengthening

		Trigger			
		A / __ C	B / C __	C / __ (X)V	D / V(X) __
Target	1: V	numerous	limited	numerous	limited
	2: C	limited	numerous	isolated	limited

a reduced vowel following the target, and Column D those in which the trigger is a reduced vowel preceding the target.

Examples representative of each cell in Table 64.1 are provided in the following brief sections.

1.1 *Type 1A* (Target V; Trigger / __ C)

(1) *Old French* (Gess 1998, 1999)

a.	*blasmer*	[blazmer]	>	[blaːmer]	'to blame'
b.	*angle*	[ãnglə]	>	[ãːglə]	'angel'
c.	*large*	[lardʒə]	>	[laːdʒə]	'wide'

The type of compensatory lengthening in (1) is common. Kavitskaya (2002: App. 1) lists 58 languages in which it is manifested. Other cases not mentioned by Kavitskaya can be found in Gordon (1999) and Beltzung (2008).

More exotic types of compensatory lengthening triggered by a following consonant are ones in which the trigger is an intervocalic consonant, or in which the target and trigger are separated by an intervening consonant, i.e. in which the triggering consonant is the second of a sequence of two intervocalic consonants (e.g. Ancient Greek *odwos > East Ionic /oːdos/ (Wetzels 1986: 310). Hayes (1989) refers to this type of compensatory lengthening as a "double flop," a term which has gained currency in the literature (see Beltzung 2008 for an extensive discussion of "exotic" types of compensatory vowel lengthening triggered by consonant loss).

1.2 *Type 1B* (Target V; Trigger / C __)

(2) *Samothraki Greek* (Katsanis 1996, as reported in Topintzi 2006; see also CHAPTER 55: ONSETS)

a.	'rafts	>	'aːfts	'tailor (MASC)'
b.	'ruxa	>	'uːxa	'clothes'
c.	'rema	>	'eːma	'stream'
d.	'protos	>	'poːtus	'first'
e.	'vrisi	>	'viːs'	'tap'
f.	me'trun	>	mi'tuːn	'they count'
g.	'extra	>	'extaː	'hostility'

The type of lengthening shown in (2) is rare and somewhat controversial (see Beltzung 2008 for an overview), as it is predicted not to occur by the framework of Moraic Phonology developed in Hayes (1989). In Samothraki Greek this process is limited to rhotics in word-initial or post-consonantal position – it does not occur when the segment is in intervocalic position or in a coda position (there is no deletion of syllable-final /r/ in Samothraki Greek). According to Beltzung (2008), the segments implicated in this type of compensatory lengthening are rhotics, pharyngeals, and laryngeals.

1.3 *Type 1C* (Target V; Trigger / __ CV)

(3) *Hungarian* (Kálmán 1972, as reported in Kavitskaya 2002)

a.	*wizi	>	viːz	'water'
b.	*tyzy	>	tyːz	'fire'
c.	*utu	>	uːt	'road'
d.	*ludu	>	luːd	'goose'
e.	*neʒi	>	neːʒ	'four'
f.	*modoru	>	modoːr	'bird'
g.	*teheni	>	teheːn	'cow'

Cases like the one illustrated here are relatively common, but appear to be more phonologically restricted and less widespread than the type shown in §1.1. Kavitskaya (2002: App. 2) lists 21 languages (neglecting to include Yapese (Jensen 1977), where it appears to be a synchronic process) in which this process occurs or has occurred, whereas she lists 58 languages (App. 1) in which the CVC type of compensatory lengthening shown in §1.1 is manifested. As indicated in Table 64.1, this type of compensatory lengthening does not always involve an intervening consonant (e.g. Old French [fiː] < [fiə]; Pope 1952: 205).

1.4 *Type 1D* (Target V, Trigger / VC __)

According to Hayes (1989: 284), this type of compensatory lengthening "appears not to exist." However, a process in Macuxi may prove problematic for this claim. In a section entitled "Compensatory length," Carson (1981: 50) states that "[w]hen a short vowel is suppressed, the vowel that immediately precedes a stop consonant in its vicinity is lengthened." Representative data are shown in (4).

(4) *Macuxi* (Carson 1981)

kasa'pan	→	ksaa'pan	'sand'
kusupa'ra	→	'ksuu'pra	'machete'
'wakiri'pe	→	'wakrii'pe	'agreeable'
'miki-'rɨ	→	'miɨ'kri	'he'

It appears in all the examples but the last one that the lengthened vowel is the one that follows the deleted one. However, the absence of otherwise expected *['mɨ'krɨɨ] in the last example may be attributed to what Pessoa (2006: 78, 2009: 117), citing Abbott (1991), refers to as the absence of "silaba alongada fonológica na última posiçao," since final syllables are already rhythmically lengthened.

According to Kavitskaya (2002: 149), lengthening in Macuxi (as described in Kager 1997) is not a case of compensatory lengthening, and "should be attributed to the properties of iambic systems" since it "happens regardless of syncope" (see also CHAPTER 44: THE IAMBIC–TROCHAIC LAW). She cites the following examples from Kager (1997: 466–467).

(5) *More examples from Macuxi*

/piripi/	(ˌpriː).('piː)	'spindle'
/waimujami/	(ˌwai).(mjaː).('miː)	'rats'

The first example is taken by Kager (1997) from Hawkins (1950: 89), and the second from Abbott (1991: 147). In both examples, the final syllable of the form in question is apparently lengthened, and there is no vowel syncope to point to as a trigger. However, with respect to the first form, no lengthening in any position is indicated by Hawkins (1950), who says nothing more about final vowels than that "[t]he last vowel in each stress contour in the basic form of any utterance is always retained," and he transcribes the form as 'pripí', with the acute accent marking "the end point of contours" (Hawkins 1950: 89).

It may be that length is present and not noted by Hawkins, but if so, it is just as easily described as phrase-final lengthening that is independent of iambic (foot-level) lengthening. According to Abbott (1991: 145), "[t]he final CV in a phonological phrase (i.e. a phrase bounded by a pause) is always long and stressed." Abbott describes rhythmically derived length on even-numbered V or CV syllables, counting from the left, as well as on final syllables. Again, though, Hawkins (1950) says nothing about lengthening in any context, and does not indicate it in any transcriptions. Nevertheless, Kager (1997) introduces foot-level lengthening systematically in forms taken from both authors, as well as syncope in forms where Abbott indicates none. Hawkins (1950) does discuss "stress contours" in which the contour consists of "a stretch of speech marked by loud stress on the last vowel," and he notes that "[w]hen more than one word occurs in a stress contour, the last vowel in each non-final word in the contour is retained." However, retention of a vowel is a far cry from lengthening.

Carson (1981) does not indicate final lengthening either. Kager (1997) chooses to ignore the data from Carson (1981), cited above, noting that she posits "lexical tone rather than stress," and that her data must be "based on a different dialect than those studied by Hawkins and Abbott" (1997: 466). In fact Carson describes lexical pitch accent (1981: 42–45), which may be "disturbed" at the phrasal level (1981: 46). If Carson is correct, that the variety she documented manifests pitch accent, then the lengthening she describes cannot appropriately be attributed to iambic lengthening.

Two points arise from the preceding discussion. First, it is not clear that rhythmic lengthening, independent of vowel reduction or deletion, does occur in the Macuxi variety described by Abbott (1991) (and Hawkins 1950, if lengthening other than in final syllables even occurs in the variety he describes). Second, the lengthening described by Carson (1981) cannot be dismissed on the grounds suggested by Kavitskaya (2002).

An important question arises with respect to the first point in the preceding paragraph: If rhythmic lengthening is always tied to vowel reduction or deletion, can it properly be considered compensatory lengthening? Kavitskaya's point is that it is not – that it is better in this case to consider it a property of iambic systems, together with rhythmic vowel deletion. It is unclear, however, why a foot-based process that can be described as CVCV > CvCV: (where "v" represents a reduced vowel) should be treated any differently than one that can be described as CVCV > CV:Cv, i.e. the fairly common type of compensatory lengthening described in (3) (1C in Table 64.1), that is uncontroversially labeled as such.

Two other types of compensatory lengthening can result from a triggering vowel preceding the target vowel, in these cases with no intervening consonant. The first of these is compensatory lengthening through glide formation (typically from high vowels), a relatively common synchronic process in Bantu languages (e.g. Ganda

/li+ato/ 'boat' → [ljaato]; Clements 1986: 47). The other type of process is also attested in Ganda, involving the deletion of non-high vowels in prevocalic position (e.g. /ka+oto/ 'fireplace (DIM)' → [kooto]; Clements 1986: 49).

1.5 *Type 2A* (Target C; Trigger / __ C)

(6) *Semitic* (Lipiński 2001: 195)

*usˤtabbit	>	usˤsˤabbit	'he imprisoned'	(Assyro-Babylonian)
*atˤtarad	>	atˤtˤarad	'I sent'	(Assyro-Babylonian)
*itˤtalaba	>	itˤtˤalaba	'he sought'	(Arabic)
*ʔətzəkkar	>	ʔəzzəkkar	'I remember'	(Ge'ez)
*jilkədenhu:	>	jilkədennu:	'he shall capture him'	(Hebrew)
*gəmaːlathu:	>	gəmaːlattu:	'she weaned him'	(Hebrew)
*wɛsfi	>	wɛssi	'awl'	(Gurage)
*nisf	>	nəsˤsˤ	'half'	(Colloquial Arabic)

This type of compensatory lengthening appears to be relatively uncommon, and like the type described in the following section, it is not formally distinguishable from total assimilation.

1.6 *Type 2B* (Target C; Trigger / C __)

(7) *Bengali* (Hayes and Lahiri 1991: 81)

bɔrʃa	~	bɔʃʃa	'rainy season'
bɔrdi	~	bɔddi	'elder sister'
bʰorti	~	bʰotti	'full'
kortʃʰe	~	ko-tʃʰe	'do-3PRES'
kor-lo	~	kol-lo	'do-3FUT'

Although this type of compensatory lengthening is quite common, unlike the type illustrated in (6), it shares with that process the lack of any formal distinctiveness from total assimilation.

1.7 *Type 2C* (Target C; Trigger / __ V)

(8) *Bulgarian* (Shishkov 2002)

'balite	>	'bal'ːte	'the bales'
er'genite	>	er'gen'ːte	'the bachelors'
ku'ʃarite	>	ku'ʃar'ːte	'the (sheep) pens'
'belezite	>	'belesːte	'the scars'
'babinata	>	'babinːta	'the grandmother's (things)'
ven'tʃiloto	>	ven'tʃilːtu	'the wedding'
ameri'kancite	>	amerː'kancite	'the Americans'
done'sa	>	donː'se	'bring (3SG)'

Cases such as the one illustrated in (8), in which a consonant is lengthened before a following reduced vowel, appear to be isolated. It is noteworthy that the consonants involved in the Bulgarian process are of relatively high sonority – only sonorants and /z/, although there is at least one attested case of synchronic compensatory lengthening in this category in which sonority does not appear to be relevant (compensatory lengthening resulting from glide formation, as in Ilokano /'luto+en/ → /lutt'w-en/ 'cook-GOAL FOCUS'; Hayes 1989: 269).

1.8 *Type 2D* (Target C; Trigger / V __)

(9) *Ganda* (adapted from Clements 1986: 62–63)

/li + kubo/ > [kkubo]
/li + tabi/ > [ttabi]
/li+ daala/ > [ddaala]

According to Clements (1986: 6), the synchronic rule deriving geminates from a CV prefix is "a restructuring of the historical situation, in which a phonetically motivated rule is replaced by a morphologically conditioned one." Clements assumes the geminates to have arisen historically from earlier *Vį sequences (į represents an upper high front vowel), "with the process giving rise to consonant gemination [being] one in which the articulation of a consonant is anticipated on a preceding postvocalic *į" (1986: 65). They are now associated with a certain class of nominal prefixes.

2 Approaches to compensatory lengthening

Most documented cases of compensatory lengthening, at least those formally distinguishable from total assimilation of adjacent consonants, involve the compensatory lengthening of vowels. Furthermore, the most common types of compensatory lengthening of vowels involve those in which the trigger follows, rather than precedes, the target – i.e. the types of cases illustrated in §1.1 and §1.3. These two types of compensatory lengthening are commonly referred to as CVC and CVCV compensatory lengthening, respectively. In this section, I focus on these most common types and, since synchronic cases of compensatory lengthening are derived from historical ones, I focus on the diachronic instantiation of the processes.

I first summarize three general approaches to compensatory lengthening, all of which have in common an implicit assumption that the phenomenon is speaker-controlled. A fourth section outlines an alternative approach put forth in Kavitskaya's (2002) quite comprehensive treatment of compensatory lengthening, which may be considered somewhat radical in proposing a strictly listener-oriented account of the process. The relevance of the various approaches to synchronic cases of compensatory lengthening, as well as to the other types illustrated in §1, is discussed in §4.

The first three approaches to be examined in this section fall into two categories, as described by Kavitskaya (2002): one that treats compensatory lengthening as a type of conservation and one that does not. The first category is the most common, and assumes that compensatory lengthening is fundamentally teleological in that its goal is to preserve length present in the input in the output string. Being the most common category, it comprises two of the three approaches: a phonetic conservation approach and a phonological conservation approach. The third approach, in a category of its own, is the non-conservation approach, which denies the existence of any intrinsic connection between the loss of a segment and the lengthening of another.

2.1 *Phonetic conservation approach*

In a phonetic conservation approach, compensatory lengthening is viewed as a goal-oriented process functioning to preserve some or all of the physical duration of lost segmental material.

Timberlake (1983) discusses a case in Slavic in which a number of modern dialects have long or tense vowels in syllables that preceded a weak jer in Late Common Slavic (see also CHAPTER 122: SLAVIC YERS). This is a case of CVCV compensatory lengthening (Timberlake does not discuss CVC cases). According to Timberlake, a long reflex of a vowel in a syllable before a Late Common Slavic weak jer "is in some way a result of the phonetic weakening and the eventual phonemic loss of the following jer vowel" (1983: 293). He suggests that:

> Late Common Slavic was subject to a constraint on the preservation of word timing, such that phonetic reduction in one syllable (containing the "weak" jer) was compensated for by increased phonetic duration in the preceding, "strong" syllable.

In Timberlake's model, compensatory lengthening takes place phonologically through re-analysis. Re-analysis depends upon both the phonemic elimination of jers and the surpassing of a "critical duration" on the part of the phonetically lengthened preceding vowel. If, when jers are "eliminated phonemically, either by identification with another vowel or by identification with null" (1983: 299), phonetically lengthened vowels are sufficiently lengthened, the latter are re-analyzed as phonemically long (or tense). Timberlake sets the critical duration for re-analysis arbitrarily at anything beyond 1.5 times "full duration (nearly or exactly 1.0 morae. [. . .] numerical values for duration [. . .] are intended to be highly approximate)" (1983: 298).

Timberlake's model is an additive one, "in which the duration of vowels is adjusted by adding or subtracting increments of duration depending on various factors." The various factors at play in Late Common Slavic were the consonant intervening between the jer and the lengthened preceding vowel; the position of the CVCə sequence in the word (final or internal); and the accent of the lengthened vowel.

The phonetic process of compensatory lengthening is described by Timberlake (1983: 298) as in (10).

(10) *Compensatory lengthening as a phonetic process*

$$/CVCə/ > [CV^{1.0+\alpha}Cə^{-\alpha}]$$

The formula in (10) states that for any reduction of value α in the phonetic length of a jer, a preceding vowel is realized at full duration (1.0) plus α.

In order to model the gradual nature of phonetic lengthening, Timberlake breaks the process down into discrete stages, arbitrarily shown in 0.2 increments, as illustrated in (11), from Timberlake (1983: 298).

(11) *Jer reduction and compensatory lengthening*

a. /CVCə/ > [CV$^{1.2}$Cə$^{-0.2}$] {$\alpha = 0.2$}
b. /CVCə/ > [CV$^{1.4}$Cə$^{-0.4}$] {$\alpha = 0.4$}
c. /CVCə/ > [CV$^{1.6}$Cə$^{-0.6}$] {$\alpha = 0.6$}
d. /CVCə/ > [CV$^{1.8}$Cə$^{-0.8}$] {$\alpha = 0.8$}

Finally, as indicated above, Timberlake assumes re-analysis at anything beyond 1.5 times full duration. Regarding the cut-off, Timberlake (1983: 299) explains:

> When reduced jers were eliminated phonemically, the phonetic phase of CL was necessarily interrupted, and the lengthened variant of a vowel in strong position had to be identified as phonemically long (tense) or short (lax).

This view of re-analysis is illustrated schematically in (12).

(12) *Phonemic analysis*

a. [CV$^{1.2}$Cə$^{-0.2}$] ⇒ /CVC/ {$\alpha = 0.2$}
b. [CV$^{1.4}$Cə$^{-0.4}$] ⇒ /CVC/ {$\alpha = 0.4$}
c. [CV$^{1.6}$Cə$^{-0.6}$] ⇒ /CV:C/ {$\alpha = 0.6$}
d. [CV$^{1.8}$Cə$^{-0.8}$] ⇒ /CV:C/ {$\alpha = 0.8$}

Timberlake's account of CVCV compensatory lengthening can be straightforwardly extended to CVC compensatory lengthening. A demonstration of this is provided in Gess (forthcoming).

Explicit criticisms of phonetic conservation approaches to compensatory lengthening are minimal. The most obvious thing to point out here is the limited relevance of this approach to synchronic cases of compensatory lengthening. That is, while the approach may be useful in showing how compensatory lengthening may arise historically, it is not well suited for modeling the processes in synchronic grammars unless the process at hand is, or at least could be, a change in progress (i.e. it is a gradient, post-lexical process). Another problem with the phonetic conservation approach, pointed out by Gess (forthcoming), is that its extension to CVC cases of compensatory lengthening, while straightforward in a mechanical sense, seems to entail at least implicitly the problematic assumption that moras associated with consonants are equivalent in duration to those associated with vowels. This problem could be overcome by making assumptions regarding the formalisms used more explicit.

2.2 *Phonological conservation approach*

In a phonological conservation approach, compensatory lengthening is viewed as a goal-oriented process functioning to preserve some aspect of the phonological representation (a suprasegmental unit) associated with the loss of segmental material. In fact, compensatory lengthening phenomena crucially informed the debate as to the best way to characterize the prosodic tier (or timing tier) assumed in autosegmental phonology, i.e. in terms of C- and V-slots, X-slots, or moras (McCarthy 1979; Clements and Keyser 1983; Hyman 1984, 1985; Levin 1985; Lowenstamm and Kaye 1986; McCarthy and Prince 1986; Hayes 1989; Beltzung 2008 (especially Chapter 3); see also CHAPTER 54: THE SKELETON).

In order to explain constraints on compensatory lengthening (e.g. that triggers in CVC cases are only coda consonants and not onset consonants [an assumption later proven problematic; Topintzi 2006; Beltzung 2008], and even more specifically that coda consonants are triggers only when they contribute to syllable weight in the language in question), Hayes (1989), in probably the most influential single article on compensatory lengthening, suggests that lengthening only occurs when deletion results in an empty prosodic position and that only a prosodic frame defined in terms of moras yields the correct typological results. (Hayes 1989: 260–261 also provides a simple and straightforward demonstration of the inability of a linear approach to account for compensatory lengthening.)

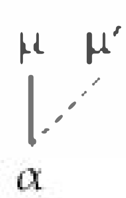
Hayes (1989) accounts for CVC compensatory lengthening as illustrated in (13), with Latin [kasnus] > [kaːnus] 'dog'.

(13) *Compensatory lengthening in* CVC *sequences* (Hayes 1989: 262)

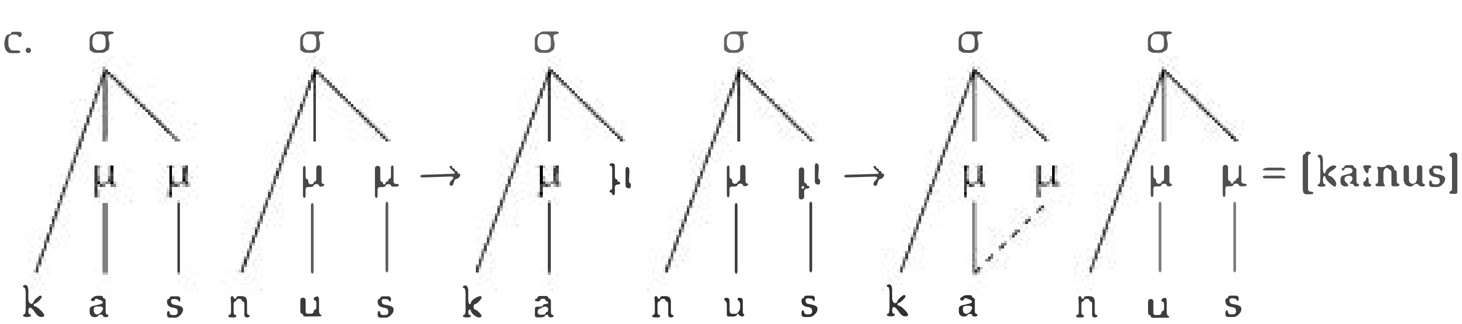
a. */s/-deletion*

$$s \rightarrow \varnothing \ / \ \_\_ \begin{bmatrix} +\text{son} \\ +\text{ant} \end{bmatrix}$$ (segmental tier only)

b. *Compensatory lengthening*



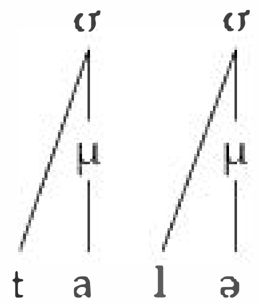
where μ' is a segmentally unaffiliated mora

c. 

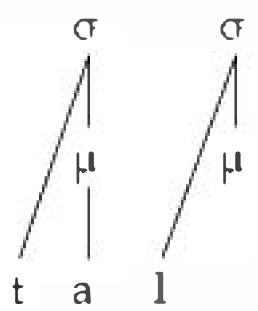
A theory assuming a prosodic frame defined in terms of X-slots can account for the example above, but not for the fact that in the same language (Latin), /s/-deletion does not trigger lengthening when it is word-initial, as in *snurus* > *nurus* (the same problem holds for the type of compensatory lengthening illustrated in §1.2). In a segmental theory based on X-slots, any deleted segment should trigger lengthening, whereas in a moraic theory only those segments that are mora bearing will do so.

The moraic theory of compensatory lengthening accounts for CVCV cases as illustrated in (14) with Middle English [talə] > [taːl] 'tale' (see Minkova 1982 for an in-depth discussion of this case).

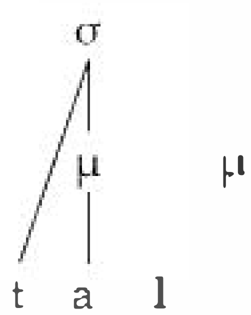
(14) *Compensatory lengthening in* CVCV *sequences* (Hayes 1989: 268–269)

a. 

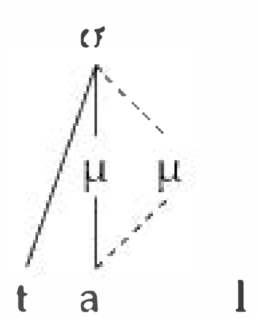
Input

b. 

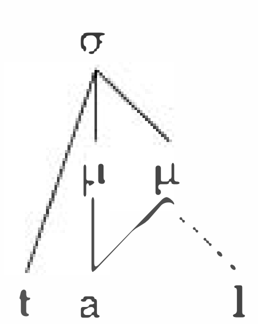
Schwa drop

c. 

Parasitic delinking

d. 

Compensatory lengthening

e. 

Resyllabification

= [taːl]

Parasitic delinking, illustrated in (14c), is a principle that eliminates ill-formed syllable structure, caused in this case by the loss of the vowel segment via schwa drop.

A very positive aspect of Hayes (1989) is that a wide range of cases of compensatory lengthening are discussed (although a potentially problematic empirical gap is discussed below). Besides the classic CVC and CVCV cases, Hayes (1989) treats the so-called "double flop" cases in which the deletion of a glide triggers compensatory lengthening of a vowel in a preceding syllable, as in Ancient Greek *odwos > /oːdos/ (1989: 265–266), and compensatory lengthening from glide formation, as in the Ilokano case mentioned earlier /ˈluto+en/ → /lutt'w-en/. He also mentions (without providing a formal treatment) "straightforward" cases like compensatory lengthening through progressive and regressive total assimilation of consonants, compensatory lengthening through prenasalization, and so-called "inverse compensatory lengthening," which involves the lengthening of a consonant triggered by the shortening or loss of a vowel (1989: 279–281).

Fox (2000) points out a number of what he sees as problems for Hayes's (1989) approach. The first is that the principle of parasitic delinking is "a radical measure which is not required in most other processes of Compensatory Lengthening." The second relates to the required linking of the vowel of the first syllable to the mora stranded by deletion of the second syllable. According to Fox (2000: 100–102), this is:

> unmotivated by the normal principles of the model, since, according to one view at least, the syllable would be perfectly well-formed without this linking; the final mora would be linked to the final consonant and is thus not left stranded.

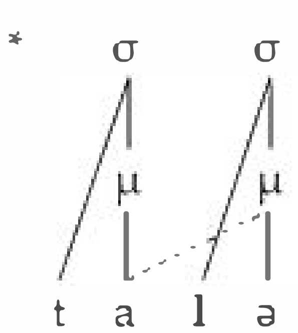
Finally, Fox (2000) suggests that Hayes's (1989) principle of mora conservation is inappropriate as a motivation for CVCV compensatory lengthening. This is because Hayes defines the mora as "the basic unit for syllable weight" (1989: 285) and syllable weight is not maintained in these cases. Rather, what is maintained is the length of the foot (2000: 101).

One might also argue that a problem for phonological conservation approaches generally is that they are ill equipped to deal with the gradual nature of diachronic compensatory lengthening. For example, the Middle English case discussed above did not take place as a discrete change in one fell swoop. According to Minkova (1982: 50):

> Before becoming identified with existing long vowels or developing into new ones, i.e. prior to the establishment of a phonological length contrast, the short vowels in the environment / __ C_1e# undergo phonetic lengthening. [. . .] In a situation where forms with and without the second syllabic element, the *-e*, are both available to the speaker, there will be a negative correlation between it and the first syllabic element. Phonetically, "the word as a whole has a certain duration that tends to remain relatively constant." (Lehiste 1970: 40)

One way to show intermediate length in moraic phonology is to have segments share a mora. In this case, a standard interpretation of the formalism prevents this because it would involve the crossing of association lines, as illustrated in (15).

(15) *Potential inadequacy of Hayes's mora conservation approach with respect to modeling gradual change:* CVCV *cases*



If one adopts a strict and unnuanced view of the ban on crossed association lines, since parasitic delinking (illustrated in (14c)) is triggered only when ill-formed syllable structure is present, the moraic account of compensatory lengthening is only able to succeed if the final vowel is entirely deleted and parasitic delinking applies. That is, the account is unable to account for the allophonic lengthening that must be assumed to precede phonemic lengthening. However, a more relaxed view might interpret the ban on crossed association lines as applying separately to distinct C and V tiers.

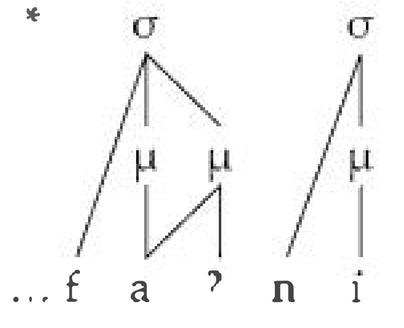
Gradual change might also be seen as a problem for the mora conservation approach in diachronic cases involving CVC compensatory lengthening. Hock (1986) mentions a case reported in Brockelman (1908) from Tunisian Arabic (with similar instances in Ge'ez and Tigrinya), in which a preconsonantal glottal stop is reduced (not deleted), with compensatory lengthening on the preceding vowel. The examples provided are shown in (16).

(16) *Compensatory lengthening triggered by segmental reduction in Tunisian Arabic* (Hock 1986: 444)

ʃeffaʔni > ʃeffaːˀni
smaʔtkum > smaːˀtkum

In autosegmental representation, the output of this process would be as illustrated in (17).

(17) *Potential inadequacy of Hayes's mora conservation approach with respect to modeling gradual change:* CVC *cases*



While the shared mora representation in (17) is adequate for representing an intermediate stage between a fully moraic glottal segment following the short vowel and a long vowel with no following glottal, one might argue that it cannot represent any more than a single such stage, whereas more stages might well be warranted. However, this potential criticism disregards the possibility of a single phonemic representation having different phonetic interpretations at different periods (or indeed across speakers at a single period).

We must also note (as have others) one apparent empirical weakness with the phonological conservation approach, as put forth in Hayes (1989). This involves cases in which compensatory lengthening is triggered by a prevocalic consonant that in normal circumstances would not be associated with a mora. Some such cases are discussed in Hock (1986), a paper cited by Hayes (1989), but without mention of these specific examples. Strangely, the cases are also problematic for Hock, although he does not treat them as such. Hock's interest in the cases is that they involve compensatory lengthening triggered not by deletion of a segment, but by its weakening only (as in the case illustrated in (17), but in intervocalic position). The first case is from Tyrone Irish (as discussed in Stockman and Wagner 1965), where:

> vowels are dialectally distinctively lengthened before the highly reduced glottal-fricative outcome of earlier voiceless fricatives (as well as before sonorant + consonant etc. and in "ordinary" CL environments, *but not in open syllables*). (Hock 1986: 443–444, emphasis added – RG)

The fact that Hock takes care to note that compensatory lengthening does not take place in open syllables suggests that the reduced segments in question might be ambisyllabic, but this is not made explicit anywhere, including in the original source. If it is the case that the segments in question are ambisyllabic, then an argument could be made that they are doubly linked – to a mora in the first syllable and to the onset of the second. The relevant data are shown in (18).

(18) *Compensatory lengthening triggered by segmental reduction in Tyrone Irish* (Hock 1986: 444)

srathar	[straːʰər]
tachas	[tɔːʰəs]

The second problematic case mentioned by Hock (1986) is from the Westphalian dialect of Soest, as reported in Holthausen (1886: 28–29). In this dialect, as illustrated in (19), compensatory lengthening occurs before highly reduced, voiced labial and velar fricatives, and before a deleted "secondary" (analogically reinserted) voiced alveolar stop.

(19) *Compensatory lengthening triggered by segmental reduction in Westphalian* (Hock 1986: 444)

*hege	'hedge'	>	(*)hiəgə	>	hiːəɣə	
*seven	'seven'	>	(*)siəv(e)n	>	siːəvn	
*snede	'slice'	>	(*)sniəde	>	sniːə	(*with* əə > ə)

Again, there is no suggestion that the consonants triggering compensatory lengthening in these cases were ambisyllabic, nor is there any reason to believe that they were. This case therefore represents an apparent problem for Hayes's (1989) mora conservation approach, aside from the issue of the gradient nature of the triggering segmental reduction. (The case also poses a problem for Hock 1986, which represents a mora conservation approach as well, although one not couched in an autosegmental framework.) As mentioned earlier, other cases of compensatory lengthening triggered by the loss of intervocalic consonants are discussed by Beltzung (2008: especially ch. 2).

Finally, the phonological conservation approach *à la* Hayes (1989), particularly in the case of CVC compensatory lengthening, has proven difficult to model in Optimality Theory (OT). The basic problem is that in order for lengthening to occur, consonants must be assigned weight before deletion happens, thus suggesting a serial analysis. Getting around this problem has necessitated the abandonment of some of the basic tenets of OT. For example, one could simply assume that consonants are moraic in the input (Sprouse 1997), but this requires a sidestepping of OT's principle of Richness of the Base, whereby output well-formedness is determined solely by constraints and their ranking, and not by restrictions on input. Other ways of handling the problem involve treatments designed to handle opacity more generally, such as stratal OT (Kiparsky 2000), which rejects strict serialism, Turbidity Theory (Goldrick 2001), Sympathy Theory (McCarthy 2003), or OT with candidate chains (McCarthy 2005; Shaw 2009), which require reference to what amounts to one or more intermediate representations.

2.3 *Non-conservation approach*

In an influential article, de Chene and Anderson (1979) take a novel approach to compensatory lengthening by rejecting the notion that such a process exists as "an independent mechanism of phonetic change" (1979: 505) (they discuss only cases of CVC compensatory lengthening). For them, putative cases of compensatory lengthening can be decomposed into two independent processes: weakening of the consonant in question to a glide, and subsequent monophthongization of the resulting vowel + glide sequence. De Chene and Anderson further contend that monophthongization will result in a long vowel only if the language in question has a pre-existing vowel length distinction. The latter claim is about a structure-preserving condition on compensatory lengthening and not about the process itself. Discussion of it is explored further in §3. For now, let us look in more detail at the first claim.

The proposal that cases labeled compensatory lengthening are in fact the result of two unrelated processes has generated much discussion. Gess (1998) points out that it has been challenged by numerous scholars, including Hock (1986), Poser (1986), Sezer (1986), and Gildea (1995). According to Gess (1998: 353), "[E]ach of these scholars provides a strong case against the view that compensatory lengthening is always decomposable into two distinct stages. The ensemble of their arguments renders this claim simply untenable." Without laboring the point, then, we simply illustrate de Chene and Anderson's hypothesis with one straightforward example. According to de Chene and Anderson (1979: 512):

> In Latin, compensatory lengthening involving loss of a dental spirant is limited in source to *Vz[C, +dent] sequences, where *z is the reconstructed allophone of *s before a voiced segment. Thus we have *ni-sd-o > *nīdus* 'nest' and *si-sd-ō > *sīdō* 'I sit down', both involving the zero grade of **sed* 'to sit' (cf. *sedeō* 'I sit').

De Chene and Anderson continue: "Our posited intermediate development involves the loss of occlusion in (preconsonantal) *[z], leading to the voiced glottal spirant [ɦ]" (1979: 512). In this type of analysis, de Chene and Anderson were not in fact alone. Jeffers and Lehiste (1979) propose the analysis in (20) for the remarkably similar change from Proto-Indo-European (PIE) *nisdo to Sanskrit /niːḍa-/.

(20) *Jeffers and Lehiste's analysis of* PIE *nisdo > *Sanskrit* /niːḍa-/ (as presented in Hock 1986: 435)

nisd-	
nizd-	voicing assimilation
nizḍ-	retroflexion
niẓḍ-	retroflex assimilation
nijḍ-	gliding
niːḍ-	contraction

In noting the similarity in analyses, Hock (1986: 435) points out that what distinguishes Jeffers and Lehiste's analysis from de Chene and Anderson's is the fact that the former is "not proposed as explanations for *all* cases traditionally labeled loss with compensatory lengthening, but only for a certain subset, however poorly that subset may be defined." (Note that Kavitskaya 2002: 38 incorrectly interprets this sentence as referring to de Chene and Anderson 1979, rather than to Jeffers and Lehiste 1979.)

Hock (1986: 435) continues by saying that he is:

> ready to concede that many instances of what has traditionally been called loss with compensatory lengthening may well be ambiguous, and can be analyzed either as weakening-cum-assimilation or as cases of loss-with-mora-retention. However, in light of the fact that there are [. . .] cases of loss with CL which cannot be explained in terms of weakening-cum-assimilation, any theory which recognizes only the latter process must be considered insufficient.

One of the examples provided by Hock that is not amenable to explanation under de Chene and Anderson's non-conservation approach is from Icelandic, and is provided in (21).

(21) *Compensatory lengthening in Icelandic* (Hock 1986: 442)

*liugan	'lie'	>	lju:ga
*keosan	'choose'	>	kjo:sa
(*)priar	'three-FEM'	>	prja:r
(*)se:an	'see'	>	sja:

In this case, which is similar to the Ganda case at the end of §1.4, although there is weakening to a glide, it involves a preceding vowel rather than a following consonant, and there is no monophthongization involved.

Hock's (1986: 435) reproach of de Chene and Anderson (1979) for proposing an alternative explanation of all cases traditionally called compensatory lengthening while neglecting to treat all types extends to other cases as well. For example, while de Chene and Anderson are aware of CVCV compensatory lengthening, they choose not to discuss it (1979: 506, n. 1).

2.4 *Kavitskaya (2002)*

Kavitskaya (2002) puts forth a model of compensatory lengthening that can be considered a radical departure from previous treatments, in that it assumes the process to be entirely listener-oriented (see CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). In this respect, her model is representative of the overall approach to phonological change espoused in Blevins (2004, 2006), Evolutionary Phonology. This is a model that rejects any explanations for historical phenomena that involve the synchronic phonologies of speakers (e.g. by assuming a role for phonological rules or markedness constraints) when there is an alternative, diachronic explanation available. This essentially removes the speaker from the story of phonological change, except as a source of variation from which potential changes may or may not take root through "innocent misperception" on the part of the listener. This variation is constrained by

> speaker-specific anatomical differences, and within the speech of a given speaker, due to phonetic transforms of speech dependent (at least) on: rate of speech; degree of physical effort involved; and the humanly physical impossibility of making exactly the same sound twice. (Blevins 2006: 125–126)

According to Kavitskaya's (2002) listener-oriented account of compensatory lengthening:

> diachronic CL through consonant loss [CVC > CV:] ultimately has its origin in the phonetic lengthening of vowels in the environment of neighboring consonants; the subsequent loss of a consonant conditioning such length causes the length to be re-analyzed as phonological. (Kavitskaya 2002: 8)

Further, according to Kavitskaya, with respect to diachronic compensatory lengthening through vowel loss [CVCV > CV:C]:

> Prior to the deletion of the final vowel, the longer vowel duration characteristic of open syllables is correctly parsed by listeners as a phonetic consequence of syllable structure in the first syllable of a CVCV sequence, and is discounted [. . .] Upon deletion of the final vowel, however, the duration of the vowel in the newly-closed syllable becomes inexplicable, since it is longer than is expected in the closed syllable. (Kavitskaya 2002: 9)

If Kavitskaya's arguments are right, then compensatory lengthening is not really compensatory in nature. For the process to be compensatory, the compensatory aspect would rely on a role for the speaker, as is assumed at least implicitly in all other models of compensatory lengthening.

In so far as Kavitskaya (2002) is representative of the Evolutionary Phonology framework proposed in Blevins (2004, 2006), it is susceptible to the general criticisms that have been leveled against that framework. Lindblom (2006) criticizes the Evolutionary Phonology framework for its reliance on so-called "extra-phonological" explanations over phonological accounts. According to Blevins (2006: 20), "principled extra-phonological explanations for sound patterns have priority over competing phonological explanations unless independent evidence demonstrates that a purely phonological account is warranted." Lindblom takes exception to this stance on the grounds that it highlights a "phonetics/phonology split" and traps the framework in "the conceptual prison of the form/substance distinction" (2006: 242).

As the title of his response to Blevins (2006) declares very loudly, Lindblom (2006) rejects the phonetics/phonology split. Lindblom admonishes us to:

> Deduce sound structure from language use. Anchor theory construction in the universal conditions under which all speech communication must take place. Start from "first principles" and not circularly from the data to be explained (cf. "markedness"). At the level of the individual user, model *phonological structure*, not as autonomous form, but as *an emergent organization of phonetic substance* acquired by each native speaker in the context of socially shared, ambient knowledge. At the population level, model this knowledge as a use- & user-dependent process that undergoes change along the historical time scale. Get rid of the distinction between "phonological" and "extraphonological." Here is a key step: *Make the "intrinsic content" an integral part of the theory from scratch*. Treat "intrinsic content" as the source that helps generate discrete structure and that constrains both synchronic and diachronic phonological patterning. (2006: 243)

In §4, we will explore how a rejection of the phonetics/phonology split might be helpful in accounting for the many types of compensatory lengthening as a unified phenomenon. We conclude this section by looking at an empirical challenge to Kavitskaya's listener-oriented approach – a synchronic case of compensatory lengthening that suggests an explanation in terms of speaker-controlled behavior.

McRobbie-Utasi (1999) provides evidence for a synchronic case of compensatory lengthening that is apparently speaker-controlled and that suggests the relevance of a principle of isochrony in a synchronic production grammar. In an acoustic analysis of quantity in Skolt Saami, McRobbie-Utasi shows a clear connection between the distribution of duration in V1, intervening C, and V2 sequences in disyllabic groups, and a phonological process realized as "an optional rule that either reduces word-final short vowels or deletes them" (1999: 111). Deletion of the final vowel is a feature of casual speech. The relevant optional rule is shown in (22).

(22) *Word-final vowel deletion in Skolt Saami* (McRobbie-Utasi 1999: 111)

V --> Ø / __ #

Vowel deletion rule
A word-final vowel is optionally deleted in Type 1–5 disyllabics.

It is important to note that in the V1, intervening C, and V2 sequences, the intervening C can be long and ambisyllabic (in four of the five types mentioned), or short and affiliated as the onset of the second syllable.

The principal relevant passage from McRobbie-Utasi (1999) is shown below, where the "stress-group locations" referred to are the V1 and following C in the relevant sequences. V2 constitutes a third "stress-group location." According to McRobbie-Utasi (1999: 114–115):

> From the [. . .] measurements an important tendency can be deduced: namely, that the presence or absence (or reduced duration) of the vowel in the second syllable has clear consequences for the distribution of duration in the first syllabic vowel and the consonant(s) following it. Thus, an increase in duration takes place as a result of compensatory lengthening. It will be recalled that second syllabic vowel durations were constant in all the structural types once they were realized as full vowels, with an average of 87 msec [. . .]; also, that durations signaling differences between the structural types and/or gradation types are manifested in the first syllabic vowel and the consonant(s) following it. The fact that the presence or absence of the second syllabic has a considerable effect on these durational distributions in the segments preceding has important implications. The durational changes noticeable in these two stress-group locations (i.e. first syllabic vowel and the consonant(s) following) must be recognized as exemplifying the phenomenon of compensatory lengthening. The absence or reduced status of the second syllabic vowel results in an increase of duration in both of the stress-group locations referred to above.

Compensatory lengthening in Skolt Saami, triggered by the reduction or deletion of a final vowel, affects both the preceding vowel and consonant in four of five types (those in which the consonant is long and ambisyllabic), and the lengthening that occurs does so in a way that precisely preserves the overall V/C ratio. In the remaining type, in which the consonant is short and syllabified as the onset of the second syllable, "reduced duration of the second syllabic vowel results in compensatory lengthening in the first syllabic *vowel* only. There is practically no durational increase in the consonant following this vowel" (McRobbie-Utasi 1999: 118).

It is difficult, although perhaps not impossible, to reconcile McRobbie-Utasi's (1999) findings with a listener-based approach. Although McRobbie-Utasi's study involved only two speakers of Skolt Saami, their behavior with respect to the 550 test words used (recorded three times by both speakers, for a total of 3079 usable tokens) was remarkably consistent. Nor do the types of sequences involved lend themselves readily to Kavitskaya's line of explanation for CVCV compensatory lengthening, since they do not involve (except for Type 3) phonetically lengthened vowels in open syllables. (As expected, V1 in Type 3 sequences is longer than in other types, both when V2 is fully realized and when it is not.) Nor has any reanalysis occurred (whatever that might look like given the sequences involved and their variety [five types]) since the trigger for compensatory lengthening is still synchronically recoverable. Rather, it appears that the speakers are guided directly or indirectly by a principle of isochrony with respect to the disyllabic group.

Other empirical problems for Kavitskaya's approach, from historical French (manifesting types 1A and 1C), are discussed in Gess (forthcoming).

3 A putative constraint on compensatory lengthening

This section briefly explores the second claim made by de Chene and Anderson (1979): that compensatory lengthening can only occur in a language with a pre-existing vowel length contrast – i.e. that it is strictly structure preserving (CHAPTER 76: STRUCTURE PRESERVATION: THE RESILIENCE OF DISTINCTIVE INFORMATION). This issue is discussed in detail in Gess (1998), which treats the very data from Old French on which de Chene and Anderson base their claim, thus adding a particularly severe blow to a claim already questioned in other work (for example, in Hock 1986, Hayes 1989, Morin 1994, and Lin 1997, as well as two further cases discussed more recently in Beltzung 2008: 20–21).

According to de Chene and Anderson (1979: 517), "a necessary condition for the development of contrastively long vowels through monophthongization is the independent existence of a length contrast in the language." With respect to historical French, de Chene and Anderson compare two distinct processes (in the ninth and sixteenth centuries) of monophthongization of the diphthong [aw]. At the earlier stage, the resulting monophthong [o] was short. However, at the later stage, the outcome was the long vowel [oː]. (Strangely, de Chene and Anderson (1979: 519) also suggest a sixteenth-century date for the loss of preconsonantal [l] – the same century in which they contend that monophthongization of the vowel + glide sequence resulting from its loss had occurred. However, Gess (1999) provides strong evidence for a much earlier date for the loss of syllable-final [l], after the latter part of the eleventh century – and many scholars assume a much earlier date still.)

The difference in outcomes in the monophthongization of derived [aw] was due, according to de Chene and Anderson, to the introduction of vowel length into the language via the loss of intervocalic consonants, in the late ninth and early tenth centuries (1979: 521). This introduction of vowel length also allowed for compensatory lengthening, according to de Chene and Anderson, following the loss (through an intermediate stage as a glide) of syllable-final [z], [s], and nasals.

Loss of the latter is incorrectly dated by de Chene and Anderson in the sixteenth century, while loss of the former, [z] and [s], is dated in the twelfth and thirteenth centuries. According to Gess (1999), loss of nasal consonants dates from the thirteenth century, and loss of [z] and [s] dates from the eleventh to the thirteenth centuries.

De Chene and Anderson (1979: 522) make the following claim with respect to the establishment of long vowels in Old French:

> There is a solid body of long vowels, however, that were established by 1100 through deletion of the consonant in original V_iCV_i sequences. In these cases, no leveling or assimilation being necessary, a long vowel is the automatic result of loss of the consonant.

They go on to provide a list of several words illustrating the relevant consonant loss and the resulting putative long vowels. However, Gess (1998) found each of the forms listed by de Chene and Anderson in twelfth- and thirteenth-century Old French poetry and, in each case, the forms are clearly treated as consisting of two syllables. Gess (1998: 358) "found many other examples of orthographic geminate vowels in 12th and 13th-century Old French poetry, all of which are treated as bisyllabic."

The fact that sequences of two vowels were still counted as bisyllabic in the thirteenth century, when the loss of [z] at the very least had occurred, with compensatory lengthening, shows that a pre-existing vowel length contrast in the language was not a prerequisite for compensatory lengthening to take place. Rather, Hayes's (1989) assumption is likely the right one, that a syllable weight distinction in the language in question is necessary and, crucially, sufficient for compensatory lengthening to take place. Gess (1998: 364) points out that from an optimality-theoretic perspective this would be a rather unsurprising consequence of the general principle of minimal violation, in this case of faithfulness to the input. While a given constraint ranking may allow for the erosion of segmental features, it may still protect prosodic structure.

4 Assessment and recent directions

§2 outlined various approaches to compensatory lengthening: a phonetic conservation approach, a phonological conservation approach, a non-conservation approach, and a listener-based approach. We saw that the non-conservation approach, proposed only in the context of CVC compensatory lengthening, is basically untenable, both because it fails to account for any other type of compensatory lengthening and because there are instances of CVC compensatory lengthening that appear not to be decomposable into the stages suggested by de Chene and Anderson (1979). This leaves us with two conservation approaches, both suggestive of a speaker-based process, and a listener-based approach.

We have noted problems with each of these approaches, which I will summarize briefly here. We have observed that the phonetic conservation approach proposed in Timberlake (1983) is most relevant to those instances of compensatory lengthening that are gradient in nature and that may be characterized as changes in progress. This approach seems ill suited for dealing with synchronic cases of compensatory lengthening that involve complete loss of the trigger. Without some refinement, the approach also has difficulty with CVC cases of compensatory lengthening given the assumption implicit in its formalism that moras associated with consonants are of equal duration to those associated with vowels.

With respect to synchrony/diachrony dichotomy, the phonological conservation approach suffers the opposite problem from the phonetic conservation approach. That is, while it can account for most, if not all, cases of synchronic compensatory lengthening, it is not ideally suited to account for compensatory lengthening as a gradual diachronic process or as a process that involves a trigger that is only reduced and not entirely lost. Other criticisms of the phonological conservation approach relate to formalisms (e.g. motivation for parasitic delinking in a rule-based approach and required mechanisms for dealing with opacity in an OT approach) and empirical weakness (its inability to account for compensatory lengthening involving non-moraic segments).

The listener-based approach is explicitly an historical approach. That is, it aims to account only for the diachronic development of compensatory lengthening and does not attempt to model synchronic instantiations of the process. For synchronic cases, it is compatible with (but also limited by) the formalisms required by the phonological conservation approach so long as the process involves a trigger that is synchronically recoverable (see Kavitskaya 2002: ch. 5). The principal problem with the listener-based approach is its inability to account for cases that do not lend themselves to re-analysis via misperception. One such case is the synchronic process in Skolt Saami, which involves a prevocalic VC complex as the target, with lengthening affecting the complex as a whole, and with VC ratios precisely maintained. Also problematic for the listener-based approach is left-to-right CVCV compensatory lengthening (1D), which is why it is so important for Kavitskaya to dismiss such cases as instances of non-compensatory, rhythmic lengthening. Further research is necessary to shed light on this particular issue.

The listener-based approach will also have difficulty accounting for left-to-right compensatory lengthening processes in which both target and trigger are consonants (2A). Note that this difficulty will obtain for an Evolutionary Phonology inspired approach whether one labels the process as compensatory lengthening or as total assimilation since the second consonant in such sequences will normally be perceptually stronger. Further challenges for the listener-based approach come from the compensatory lengthening of consonants triggered by either a following vowel (2C) or a preceding one (2D). Whether the approach can be developed sufficiently to meet these challenges will be interesting to see.

It may be worthwhile for future work to explore the compatibility of the different approaches to compensatory lengthening. Note, for example, that a phonetically based speaker-oriented analysis neither denies a role for the listener, nor necessarily discounts phonological (e.g. moraic) structure. Timberlake (1983), for example, makes clear reference to the mora as the "full duration" of a vowel. Hock (1986: 432), who does not cite Timberlake (1983), also points out that "at least some traditional historical linguists have offered a phonetic explanation of CL in terms of the concept 'mora'." Note that mora-based phonological approaches do not necessarily deny a gradual, phonetic aspect either, at least those prior to autosegmentalist accounts. Minkova (1982) provides a clear phonological account of compensatory lengthening in Middle English, based on syllables and rhythmic units (metrical feet) "described [. . .] with reference to their phonological/moric composition" (1982: 48), but is careful to "complete the picture by adding some considerations of purely phonetic nature" (1982: 50). Minkova also touts as an advantage of her revised environment for Middle English Open Syllable Lengthening the fact that "it is the only way in which the shift from allophonic to phonemic length of the stressed vowel can be accounted for" (1982: 51).

Hock (1986: 434) cites the "striking extent" to which historical evidence coincides with "fine-phonetic" experimental data. Indeed, Hock goes even further (1986: 445), citing the apparent fact "that CL may set in *before* the complete loss of a segment, simply as the result of TC [temporary compensation] for the *reduction* of the segment" (emphasis in the original). An important consequence of this is that:

> the situation just described requires an important modification of the notion "mora": Rather than referring to a temporal unit measureable in terms of segment length, it must – at least for CL – be permitted to refer to time spans which are fractions of ordinary segment length. (Hock 1986: 445)

This view of the "mora" is entirely in keeping with an analysis along the lines of the one proposed in Timberlake (1983). It is also in keeping with the spirit of Lindblom's (2006) view of phonological structure as non-autonomous and emergent from phonetic substance, at least if we assume both that the mora is an abstract temporal unit *and* that reference is permitted to time spans that may be fractions thereof.

A rejection of the phonetics/phonology split with respect to the mora may be the only way to achieve real explanatory adequacy with respect to compensatory lengthening. It allows us to explain the gradual nature of compensatory lengthening – a clearly phonetic aspect of the process. On the phonological side, it also accounts for the fact that CVC compensatory lengthening tends to occur mostly in languages with moraic consonants and for the fact that the process in general functions to preserve moraic structure. In conceptual terms, it is the phonological status of the mora, as an abstract unit of weight functioning in the grammar, that provides the motivation for preserving it when an associated segment is subject to gradual reduction and eventual elimination. On the other hand, it is the physical timing associated with moraic elements that guides the actual articulatory implementation of reduction with concomitant compensatory lengthening, a process that is gradual (and variable) in nature. Since all segments have physical timing associated with them, the only (unsurprising) assumption we have to make is that preservation of timing associated with weight-bearing units is generally privileged over the preservation of timing associated with units that do not bear weight.

Recent work by Topintzi (2006) and Beltzung (2008) demonstrates the continuing relevance of compensatory lengthening for phonological theorizing. It has directly tackled the problems that compensatory lengthening poses for OT and manages to maintain the basic tenets of the framework. Both pieces of work must be categorized as phonological conservation approaches, both seek to expand the empirical coverage of previous approaches (notably to account for non-moraic consonant triggers of compensatory vowel lengthening), and, interestingly, both demonstrate the need for formal appeal to the preservation of segment positions in addition to moras. Both appear compatible, therefore, with a rejection of the phonetics/phonology split as described above (whether or not the authors themselves agree with such a rejection). Both also appear compatible with a potential complementary, phonetically based OT approach that might focus on the functional motivation for and phonetic implementation of non-categorical reduction with concomitant lengthening, as it has for assimilation (Jun 1995) and lenition (Kirchner 1998, 2004; Gess 2003, 2004, 2009) (again, whether or not the authors themselves agree with such a move, which is not part of the OT orthodoxy). It seems safe to conclude that compensatory lengthening will continue to be a topic of some interest in the phonological community.

REFERENCES

Abbott, Miriam. 1991. Macushi. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.) *Handbook of Amazonian languages*, vol. 3, 23–160. Berlin & New York: Mouton de Gruyter.

Beltzung, Jean-Marc. 2008. L'allongement compensatoire dans les representations phonologiques: Nature, contraintes et typologie. Ph.D. dissertation, Université Paris III, Sorbonne-Nouvelle.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32. 117–166.

Brockelmann, Carl. 1908. *Grundriss der vergleichenden Grammatik der semitischen Sprachen*, vol. 1. Hildesheim: Georg Olm.

Carson, Neusa M. 1981. Phonology and morphosyntax of Macuxi (Carib). Ph.D. dissertation, University of Kansas.

Chene, Brent de & Stephen R. Anderson. 1979. Compensatory lengthening. *Language* 55. 505–535.

Clements, G. N. 1986. Compensatory lengthening and consonant gemination in LuGanda. In Wetzels & Sezer (1986), 37–77.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Fox, Anthony. 2000. *Prosodic features and prosodic structure: The phonology of suprasegmentals*. Oxford: Oxford University Press.

Gess, Randlall. 1998. Compensatory lengthening and structure preservation revisited. *Phonology* 15. 353–366.

Gess, Randall. 1999. Rethinking the dating of Old French syllable-final consonant loss. *Diachronica* 16. 261–296.

Gess, Randall. 2003. On re-ranking and explanatory adequacy in a constraint-based theory of phonological change. In D. Eric Holt (ed.) *Optimality Theory and language change*, 67–90. Dordrecht: Kluwer.

Gess, Randall. 2004. Phonetics, phonology and phonological change in Optimality Theory: Another look at the reduction of three-consonant sequences in Late Latin. *Probus* 16. 21–41.

Gess, Randall. 2009. Reductive sound change and the perception/production interface. *Canadian Journal of Linguistics* 54. 229–253.

Gess, Randall. Forthcoming. Compensatory lengthening in Old French: The role of the speaker. In Deborah Arteaga (ed.) *Old French: The state of the research*. Dordrecht: Kluwer.

Gildea, Spike. 1995. A comparative description of syllable reduction in the Cariban language family. *International Journal of American Linguistics* 61. 62–102.

Goldrick, Matthew. 2001. Turbid output representations and the unity of opacity. *Papers from the Annual Meeting of the North East Linguistic Society* 30(1). 231–245.

Gordon, Matthew. 1999. Syllable weight: phonetics, phonology, and typology. Ph.D. dissertation, University of California, Los Angeles.

Hawkins, W. Neil. 1950. Patterns of vowel loss in Macushi (Carib). *International Journal of American Linguistics* 16. 87–90.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce & Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9. 47–96.

Hock, Hans Henrich. 1986. Compensatory lengthening: In defense of the concept "mora." *Folia Linguistica* 20. 431–460.

Holthausen, Ferdinand. 1886. *Die Soester Mundart*. Norden & Leipzig: Soltau.

Hyman, Larry M. 1984. On the weightlessness of syllable onsets. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 10. 1–14.

Hyman, Larry M. 1985. *A theory of phonological weight*. Dordrecht: Foris.

Jeffers, Robert J. & Ilse Lehiste. 1979. *Principles and methods for historical linguistics*. Cambridge, MA: MIT Press.

Jensen, John T. 1977. *Yapese reference grammar*. Honolulu: University Press of Hawaii.

Jun, Jongho. 1995. Perceptual and articulatory factors in place assimilation: An Optimality Theoretic approach. Ph.D. dissertation, University of California, Los Angeles.

Kager, René. 1997. Rhythmic vowel deletion in Optimality Theory. In Iggy Roca (ed.) *Derivations and constraints in phonology*, 463–499. Oxford: Clarendon.

Kálmán, Bela. 1972. Hungarian historical phonology. In Loránd Benkő & Samu Imre (eds.) *The Hungarian language*, 49–83. The Hague: Mouton.

Katsanis, Nikolaos. 1996. *Το γλωσσικό ιδίωμα της Σαμοθράκης* [The dialect of Samothraki Greek]. Δημος Σαμοθράκης [Municipality of Samothraki].

Kavitskaya, Darya. 2002. *Compensatory lengthening: Phonetics, phonology, diachrony*. London & New York: Routledge.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–365.

Kirchner, Robert. 1998. An effort-based approach to consonant lenition. Ph.D. dissertation, University of California, Los Angeles. Published 2001. New York & London: Routledge.

Kirchner, Robert. 2004. Consonant lenition. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 313–345. Cambridge: Cambridge University Press.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, Ilse. 1977. Isochrony reconsidered. *Journal of Phonetics* 5. 253–263.

Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.

Lin, Yen-Hwei. 1997. Syllabic and moraic structures in Piro. *Phonology* 14. 403–436.

Lindblom, Björn. 2006. Rejecting the phonetics/phonology split. *Theoretical Linguistics* 32. 237–243.

Lipiński, Edward. 2001. *Semitic languages: Outline of a comparative grammar*. 2nd edn. Leuven: Peeters.

Lowenstamm, Jean & Jonathan Kaye. 1986. Compensatory lengthening in Tiberian Hebrew. In Wetzels & Sezer (1986), 97–132.

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.

McCarthy, John J. 2003. Sympathy, cumulativity, and the Duke-of-York gambit. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*, 23–76. Cambridge: Cambridge University Press.

McCarthy, John J. 2005. Candidate chains. Paper presented at the 2nd Old World Conference on Phonology, University of Tromsø.

McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.

McRobbie-Utasi, Zita. 1999. *Quantity in the Skolt (Lappish) Saami language: An acoustic analysis*. Bloomington: Indiana University.

Minkova, Donka. 1982. The environment for open syllable lengthening in Middle English. *Folia Linguistica Historica* 3. 29–58.
Morin, Yves-Charles. 1994. Phonological interpretations of historical lengthening. In Wolfgang U. Dressler, Martin Prinzhorn & John R. Rennison (eds.) *Phonologica 1992*, 135–155. Turin: Rosenberg & Selier.
Pessoa, Katia N. 2006. Fonologia Taurepang e comparação preliminar da fonologia de línguas do grupo Pemóng (família Caribe). M.A. thesis, Universidade Federal de Pernambuco.
Pessoa, Katia N. 2009. O accento ritmico na língua Taurepang (família Karib). *Guavira* 9. 114–126.
Pope, Mildred K. 1952. *From Latin to Modern French with especial consideration of Anglo-Norman: Phonology and morphology*. Manchester: Manchester University Press.
Poser, William J. 1986. Japanese evidence bearing on the compensatory lengthening controversy. In Wetzels & Sezer (1986), 167–186.
Sezer, Engin. 1986. An autosegmental analysis of compensatory lengthening in Turkish. In Wetzels & Sezer (1986), 227–250.
Shaw, Jason. 2009. Compensatory lengthening via mora preservation in OT-CC: Theory and predictions. *Papers from the Annual Meeting of the North East Linguistic Society* 38. 323–336.
Shishkov, Peter. 2002. Elision of unstressed vowels in the Erkeč dialect. Retrieved from http://escholarship.org/uc/item/7tm865v7.
Sprouse, Ronald. 1997. A case for enriched inputs. Ms. (ROA-193.)
Stockman, Gerard & Heinrich Wagner. 1965. Contributions to a study of Tyrone Irish. *Lochlann* 3. 43–236.
Timberlake, Alan. 1983. Compensatory lengthening in Slavic 2: Phonetic reconstruction. In M. S. Flier (ed.) *American Contributions to the 9th International Congress of Slavists*, vol. 1: *Linguistics*. 293–319.
Topintzi, Nina. 2006. A (not so) paradoxical instance of compensatory lengthening. *Journal of Greek Linguistics* 7. 71–119.
Wetzels, W. Leo. 1986. Phonological timing in Ancient Greek. In Wetzels & Sezer (1986), 279–344.
Wetzels, W. Leo & Engin Sezer (eds.) 1986. *Studies in compensatory lengthening*. Dordrecht: Foris.

65 Consonant Mutation

Janet Grijzenhout

1 Introduction

The phenomenon of "consonant mutation" occurs in a wide array of unrelated languages and comprises changes that are also known as "consonant weakening" (or *lenition*), "consonant strengthening" (or *fortition*), and "nasalization." In this chapter, "consonant mutation" will be defined as a change in one phonetic property of a consonant that affects its degree of sonority and that does not depend on the position of the consonant within a prosodic domain (i.e. neutralization and enhancement phenomena are excluded), nor on the position immediately adjacent to a segment with which it forms a natural class (i.e. progressive and regressive voicing and place assimilations are not regarded as instances of "consonant mutations"). More specifically, the term "consonant mutation" refers to a class of processes by which a consonant turns into a segment with a different degree of voicing, continuancy, or nasality that is not due to neutralization or assimilation to a neighboring segment of the same natural class.[1]

Some types of consonant mutation can be described as alternations that take place in a particular phonological environment; for instance, an oral stop may turn into a fricative between a sonorant consonant and a vowel. Other types of consonant mutation take place in a certain morphological or lexical context; for example, stem-initial oral stops are realized as continuants under certain morphosyntactic conditions, while continuants are deleted or realized as a laryngeal sound under the same conditions in Modern Irish (e.g. Ní Chiosáin 1991; CHAPTER 117: CELTIC MUTATIONS). The interesting aspect of consonant mutation in general is a diachronic one: what starts out as a purely phonological alternation induced by neighboring segments may gradually turn into a morphological alternation for which the phonological context is no longer transparent (CHAPTER 93: SOUND CHANGE). In the course of this chapter, we will encounter various examples of such developments.

[1] Note that changes in voicing, continuancy, or nasality make a consonant either more or less sonorant. Consonant mutation processes thus have in common that they alter a consonant's degree of sonority.

Typically, mutations are "scalar." In the languages of the world, we find consonant mutations where a consonant's degree of stricture decreases (e.g. in Archaic Irish and Finnish an underlying geminate stop is realized as a singleton stop, while an underlying singleton stop is realized as a continuant sound) or increases. Another example of a scalar mutation is one in which a consonant's degree of laryngeal specificity and/or nasality increases (e.g. in Old Irish, an aspirated voiceless stop is realized as an unaspirated one in the same context in which an unaspirated oral stop is realized as a prenasalized stop).

This chapter first discusses possible consonant alternations in more detail (§2). As examples of languages that have relatively many types of consonant mutations (i.e. spirantization, gemination, nasalization, and/or prenasalization), we discuss Southern Paiute (§3) and Fula (§4). Balto-Finnic, Sami, and some Australian languages show scalar mutations (§5). §6 points out the merits and drawbacks of some theoretical accounts of consonant mutation that exist in phonological literature. §7 summarizes the discussion.

2 Consonant alternations within prosodic and morphological domains

Consonants are highly adaptable elements that may change their properties for a variety of reasons. This section will focus on some phonological and morphological environments that may trigger a change in one phonetic property of a consonant. We start with consonant alternations that are characterized by the fact that a phonological opposition is neutralized in a certain prosodic environment, viz. final devoicing and debuccalization. We will also briefly consider consonant alternations that occur at the left edge of a prosodic domain. Next, five consonant alternations that are not triggered merely by a prosodic environment (i.e. that are independent of the position within a prosodic domain) and that fall under the rubric of "consonant mutation" are introduced: (a) stopping (Soninke), (b) obstruent voicing (Burmese), (c) spirantization or fricativization (Djapu; the first stage of Grimm's Law), (d) devoicing (the second stage of Grimm's Law), and (e) deaspiration (the third stage of Grimm's Law). Other types of consonant mutation that are frequently encountered in languages – both diachronically and synchronically – are gemination, degemination, and (pre)nasalization. The latter phenomena will be discussed in later sections.

Many consonant alternations are characterized by the fact that a phonological opposition is neutralized in a certain prosodic environment. In a variety of unrelated languages, e.g. Catalan, Czech, Dutch, German, Ojibwa, Polish, Russian, Turkish, and Wolof, we find that the opposition between voiced and voiceless obstruents is neutralized in one particular environment only, viz. at the end of a prosodic domain (usually the syllable; see Brockhaus 1995 and CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). In other languages, place of articulation contrasts are neutralized at the end of a prosodic domain – usually the syllable coda – and this phenomenon is known as "debuccalization" (CHAPTER 80: MERGERS AND NEUTRALIZATION). In some generative frameworks, alternations at the end of a prosodic domain are described as processes where consonants "lose" their underlying marked specification for laryngeal features or place of articulation features.[2] Conversely, in other positions of the word, consonants may become reinforced phonetically, e.g. by initial aspiration (which is seen as a form of enhancement, for instance, by Keyser and Stevens 2006: 42ff.). Consonant alternations that involve neutralization of an opposition in a particular prosodic context (e.g. the contrast between voiced and voiceless obstruents in Dutch and German is neutralized at the end of a prosodic domain), or that involve adding a feature to enhance an opposition in a particular prosodic domain (e.g. the contrast between voiced and voiceless stops is enhanced by adding aspiration for the voiceless plosives at the left word edge in English) do not fall under the category of "consonant mutations" as understood here.

In the examples in (1), stem-final consonants optionally change their laryngeal properties to become more similar to the neighboring consonants within a phonological phrase (e.g. Berendsen 1983 for Dutch) and in (2) the consonant /n/ changes its place of articulation to the same place of articulation as the following obstruent (CHAPTER 81: LOCAL ASSIMILATION).

(1) *Regressive laryngeal assimilation in obstruents*

a. *Dutch*

zeep + doos	→	*zee*[bd]*oos*	soap + box	'soap-box'
kas + boek	→	*ka*[zb]*oek*	cash + book	'cash-book'
zak + doek	→	*za*[gd]*oek*	pocket + cloth	'handkerchief'

b. *Hungarian*

zseb + kendo	→	*zse*[pk]*endõ*	pocket + cloth	'handkerchief'

(2) *Regressive place assimilation of* n- *(marker of noun classes* 9 *and* 10*) in Kisukuma* (data from Batibo 2000: 169)

n + buli	→	mbuli	'goat'
n + dama	→	ndama	'calf'
n + guzu	→	ŋguzu	'strength, energy'

Cases where consonants change a laryngeal or place feature under the influence of an adjacent consonant within a certain prosodic domain are most commonly referred to as "assimilations" rather than "mutations," and these processes are relatively easy to describe in theoretical frameworks, e.g. in autosegmental theory as spreading of laryngeal or place features or in Optimality Theory (OT) as phenomena that are the result of ranking AGREE[feature] and *[αfeature] constraints higher than the corresponding IDENT[feature] constraint (e.g. Lombardi 1996). The reason for assimilation is not to increase or decrease the degree of sonority of a segment, but rather to become "more similar" with respect to laryngeal or place properties to an immediately adjacent obstruent (in the Dutch or Hungarian examples) or a stop (in Kisukuma).

[2] Note that the assumption about laryngeal neutralization being a case of "weakening" in the sense that the consonant "loses" an underlying feature is highly controversial, as can be seen, for example, by the German terminology *Auslautverhärtung* ("final hardening") for syllable-final devoicing. Foley (1970), for instance, claims that a change from voiced to voiceless obstruent should be considered a case of strengthening ("fortition") rather than weakening ("lenition"). For further discussion on the issue of what exactly constitutes strengthening or weakening, I refer to CHAPTER 66: LENITION.

The set of data below illustrates a process whereby the oral stricture of initial consonants of nouns increases, i.e. voiceless fricatives or voiced continuants are realized as oral stops (3a), /l/ (3b), or nasal stops (3c) after nasal segments:[3]

(3) *Consonant alternations after nasals in Soninke*

a.	fare	'donkey'	n pare	'my donkey'
	si	'horse'	n ʧi	'my horse'
	xore	'charcoal'	n gore	'my charcoal'
b.	raqqe	'mouth'	n laqqe	'my mouth'
c.	wulle	'dog'	n ŋulle	'my dog'
	jaaxe	'eye'	n ɲaaxe	'my eye'

In the examples in (3), all continuant consonants are affected and the trigger of the change is always the same, i.e. a nasal element. Intuitively, we may formulate the consonant alternation as a process that increases a consonant's oral stricture after a nasal stop or nasalized vowel. Nevertheless, it is not easy to describe this process as one in which a single feature of the stem-initial consonant changes under the influence of a preceding nasal vowel or stop.[4] Furthermore, some nouns (especially names) "resist" change and do not participate in the consonant alternation process. Moreover, under some morphological conditions, the nasal trigger is absent, but the change takes place nonetheless, for instance in some imperative forms, e.g. /pagu/ 'fill up!' and /ʧi/ 'shave!' (cf. /si/ 'to shave'). According to Kendall and Bird (1982), the language is thus in a transitional stage in which the phonologically triggered process of consonant change has developed into a process that is no longer purely phonological; there are exceptions as well as overapplications, i.e. cases where the consonantal change takes place without an overt phonological trigger.

We next consider another type of consonant alternation where the phonological context determines the shape of a consonant: in Burmese, voiceless stops are voiced in intervocalic position (4a), (4b) or following a nasal (4c).

(4) *Intersonorant stop gradation in Burmese* (data from Campbell 1995: 98–102)

a.	θwa + pa	→	θwàba	'please go'
b.	θwa + tɔ + lu	→	θwàdɔ̀lu	'the man who is going'
c.	kauŋ + kauŋ	→	kauŋgauŋ	'to be good'

[3] Soninke is a Mende language spoken in West Africa. All Soninke data presented in this chapter are from Kendall and Bird (1982). The nasals in (3a) assimilate in place of articulation to the following segment. Kendall and Bird (1982: 1, 3) state that the same consonant alternation occurs both after nasal consonants and after nasalized vowels, e.g. /r/ in /ri/ 'to come' is realized as /l/ after a nasalized vowel in /nili/ 'I came'.

[4] In *SPE* (Chomsky and Halle 1968), we could formulate a rewrite rule [+continuant] → [–continuant] / [+nasal] __, with some additional rules to account for the fact that voiced continuants turn into a lateral or nasal stop. In autosegmental phonology, there is no single feature that nasal consonants and vowels have in common that could spread unto the following consonant. In OT, it is possible to formulate a constraint that bans continuant consonants after a nasal consonant or vowel (e.g. *[+nas][+cont, –voc]), but this constraint would be an ad hoc one and leaves open the possibility of other ad hoc constraints such as *[+strid][–cont] or *[+nas][lab], etc.

In the Australian language Djapu, stops are only realized as such in lexical words after obstruents and nasal stops. Thus the dative suffix /-ku/ appears as such in [buurut̪ʲ-ku] 'mosquito'; when preceded by a vowel or a liquid, however, the labial and velar stops become a labial-velar glide (e.g. [ŋajmil-wu] 'Ngaymil clan') and the dental stop becomes a palatal glide (Morphy 1983).

Cases where consonants change in an intersonorant environment are most often referred to as "gradations" or "mutations." When glides become fricatives, when underlying continuants becomes non-continuant segments (as in Soninke), or when singleton stops are geminated, the mutation is often referred to as "consonant hardening" or "fortition." If the mutated consonant increases its degree of sonority, this type of mutation is often referred to as "consonant weakening" or "lenition." The environment for gradation in Soninke can be characterized as "after a nasal"; gradation in Burmese is "in intervocalic position or between a nasal stop and a vowel," whereas the environment for gradation in Djapu is "in intervocalic position or between a liquid and a vowel." CHAPTER 66: LENITION provides more examples of intervocalic voicing and spirantization.

The interesting problem that consonant mutations pose for phonological theory is that they change the degree of sonority of a segment (CHAPTER 49: SONORITY) and that they are not easily accounted for by means of autosegmental processes such as spreading, inserting, or deleting a phonological feature or class of features within a natural phonological context (CHAPTER 14: AUTOSEGMENTS). Neither is it easy to account for them by means of well-formedness constraints that correctly predict surface forms in non-mutation contexts and the corresponding alternating forms in mutation contexts. Consider in this respect that a possible constraint that penalizes intersonorant voiceless stops could be *[+son][−son, −voice][+son] (i.e. no voiceless obstruent in between two sonorants). This constraint would be violated in a potential output [θwàpa]. If the markedness constraint in question is ranked higher than the faithfulness constraint IDENT[voice] – which, presumably would be the case in Burmese – the output [θwàba] would win. Now consider the fact that the winning candidate in Burmese would be a form that is disallowed in Spanish (see e.g. Harris 1969). Whereas Spanish allows intervocalic voiceless stops – suggesting that *[+son][−son, −voice][+son] is low-ranked – it does not have output forms with intervocalic voiced stops (resembling the winning candidates for Burmese). In an optimality-theoretical framework, we could again propose a constraint, e.g. *[+son][−son, −cont, +voice][+son] (i.e. no voiced stop between two sonorants), which would have to be ranked higher than a faithfulness constraint, e.g. IDENT[−son, −cont] to generate the correct output for Spanish. Apart from the fact that there is an ad hoc flavor to the OT accounts suggested immediately above, a further complicating factor is the fact that some languages exhibit both consonant alternations in their grammars. In Northern Corsican, for example, voiceless stops become voiced where voiced stops spirantize. Moreover, we find exactly the same consonant alternations in contexts that cannot be described in a straightforward way as being "inter-sonorant." Rather, as will be shown in subsequent sections, the same consonant alternations as described here appear in different morphological contexts or are lexicalized in various unrelated languages.

The cases presented so far all reflect synchronic processes. The interest in consonant mutations, however, first arose with respect to diachronic changes such as the first consonant shift in West Germanic (also known as Grimm's Law) and early Celtic consonant mutations (Pedersen 1897; Thurneysen 1898).[5]

Grimm's Law is often formulated as follows: the Indo-European stops /p t k kʷ/ spirantized and became fricatives, the unaspirated stops /b d g gʷ/ became voiceless aspirated stops, and so-called "breathy voiced" stops /bʰ dʰ gʰ gʷʰ/ were replaced by voiced unaspirated stops (CHAPTER 73: CHAIN SHIFTS). Iverson and Salmons's (1995) account of the first shift in Grimm's Law runs as follows. The voiceless aspirated stops became voiceless fricatives when aspiration was audible (i.e. only when the stop was released, so that stops following /s/, geminates and stop–stop clusters did not undergo the shift). The unaspirated stops – in which voicing was optional – had a "stronger and longer" closure phase, which made them unlikely candidates for fricativization. Thus, the outcome of the first shift in Grimm's Law is a system in which we find fricatives (specified for [stiff vocal folds]), unaspirated stops (unspecified for laryngeal features), and voiced aspirated stops (specified for [slack vocal folds, spread glottis]). The second stage of Grimm's Law is the process whereby the contrast between stops unmarked for laryngeal features (/b d g gʷ/) and stops marked as "slack with aspiration" (/bʰ dʰ gʰ gʷʰ/) is increased by introducing [stiff vocal folds] (which is available already in the fricative series) for the unmarked stops. In the system where [stiff vocal folds] stops contrast with [slack vocal folds] ones, there is no need to maintain the aspiration contrast and hence the feature [spread glottis] for breathy voiced stops gradually loses its place in the stop system. The three subsequent mutations may thus be characterized as follows: (a) fricativization of voiceless stops (/p t k kʷ/ → /f θ x xʷ/), (b) fortition or devoicing of lax stops (/b d g gʷ/ → /p t k kʷ/), and (c) deaspiration of aspirated lax stops (/bʰ dʰ gʰ gʷʰ/ → /b d g gʷ/).

This section has introduced five consonant alternations that affect the degree of sonority and that fall under the rubric of "mutations": stopping, obstruent voicing, spirantization or fricativization, devoicing, and deaspiration. Other mutations that are frequently encountered in languages – both diachronically and synchronically – are gemination, degemination, and (pre)nasalization. In the following sections we will discuss informative examples of synchronic consonant mutations. We start with Southern Paiute, a language that exhibits spirantization, gemination, and nasalization.

3 Southern Paiute consonant mutations

Sapir (1930) describes the Shoshonean dialect Kaibab Paiute, as it was spoken in southwestern Utah and northwestern Arizona during the 1910s. In this dialect, consonants can appear in one form when in suffix-initial position after a consonant or in word-initial position, and in various alternating forms when – by the process of derivation or compounding – they are immediately preceded by a vowel. The particular process that a suffix undergoes is dependent on the lexical item it is attached to and is not phonologically predictable in any obvious way.

[5] For a discussion on the history and present-day exponents of Celtic consonant mutations the reader is referred to CHAPTER 117: CELTIC MUTATIONS.

(5) *Southern Paiute consonant mutations* (Sapir 1930: 62)[6]

underlying	*spirantized*	*geminated*	*nasalized*
p	β	p:	mp
t	ɹ	t:	nt
k	ɣ	k:	ŋk
kʷ	ɣʷ	kʷ:	ŋkʷ
ts/ʧ	ts/ʧ	ts:/ʧ:	nts/nʧ
s/ʃ		s:/ʃ:	
m	ŋʷ	m:	m:
n		n:	n:

Under spirantization, oral stops change into voiced continuants, affricates do not change, and the labial nasal becomes a "back palatal" labialized nasal consonant (represented as /ŋʷ/ by Sapir). Harms (1966) attempts to describe Paiute spirantization as a regular phonological rule ([+consonantal, –vocalic, –strident] → [+continuant, +voice] / [–consonantal, +vocalic, +voice] __), but such a rule only accounts for part of the alternation, and leaves the change from labial nasal to dorsal labialized nasal unexplained. Before we turn to another analysis of Southern Paiute consonant mutations, we first introduce some examples.

To illustrate the effect of mutation, Sapir (1930: 63, 67) mentions the verbalizing suffix /-ka/ and the agentive suffix /-pi/, with their respective mutated initial consonants following adjectival (6a)–(6c) and nominal stems (6d)–(6e):

(6) *Consonant mutations affecting morpheme-initial /k/ and /p/ in Southern Paiute*

a.	aŋka	+	-ka	→	aŋkaɣa	'to be red'	*spirantization*
b.	kụʧa	+	-ka	→	kụʧak:a	'to be gray'	*gemination*
c.	paɨ	+	-ka	→	paɨŋka	'to be smooth'	*prenasalization*
d.	nɔ	+	-pi	→	nɔvi	'carrier'	*spirantization*
e.	taŋa	+	-pi	→	taŋampi	'kicker'	*prenasalization*

Note that the stem /aŋka/ 'red' triggers spirantization in a following suffix (6a), but is followed by a geminated consonant in compounds (7):

(7) *Consonant mutations affecting stem-initial /k/ and /p/ in Southern Paiute compounds*

aŋka	+	kani	→	aŋkak:anị	red + house	'red house'
aŋka	+	pajɨ	→	aŋkap:aji	red + fish	'trout'

As a possible explanation for the difference in the choice of consonant mutation between derivation and compounding in some nouns, Sapir (1930: 70) suggests that "the tendency to use geminated consonants in composition is probably due to the greater phonetic similarity thus brought about between a simplex and its compound."

[6] The consonants and the consonant alternations mentioned by Sapir (1930) are represented here by means of the corresponding IPA symbols. According to Sapir, spirantized consonants following voiced vowels are voiced, and those following voiceless vowels are voiceless. I present the voiced allophones only. The glides /w/ and /j/ and the nasal /ŋ/ do not occur in initial positions and are hence not affected by mutations. The glottal stop does not undergo mutations. The three blanks in (5) indicate that [s/ʃ] and [n] do not have a spirantized counterpart and that [s/ʃ] lacks a nasalized form.

McLaughlin (1984: 70–71) notes that prefixes are followed by one fixed mutation, i.e. all prefixes affect the initial consonant of a following stem, and Southern Paiute distinguishes "spirantizing," "geminating," and "nasalizing" prefixes. In compounds, noun stems may trigger one of three mutations, but in most cases nouns in compounds trigger gemination only. In verbal compounds, the verb of the second member has its initial consonant geminated in the majority of cases. To McLaughlin, these facts suggest that prefixes show the strongest reflection of an earlier stage in the language, viz. one in which a prefix-final phoneme triggered a change in the following consonant. The phoneme in question is lost, but the mutation that it triggered is still in effect. Nouns also show this effect, but are in a transient stage; nominal stems are cautiously on their way to induce one particular mutation only, i.e. gemination. Verbs have developed even further and generally geminate a following consonant.

To account for Southern Paiute consonant mutations, McLaughlin (1984) proposes that all prefixes, most noun stems, some adjectival stems, and a few verb stems end either in (a) a vowel, (b) an unspecified stop C, or (c) an unspecified nasal N.[7] The last two are reflexes of final stops and nasals that used to be present in earlier stages of the language. If a stem is preceded by a prefix or another stem ending in a vowel, its first consonant will undergo a rule of spirantization. A suffix or stem following an unspecified stop or nasal, however, will undergo a place assimilation rule (e.g. /taŋaN + pi/ → [taŋampi]). Finally, stems following an unspecified stop will also be subject to a lengthening rule, so that the result will be a geminate consonant. Another rule may subsequently degeminate the stop in question when it follows an unstressed vowel.

McLaughlin's analysis of spirantization thus involves a regular spirantization rule, which turns stops into voiced spirants in inter-sonorant position. The analysis of gemination and nasalization, on the other hand, depends on "ghost" elements, i.e. a final stop or nasal that never surfaces, but has a geminating or nasalizing effect if another consonant follows in the next morpheme (i.e. a suffix in the case of derivation and a stem in the case of compounding). These processes are thus lexical in the sense that in the lexicon, the morphemes in question (prefixes, most noun stems, and some adjectival stems) end either in a vowel or in an unspecified stop or nasal. The mutation triggered by verb stems is becoming more and more morphologized; i.e. morphological leveling is producing ever larger numbers of geminating verb stems.

Since oral stops become continuants under spirantization, we would perhaps expect the labial nasal stop to turn into a labial approximant [w] (CHAPTER 28: THE REPRESENTATION OF FRICATIVES), as is the case in Celtic (see CHAPTER 117: CELTIC MUTATIONS). The question arises why this is not the case. With respect to the quirky character of spirantized /m/, McCarthy and Prince (1995: 349ff.) make the following suggestion. They first observe that the segments [w] and [ŋʷ] are in complementary distribution: the former is found word initially and the latter occurs post-vocalically.[8] This distribution follows from the following

[7] For similar proposals to account for Celtic consonant mutation, see CHAPTER 117: CELTIC MUTATIONS.

[8] The only exception is the context of reduplication, where "w" surfaces between vowels due to an identity constraint that says that a base and a reduplicated form cannot have different values for the feature [nasal]. The fact that [ŋʷ] is not allowed to occur in word-initial positions (*]ŋʷ) rules out a candidate like hypothetical *[ŋʷa-ŋʷaxipija]. In *[wa-ŋʷaxipija] the constraint IDENT-BR[nasal] is violated and the candidate [wa-waxipija] is the winner even though it has a "w" after a vowel (i.e. the constraints mentioned in this footnote are ranked as follows: *]ŋʷ, IDENT-BR[nasal] >> *VwV).

ranking of constraints: *VwV >> *ŋʷ >> *w. Even though this ranking accounts for the fact that we are more likely to find [ŋʷ] rather than [w] in intervocalic position, it is not immediately clear why /m/ alternates with [ŋʷ] in the first place, i.e. why /m/ surfaces as a nasal with a "back palatal" place of articulation in a spirantization context. If spirantization means that oral or nasal stops are realized as continuants, and if there is an operation or constraint (*VwV) that prevents the nasal /m/ from being realized as the corresponding continuant [w], it is still not evident that /m/ undergoes a transformation whereby it turns into a labialized dorsal nasal stop. I will leave this issue for further research.

4 Fula consonant mutation

Fula (also known as Fulani, Fulbe, Fulfulde, Pular, Pulaar, and Peul) is spoken in West Africa; the majority of the speakers live in Nigeria (Campbell 1995: 178). In all dialects of Fula, initial consonants of verbal radicals and nominal, numeral and verbo-nominal stems can have different forms. In contrast to Southern Paiute – where stem-initial consonants change their form depending on the mutating requirements of the preceding prefix or stem – one cannot argue that in Fula a segment has a different surface form depending on a preceding vowel or a morpheme specified for a mutating feature. Arnott (1970) shows that in the Gombe dialect of Fula, stem-initial consonants surface as homorganic stops, continuants or prenasals, depending on the adjective or noun class.

(8) *Fula consonant mutations* (based on Arnott 1970: 42–43)[9]

stop	*spirant*	*prenasal*
p	f	p
b	w	mb
d	r	nd
ʃ[10]	s	ʃ
ɟ	j	ɲɟ
k	h	k
g	j/w[11]	ŋg

According to Arnott (1970), adjectives and nouns are marked as belonging to one of 25 possible classes. Class membership is indicated by a suffix[12] and a particular manner of articulation of the initial consonant of the stem: 11 classes are marked by a stem-initial voiceless or voiced stop, six classes are marked by

[9] The consonant alternations mentioned by Arnott are represented here by means of the corresponding IPA symbols. Note that the glottalized consonants /ɓ ɗ ƴ/, the nasals /m n ɲ ŋ/, and the coronal consonants /t c l/ are "invariable" and do not alternate.

[10] Note that this sound is a fricative, but in the mutation system it functions as a stop. In most analyses of Fula consonant mutations that can be found in the literature, this sound is usually transcribed as the palatal stop /c/, but this does not tally with Arnott's (1970) description of this sound.

[11] The palatal glide is found before the front vowels /i e/ and the labial-velar one before the back vowels /u o a/.

[12] Suffixes belong to certain grades, indicated as grades A, B, C, or D in Arnott (1970). In the phrase /gude daneeje/ 'white cloths', both the noun and the adjective belong to class 24 (i.e. one of the classes that are marked by an initial stop), but the noun /gude/ 'cloths' has a grade A suffix /-e/ and the adjective /daneeje/ 'white' has a grade B suffix /-je/.

an initial spirant and eight are marked by an initial voiceless oral stop or a prenasalized stop. The marker of noun class 1 (singular nouns referring to persons), for instance, is a suffix ending in the vowel /o/ and a stem-initial stop. The marker of noun class 2 (plural nouns referring to persons) is a spirant-initial stem consonant and the suffix /-ɓe/. The marker of class 7 is a suffix ending in the vowel /a/ and a stem-initial oral voiceless stop or prenasalized stop for consonants that alternate with a voiced stop in class 1:

(9) *Stem-initial consonant alternations in Fula nouns* (data from Arnott 1970: 98–99)

class 1	*class 2*	*class 7*	
pull-o	ful-ɓe	pul-a	'Fula'
beer-o	weer-ɓe	mbeer-a	'host'
dim-o	rim-ɓe	ndim-a	'free man'
ʃook-o	sook-ɓe	ʃook-a	'poor man'
ɟuul-ɗo	juul-ɓe	ɲɟuul-ŋga	'Moslim'
kor-ɗo	hor-ɓe	kor-ga	'female slave'
gim-ɗo	jim-ɓe	ŋgim-ŋga	'person'

In an autosegmental framework that assumes radical underspecification (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION), Wiswall (1989) proposes that Fula voiceless continuants are underlyingly specified for [−voice], but not for manner of articulation. The voiceless stops are specified for [−voice] as well, and all underlying oral stops have a specification for [−continuant]. Underlying non-alternating nasal stops have a specification for [+sonorant] and [+nasal]. Wiswall furthermore suggests that noun class markers in Fula have floating features that associate to the initial consonant of the stem (CHAPTER 82: FEATURAL AFFIXES); the stop classes have a floating [−continuant] feature (which is problematical for the sound /ʃ/), the prenasal classes have a floating [+nasal] feature (which presumably cannot dock onto sounds underlyingly specified as being [−voice]), but the spirant classes do not have a floating feature. After association of the floating features, redundancy rules (e.g. [−continuant] → [−sonorant]), and default rules (Ø → [+continuant]) fill in the unspecified feature values. Grijzenhout (1991) points out some problems relating to the rule ordering (CHAPTER 74: RULE ORDERING) that Wiswall has to assume, especially related to the account of prenasals, which would involve a late counterintuitive redundancy rule [+nasal] → [−sonorant].

For an alternative account of the data involving the theory of "charm and government," the reader is referred to work by Paradis (e.g. Paradis 1987a, 1987b, 1992). Elzinga (1996) provides an OT account of Fula consonant mutation, which makes use of alignment and parsing constraints on the mutating features [continuant] and [nasal]. He furthermore employs morpheme constraints, i.e. constraints that indicate the type of stem ("invariable," "partially variable," or "fully variable"). The morpheme constraints are placed in between the feature alignment and parsing constraints, so that invariable stems and suffixes are ranked higher in the hierarchy than the mutation constraints, partially variable stems and suffiixes are ranked below some mutation constraints and above others, and fully variable stems and suffiixes are ranked lower than the mutation constraints.

We now turn to cases of scalar mutations, i.e. mutations that cause a change from one underlying consonant to another consonant, which also mutates into a different one.

5 Balto-Finnic and Sami consonant gradation

Most languages belonging to the Balto-Finnic group (i.e. Finnish, Estonian, Votic, Ingrian, and Karelian; excluding Livonian and Veps), as well as northern and eastern dialects of the closely related Sami group, exhibit consonant gradations. It is striking that a similar phenomenon, by which long stops were shortened and short stops were turned into spirants, occurred in the history of Iwaidjan languages in Australia (e.g. Evans 1998). The phonological condition that generated consonant gradation in languages belonging to the Balto-Finnic group was originally the following: after a vowel or sonorant consonant in a stressed syllable, stops in the so-called "strong grade" that appeared in the onset of a syllable that was closed by certain inflectional or derivational endings were mutated such that they appeared in the corresponding "weak grade." Under this condition, underlying long (or "geminate") stops were reduced to short (or "singleton") stops and underlying short stops were replaced by voiced and fricated consonantal variants. Thus, in the Balto-Finnic languages, long /pː/ alternated with short /p/ while short /p/ alternated with /b/ or /v/, and long /tː/ alternated with short /t/ while short /t/ alternated with /d/ or /ð/. Similarly, in the history of Iwaidjan languages, long intervocalic stops were shortened and short intervocalic stops became approximants or liquids (e.g. /pː/ → /p/ → /w/ and /cː/ → /c/ → /j/).

Different varieties of Sami and Estonian now have a three-way opposition between the strongest grade or quantity, the strong grade, and the weak grade, but the phonemes involved in the alternations are somewhat different for each language or language variety. We most often find that in contexts where gradation applies in these languages, (a) "overlong stops" (usually written as <pp tt kk>, corresponding to the sounds /pːː tːː kːː/) are realized as "long stops," (b) "long stops" (orthographic <p t k>, corresponding to the geminates /pː tː kː/) are realized as "short" consonants, and (c) underlyingly "short stops" (i.e. those with the shortest closure duration, usually written as <b d g>, corresponding to unaspirated /p, t, k/) spirantize.

Consonant gradation is more or less regular in most dialects of Sami. In many Sami dialects, consonant gradation was extended to consonants that had not been subject to gradation in earlier stages of the language and now also affects sonorant consonants; see (10d). Gradation applies when an affix closes an unstressed syllable. Moreover, gradation still occurs in genitive forms – see (10c) and (10d) – even though the original inflectional ending [-n] has been lost in some variations of Sami, so that the final syllable is no longer closed:

(10) *Sami consonant gradation* (Gordon 1998; Campbell 2004: 322)[13]

	non-gradated		*gradated*	
a.	bapppa	'priest (NOM SG)'	bappast	(ELAT SG)
b.	loppe	'permission (NOM SG)'	lobest	(ELAT SG)
c.	jokkâ	'river (NOM SG)'	jogâ	(GEN SG)
d.	guolle	'fish (NOM SG)'	guole	(GEN SG)

[13] Recall that orthographic <b g> in examples (10b) and (10c) correspond to short unaspirated /p k/.

Interestingly, in some Sami languages, the original gradation process now also works in reverse; single consonants are geminated in open syllables:

(11) *Sami "reversed" consonant gradation; i.e. gemination* (cf. Uralic languages 2010)

	geminated			*gradated*	
a.	ʧuotte	*ʧuote	'hundred (NOM SG)'	ʧuoðe	'hundred (GEN SG)'
b.	borra		'eats'	borâın	'I eat'

Thus, by "reversed gradation," the contrast between two related forms is enhanced: instead of a short stop–fricative contrast, we now have a geminate–fricative contrast in (11a).

The examples below illustrate the phenomenon of consonant gradation for Modern Finnish (e.g. Karttunen 1970; Skousen 1972; Keyser and Kiparsky 1984; Vainikka 1988); geminate stops degeminate (12a) in the same environment in which the singleton stops lessen their degree of stricture and become continuants (12b),[14] assimilate to a preceding sonorant consonant with the same place features as the stop in question (12c), or are not realized (12d):[15]

(12) *Finnish consonant gradation in closed syllables*

	underlying form (nominative)	*gradation (genitive)*	
a.	*lappu*	*lapun*	'piece of paper'
	matto	*maton*	'rug'
	kukka	*kukan*	'flower'
b.	*tapa*	*tavan*	'custom'
	mato	*madon*	'worm'
c.	*rampa*	*ramman*	'lame'
	lintu	*linnun*	'bird'
d.	*poika*	*pojan*	'boy'
	selkä	*selän*	'back'

As is the case in some variations of Sami – see (10c), (10d), and (11a) above – the original genitive ending /-n/ has been lost in Estonian, but gradation still occurs in genitive forms:

(13) *Estonian consonant gradation in genitive nouns* (Harms 1962)

	nominative	*genitive*	
a.	*leib*	*leiva*	'bread'
b.	*madu*	*mao*	'snake'
c.	*lind*	*linnu*	'bird'
d.	*selg*	*selja*	'back'

[14] Finnish <v> represents an approximant; <d> is a dental sonorous element whose value varies from dialect to dialect; see e.g. Vainikka (1988). Here and in what follows, I abstract away from dialectal variation.

[15] Finnish <j> represents /j/; see e.g. Vainikka (1988).

We again witness how a phonologically triggered mutation gradually changes into a morphologically triggered mutation. The original consonantal morphological ending caused the syllable in question to become closed and thus provided the phonological environment for gradation to take place. When the consonantal ending disappeared in the course of history, gradation still took place. In present-day Estonian, consonant gradation is thus triggered by morphology rather than phonology.

The type of consonant gradation presented in this section is not unique to Balto-Finnic languages and varieties of Sami. Similar processes are found in some languages belonging to the Samoyedic branch of the Uralic language family (in particular Nganasan and Selkup; see CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE for examples and an OT account of consonant gradation in Nganasan).

6 Mutation and phonological representations

Our first examples of consonant mutation – i.e. stopping after nasal vowels and consonants in Soninke, inter-sonorant voicing in Burmese, and inter-sonorant spirantization in Djapu – have been accounted for in phonological literature as cases where one feature of a preceding segment (e.g. [–continuant] in the case of nasal stops, [+voice] in the case of vowels and sonorant consonants, and [+continuant] in the case of vowels) spreads to the target consonant.[16] The first drawback of such an analysis is the obvious fact that spreading of [±continuant] is hardly attested in obstruents (CHAPTER 13: THE STRICTURE FEATURES); cases where an oral stop induces a change in a preceding or following fricative (e.g. hypothetical /t + fare/ → [tpare]) are hardly – if ever – attested. Moreover, even though we find relatively many cases where vowels cause spirantization of stops, fricatives never cause spirantization of adjacent stops (see Wetzels 1991), and this makes the suggestion that intervocalic spirantization is analyzed best as a case of spreading the feature [+continuant] problematic (CHAPTER 28: THE REPRESENTATION OF FRICATIVES).

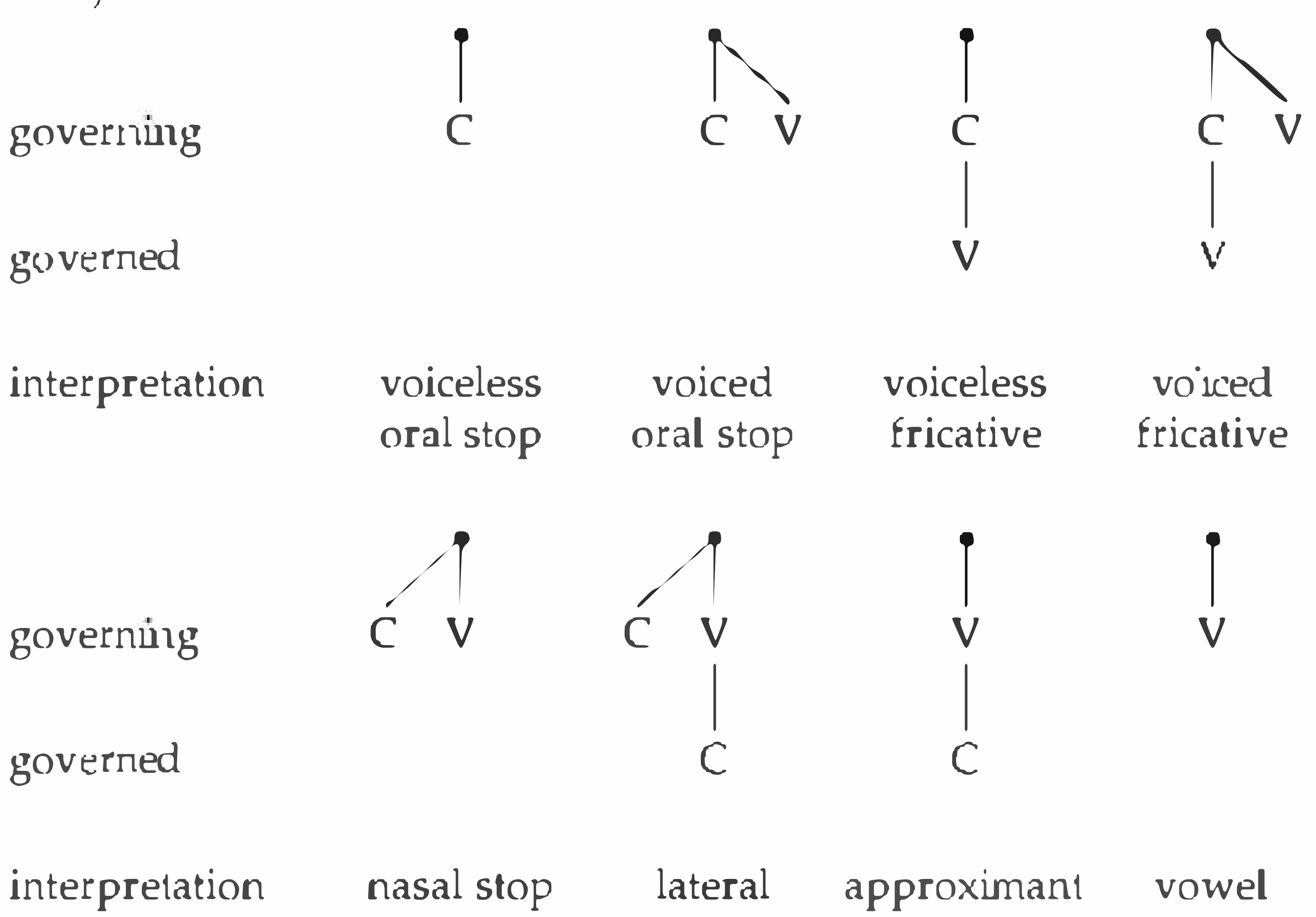
Furthermore, a spreading analysis is problematic for cases where there is no phonological segment that triggers the mutation. In those cases, e.g. Southern Paiute gemination, McLaughlin (1984) and others have suggested underlying underspecified segments that usually do not surface and only affect a following segment in mutation contexts, resulting in, for instance, a geminated consonant. Others, e.g. Wiswall (1989), propose "floating" features such as [+nasal] to account for initial prenasalization in, for instance, Fula (CHAPTER 82: FEATURAL AFFIXES). Mutation, then, is viewed as the change of one sound into another sound due to feature insertion; it is unpredictable which feature will cause a change under which circumstances in which language. The problem with this particular approach is the existence of the "scalar" mutations witnessed in Northern Corsican, Finnish, or Sami, for instance, which cannot be accounted for by means of insertion of a

[16] Thus, in these accounts, strengthening and weakening are treated as local processes where the target sound is adjacent to another sound that spreads one of its features to the target. Cf. Harris (1994), who proposes an account where an element expressing full closure or aperiodic energy (as for aspiration) spreads from one position to another in the case of strengthening, and where an element is delinked or deleted under lenition, so that the result is a "weaker" – i.e. less complex – segment.

single "floating" feature. Scalar mutations are changes in underlying consonants that result in more sonorant segments (e.g. voiceless stop → voiced stop → voiced continuant), or in segments with a higher degree of oral aperture (e.g. geminate stop → singleton stop → continuant consonant, with perhaps eventual segmental loss); feature-based analyses seem to be unable to capture this aspect.

Van der Hulst and Ewen (1991) suggest that the scalar nature of consonant mutation is most adequately accounted for by a framework in which sonority (CHAPTER 49: SONORITY) is expressed by C and V nodes that can either function as governing or governed nodes. In their system, the interpretation of a governing "C" is "some degree of oral closure that characterizes an obstruent" ([−sonorant]) and the interpretation of a governing "V" is "sonorant." The two elements "C" and "V" can also be adjoined to a governing element; an adjoined "C" is interpreted as "oral closure for sonorants" (in the case of nasal stops and laterals), and an adjoined "V" is interpreted as "periodic sound source" or "vocal cord vibration" (as expressed in feature-based frameworks by the feature [+voice]). An element "V" that is governed by an element "C" is interpreted as "continuous airflow" ([+continuant]):

(14) *C and V components in phonological representations* (van der Hulst and Ewen 1991)



Van der Hulst and Ewen (1991) further view the consonant mutations referred to as "inter-sonorant lenitions" as the result of imposing the element "V" from neighboring sonorant segments onto the consonant in question. Thus, in inter-sonorant positions, "V" can be added to the structure of a consonant either by adjunction – to give a voiced obstruent (/p/ → [b] and /f/ → [v]) – or by subjunction, to give a continuant (/p/ → [f] and /b/ → [v]). Within this theory, we might thus account for Northern Corsican consonant mutation as a process by which the element "V" is (a) adjoined to "C" segments that do not have an adjoined node and (b) subjoined to segments that involve "C" and an already adjoined "V" (thus, /p/ → [b], while /b/ → [β]).

Under this approach, another scalar mutation may involve a process by which the element "V" is adjoined to "C" segments that do not have an adjoined node and by which the element "V" is turned into the head element, so that the element "C" becomes the adjunct (thus, /p/ → [b], while /b/ → [m], as in e.g. Modern Irish initial nasalization).

The advantage of the approach suggested by van der Hulst and Ewen (1991) is clearly the fact that consonant mutations generally known as "lenitions" can be described as a more or less unified process involving an increase in the dominance of the element "V" (i.e. sonority). An account of stopping after nasals as observed in Soninke is less straightforward in this framework. Also, it is not immediately clear how the different mutations that do not have an overt phonetic trigger – such as the ones found in Southern Paiute or Fula – can be explained, and it is even less obvious how graded mutations in the Balto-Finnic and Sami languages would fit in this picture.

Another proposal for the representation sonority – especially sonority related to the degree of oral stricture – is formulated by Steriade (1993, 1994). Steriade defines slots to which laryngeal, place, and other features attach in terms of degrees of oral aperture. Released stops and affricates are viewed as sequences of a phase with complete oral closure (zero aperture, i.e. A_0) followed by a release phase (A_{fric} or A_{max}, also referred to as A_{rel} below).

(15) *The phonological representation of consonantal segments in Aperture Theory* (Steriade 1993, 1994)

oral stricture	A_0 A_{max}	A_0 A_{fric}	A_{fric}
place	[labial]	[labial]	[labial]
interpretation	labial oral released stop	labial affricate	labial fricative
nasality	[nasal]	[nasal]	
oral stricture	A_0 A_{max}	A_0 A_{max}	A_{max}
place	[labial]	[labial]	[labial]
interpretation	labial prenasal stop	labial nasal stop	labial approximant

Grijzenhout (1995, 1996) uses this framework to describe lenition as a form of loosening oral stricture and fortition as a procedure that increases oral stricture. The process by which a fricative is realized as a stop in Soninke can thus be described as one by which A_0 is inserted; the process by which a stop is realized as a fricative, e.g. in Spanish, involves deleting the A_0 slot from the representation:

(16) a. *Stopping (increase stricture) as insertion of* A_0

A_{rel} → A_0 A_{rel} (e.g. /f/ → [p] in Soninke)

b. *Spirantization (reduce stricture) as deletion of* A_0
$A_0\ A_{rel} \rightarrow A_{rel}$ (e.g. /b/ → [β] in Spanish)

Assuming that the articulation of long stops or geminates involves a long closure phase, long stops can be represented as elements with two A_0 nodes (17) (CHAPTER 37: GEMINATES). Southern Paiute gemination thus involves the same procedure as indicated above for Soninke stopping, i.e. insertion of A_0. Under consonant gradation, constriction loosens, which can be described as the loss of an A_0 slot (17b):

(17) a. *Gemination (increase stricture) as insertion of* A_0

short stop		*long stop*		
$A_0\ A_{rel}$	→	$A_0\ A_0\ A_{rel}$		
/p/	→	[pː]		Southern Paiute

b. *Consonant gradation (reduce stricture) as deletion of* A_0

long stop		*short stop*		*approximant*	
$A_0\ A_0\ A_{max}$	→	$A_0\ A_{max}$	→	A_{max}	
/pː/	→	[p]			Finnish
		/p/	→	[w]	

Other cases where single released stops are realized as fricatives also often involve a preceding vowel. Examples are Tigrinya and Biblical Hebrew post-vocalic spirantization; stops become more sonorant in the context of vowels, and the phonetic effect is that they are realized as fricatives. In Biblical Hebrew, post-vocalic singleton /p/ is realized as [f], /k/ is realized as [x], /t/ as [θ], /b/ as [v], /g/ as [ɣ], and /d/ as [ð] (Sampson 1973; Kenstowicz 1994: 53, 411, 417):[17]

(18) *Biblical Hebrew post-vocalic spirantization of single short stops*

/paːgaʃ/	[paːɣaʃ]	'meet (PERF)'
/ji-pgoːʃ/	[jifgoːʃ]	'meet (IMPERF)'
/kaːtab/	[kaːθav]	'write (PERF)'
/ji-ktob/	[jixtov]	'write (IMPERF)'
/gaːdal/	[gaːðal]	'become great (PERF)'

Note that in the same context, geminate stops are not affected, i.e. post-vocalic geminate stops do not alter:

(19) *Biblical Hebrew post-vocalic geminate stops*

/sappir/	[sapːir]	*[safpir]	'sapphire'
/gibboːr/	[gibːoːr]	*[givboːr]	'hero'
/giddeːl/	[gidːeːl]	*[giðdeːl]	'magnify (PERF)'

17 One reviewer asked why Biblical Hebrew and Tigrinya could not be accounted for as spread of [+continuant]. First, it is not obvious in current phonological theory that vowels are underlyingly specified for this feature. Second, fricatives – i.e. segments that are specified for the feature [+continuant] – do not trigger the spirantization process.

In Aperture Theory, we can capture the fact that vowels cause a change in oral stops and, at the same time, explain that a fricative does not cause a change in an oral or nasal stop (*/aspa/ → [asfa]; see Wetzels 1991). The change from released stop to fricative in the context of a preceding vowel involves decrease of constriction, which is expressed by deletion of an aperture node, as in (16b) and (20). After deletion of the aperture node for complete obstruction in the oral tract (i.e. A_0), the place feature is associated to the A_{rel} node and the result is a single fricative.

(20) *Post-vocalic spirantization as reduction of complete constriction in single stops (represented as deletion of an aperture node)*

a. A_0 A_{rel} | [labial] /p/

b. A_{rel} | [labial] /f/

Geminate stops do not undergo this process in Biblical Hebrew, due to the "uniformity condition" (Hayes 1986), which says that if a certain environment incurs a change in a feature or node of a particular segment – in this case a vowel that incurs omission of an aperture node of a following stop – every dominating slot linked to that feature (or node) must satisfy that environment. The second A0 slot dominating the place feature is not adjacent to a vowel, and deletion of the aperture node in (21) is therefore blocked.

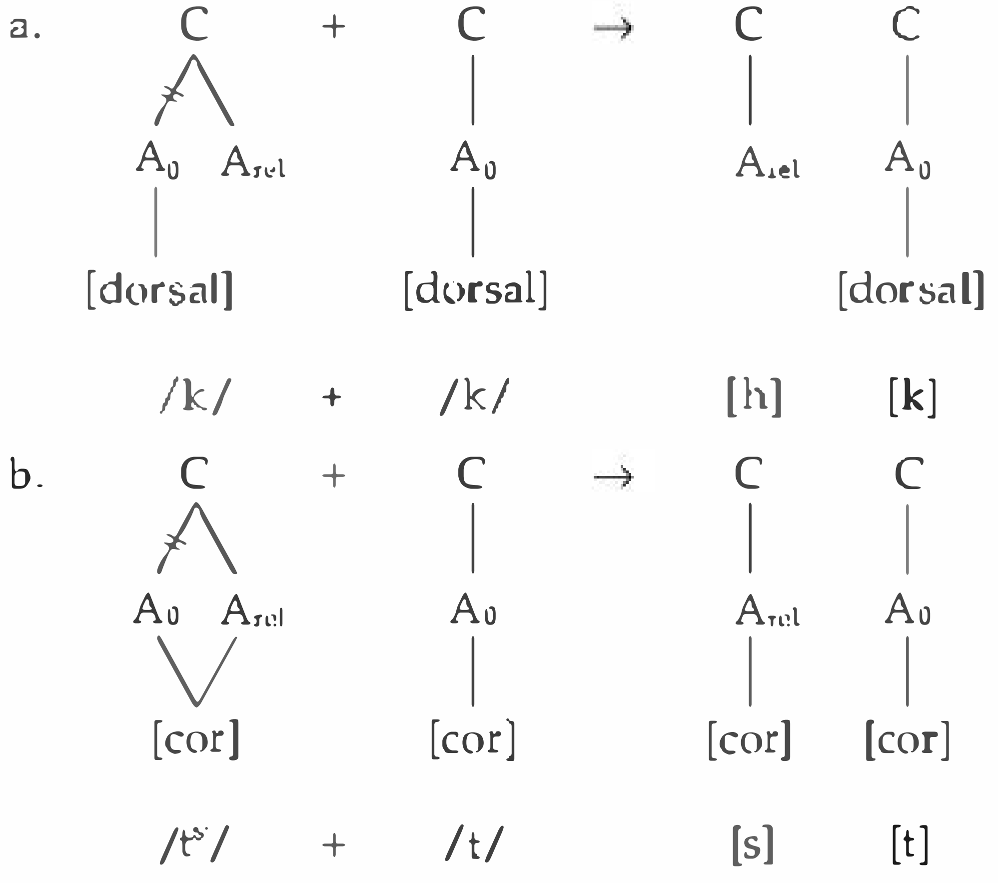
(21) *Geminate stop: Two aperture slots for closure sharing one place feature*

A_0 A_0 A_{rel} \/ [labial] /p:/

Note that the "uniformity condition" does not apply in the case of consonant gradation in Balto-Finnic languages and Sami, because the process of gradation is not triggered by a preceding vowel. In the languages discussed in §5, no vowel incurs gradation (or omission of an aperture node). Rather, the morphology determines whether or not gradation takes place, so that our account of this process is as shown in (17b) above.

Another case where Aperture Theory offers a more straightforward explanation than feature-based theories is Yukatec Maya degemination. In Yukatec Maya, a sequence of two homorganic stops is illicit. When two stops become adjacent due to a morphological process, the first one is realized as the placeless sound /h/ and the other retains its place of articulation (/k + k/ → [h k]; /t + t^j/ → [h t^j]). In cases where an affricate is immediately followed by an oral stop, it is realized as the corresponding fricative (/t^s + t/ → [s t]; /$t^ʃ$ + t/ → [ʃ t]); see McCarthy 1988; Lombardi 1990; Padgett 1991: 358–362. Under the assumption that this process involves delinking of the feature [–continuant], it is curious that oral stops do not retain their place of articulation when they are realized as a continuant, whereas affricates do. Under the proposal advocated here, this comes as no surprise. If a language disallows two adjacent stops, one of the aperture nodes for complete closure is deleted and in Yukatec Maya it is deleted together with the elements it dominates; i.e. when in oral stops the node for complete obstruction in the oral cavity (i.e. A_0) together with the node for the location of obstruction (i.e. the place feature) is deleted, an "empty" consonantal position remains that is phonetically realized as placeless /h/ (22a); conversely, when in affricates the A_0 node is deleted, a A_{rel}–Place association (which characterizes fricatives) remains (22b):

(22) *Delinking of one aperture node triggered by the OCP in Yukatec Maya*



In Aperture Theory, mutations that involve voicing or nasalization may be described as processes whereby the features [voice], [nasal], and/or Sonorant Voicing are attached to one of the aperture slots (Grijzenhout 1995, 1996). The framework of Aperture Theory encounters problems in accounting for the type of gradation that is attested in Northern Corsican, where voiceless stops alternate with voiced ones in the same context in which voiced stops alternate with continuants. Voiced stops have a shorter closure duration than voiceless ones, and the intuition behind lenition processes is thus that they reduce closure phases of stops (by degemination, voicing, reduction to incomplete closure, or even reduction to segment loss).

7 Summary and conclusion

This chapter discussed some cases of consonant mutation. By consonant mutation we understand here a change in a consonant that is not the result of neutralization (e.g. syllable-final obstruent devoicing) or assimilation (e.g. nasal place assimilation). Rather, consonant mutation is viewed as a process that increases or decreases the degree of sonority and/or the length and degree of oral stricture of a segment when there is: (i) a purely phonological context in which the mutation always takes place irrespective of speech style, e.g. inter-sonorant voicing in Burmese; intersonorant spirantization in Djapu and Spanish; or (ii) a "mixed" morphophonological environment, e.g. stopping in Soninke, where the mutation is sometimes brought on by a preceding nasal consonant or vowel, but may also have a morphological function, and consonant gradation in Estonian and Sami, where the consonant alternations sometimes take place in well-defined phonological contexts and sometimes in a context where the mutation has a grammatical function; or (iii) a morphosyntactic environment that induces mutation, e.g. word-initial spirantization in Southern Paiute and stem-initial spirantization in Fula.

Even though, from a historical perspective, consonant mutations may originally have had a phonological trigger (usually a preceding vowel or nasal), the fact that they also occur independently of a phonological environment is the first indication that an account that relies on the phonological contexts may be flawed. Moreover, accounts that regard mutations as phonological processes that involve spreading of a feature from a vowel or nasal onto the mutating consonant encounter problems when explaining why vowels and nasals spread these segmental features rather than fricatives and oral stops.

§6 discussed one theory that elegantly accounts for those mutations that involve increasing or decreasing sonority and one theory that provides an insightful account of mutations that involve increasing or decreasing oral stricture. Van der Hulst and Ewen (1991) propose to represent segments by means of C and V elements. In their view, sonority-increasing mutations involve adding a V element either as a governing node, an adjunct, or a governed node; sonority-decreasing mutations would presumably involve adding a C element or taking away a V element. §6 showed that this framework not only works for mutations that alter one class of segments into another class of segments (e.g. when voiceless stops become voiced, when stops become fricatives, or when voiced oral stops become nasal stops), but also for scalar mutations that induce a change in different classes of segments (e.g. when underlying voiceless stops become voiced while underlying voiced stops are nasalized or when underlying voiceless stops become voiced stops in the same environment where voiced stops are spirantized). This theory is less successful, however, in explaining other types of consonant gradations, for instance those where geminates become singletons in the same context where singletons spirantize (Balto-Finnic and Sami consonant gradation).

Steriade (1993, 1994) proposes using aperture nodes indicating the degree of oral stricture in phonological representations. This theory is useful in explaining processes where oral stricture increases (e.g. when continuants become stops or when the closure duration of short stops is extended) or decreases (e.g. when geminates degeminate where short stops spirantize). The downside of the theory is that it encounters difficulties in explaining scalar mutations of the type where underlying voiceless stops are voiced while underlying voiced stops are nasalized (e.g. Modern Irish initial nasalization), or when underlying voiceless stops are voiced in the same environment in which voiced stops are spirantized (e.g. Northern Corsican intervocalic consonant gradation).

§6 thus focused on two theories that are an improvement compared to feature-based accounts in the sense that they are able to explain scalar mutations that take place without an overt phonological trigger. However, the mutations that the one theory elegantly accounts for pose a puzzle for the other theory and vice versa. The problem for phonological theory thus remains the fact that there is as yet no unified account for the different types of consonant mutations that we can describe in layman's terms as changes that increase or decrease the level of sonority (a) expressed by laryngeal and nasal configurations, or (b) expressed by changes in oral aperture. Even though the description of consonant mutation is thus relatively simple, a phonological account is not.

ACKNOWLEDGMENTS

I would like to thank the editors for their invitation to contribute to the *Companion*. Thanks are also due to Colin Ewen, Harry van der Hulst, Glyne Piggott, Holly Winterton, and Wim Zonneveld for helpful discussions on the topic of consonant mutation. I also appreciate the comments of Bernhard Wälchli and two anonymous reviewers.

REFERENCES

Arnott, David W. 1970. *The nominal and verbal systems of Fula*. Oxford: Oxford University Press.

Batibo, Herman. 2000. System in the sounds of Africa. In Vic Webb & Kembo-Sure (eds.) *African voices: An introduction to the languages and linguistics of Africa*, 160–196. Oxford: Oxford University Press.

Berendsen, Egon. 1983. Final devoicing, assimilation, and subject clitics in Dutch. In Hans Bennis & W. U. S. van Lessen Kloeke (eds.) *Linguistics in the Netherlands 1983*, 21–29. Dordrecht: Foris.

Brockhaus, Wiebke. 1995. *Final devoicing in the phonology of German*. Tübingen: Niemeyer.

Campbell, George L. 1995. *Concise compendium of the world's language*. London & New York: Routledge.

Campbell, Lyle. 2004. *Historical linguistics: An introduction*. 2nd edn. Cambridge, MA: MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Elzinga, Dirk. 1996. Fula consonant mutation and morpheme constraints. *UC Irvine Working Papers in Linguistics* 2. 43–57.

Evans, Nicholas D. 1998. Iwaidja mutation and its origins. In Jae Jung Song & Anna Siewierska (eds.) *Case, typology and grammar: In honour of Barry J. Blake*, 115–149. Amsterdam & Philadelphia: John Benjamins.

Foley, James. 1970. Phonological distinctive features. *Folia Linguistica* 4. 87–92.

Gordon, Matthew. 1998. A fortition-based approach to Balto-Fennic-Sámi consonant gradation. *Folia Linguistica Historica* 18. 49–79.

Grijzenhout, Janet. 1991. Fula initial consonant mutation. In Peter Coopmans, Bert Schouten & Wim Zonneveld (eds.) *OTS Yearbook 1991*, 33–47. Utrecht: Research Institute for Language and Speech, University of Utrecht.

Grijzenhout, Janet. 1995. Irish consonant mutation and phonological theory. Ph.D. dissertation, University of Utrecht.

Grijzenhout, Janet. 1996. Consonant weakening processes and aperture theory. *Toronto Working Papers in Linguistics* 15. 105–124.

Harms, Robert T. 1962. *Estonian grammar*. Bloomington: Indiana University & The Hague: Mouton.

Harms, Robert T. 1966. Stress, voice, and length in Southern Paiute. *International Journal of American Linguistics* 32. 228–235.

Harris, James W. 1969. *Spanish phonology*. Cambridge, MA: MIT Press.

Harris, John. 1994. *English sound structure*. Oxford: Blackwell.

Hayes, Bruce. 1986. Inalterability in CV phonology. *Language* 62. 321–351.

Hulst, Harry van der & Colin J. Ewen. 1991. Major class and manner features. In Pier Marco Bertinetto, Michael Kenstowicz & Michele Loporcaro (eds.) *Certamen phonologicum II: Papers from the 1990 Cortona Phonology Meeting*, 19–41. Turin: Rosenberg & Sellier.

Iverson, Gregory K. & Joseph C. Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12. 369–396.

Karttunen, Frances E. 1970. Problems in Finnish phonology. Ph.D. dissertation, Indiana University, Bloomington.

Kendall, Martha B. & Charles Bird. 1982. Initial consonant change in Soninké. *Anthropological Linguistics* 24. 1–13.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Keyser, Samuel J. & Paul Kiparsky. 1984. Syllable structure in Finnish phonology. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*, 7–31. Cambridge, MA: MIT Press.

Keyser, Samuel J. & Kenneth N. Stevens. 2006. Enhancement and overlap in the speech chain. *Language*. 33–63.

Lombardi, Linda. 1990. The nonlinear organization of the affricate. *Natural Language and Linguistic Theory* 8. 375–425.

Lombardi, Linda. 1996. Restrictions on direction of voicing assimilation: An OT account. *University of Maryland Working Papers in Linguistics* 4. 88–112.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.

McLaughlin, John E. 1984. A revised approach to Southern Paiute phonology. *Kansas Working Papers in Linguistics* 9. 47–79.

Morphy, Frances. 1983. Djapu, a Yolngu dialect. In R. M. W. Dixon & Barry J. Blake (eds.) *The handbook of Australian languages*, vol. 3, 1–188. Amsterdam: John Benjamins.

Ní Chiosáin, Máire. 1991. Topics in the phonology of Irish. Ph.D. dissertation, University of Massachusetts, Amherst.

Padgett, Jaye. 1991. Stricture in feature geometry. Ph.D. dissertation, University of Massachusetts, Amherst.

Paradis, Carole. 1987a. Glide alternations in Pulaar (Fula) and the theory of charm and government. In David Odden (ed.) *Current approaches to African linguistics 4*, 327–338. Dordrecht: Foris.

Paradis, Carole. 1987b. Strata and syllable dependencies in Fula: The nominal class. *Journal of African Languages and Linguistics* 9. 123–139.

Paradis, Carole. 1992. *Lexical Phonology and Morphology: The nominal classes in Fula*. New York: Garland.

Pedersen, Holger. 1897. *Aspirationen i Irsk*. Copenhagen: Spirgatis.

Sampson, Geoffrey. 1973. Duration in Hebrew consonants. *Linguistic Inquiry* 4. 101–104.

Sapir, Edward. 1930. Southern Paiute, a Shoshonean language. *Proceedings of the American Academy of Arts and Sciences* 65. 1–296.

Skousen, Royal. 1972. On capturing regularities. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 8. 567–577.

Steriade, Donca. 1993. Closure, release, and nasal contours. In Marie K. Huffman & Rena A. Krakow (eds.) *Nasals, nasalization, and the velum*, 401–470. Orlando: Academic Press.

Steriade, Donca. 1994. Complex onsets as single segments: The Mazateco pattern. In Jennifer Cole & Charles W. Kisseberth (eds.) *Perspectives in phonology*, 203–291. Stanford: CSLI.

Thurneysen, Rudolf. 1898. Review of Holger Pedersen (1897). *Anzeiger für indogermanische Sprach- und Altertumskunde. Beiblatt zu den Indogermanischen Forschungen* 9. 42–48.

Uralic languages. 2010. In *Encyclopaedia Britannica*. Available (August 2010) at www.britannica.com/EBchecked/topic/619069/Uralic-languages.

Vainikka, Anne. 1988. Finnish consonant gradation unified. Unpublished ms., University of Massachusetts, Amherst.

Wetzels, W. Leo. 1991. Contrastive and allophonic properties of Brazilian Portuguese vowels. In Dieter Wanner & Douglas A. Kibbee (eds.) *New analyses in Romance linguistics*, 77–99. Amsterdam & Philadelphia: John Benjamins.

Wiswall, Wendy J. 1989. Fula consonant gradation: In support of radical underspecification. *Proceedings of the West Coast Conference on Formal Linguistics* 8. 429–444.

87 Neighborhood Effects

Adam Buchwald

1 Introduction

The organization of lexical knowledge at the phonological level has long been thought to incorporate some encoding of similarity neighborhoods: in the network structure of the lexicon, words that share phonological components are more closely connected to one another than words that are different. The notion of a "neighbor" as a similar word has provided tremendous insight into a variety of psycholinguistic phenomena related to spoken word recognition and spoken word production. The present chapter explores this notion of neighbor, focusing on the characterization of neighbors of a target word as the other words that are activated when that target is active. As we will see, this notion allows us to predict inhibitory or facilitatory effects depending on the task. These effects are well documented across tasks and participant populations. Following a brief description of the notion of the mental lexicon and competition in lexical access, we review research in spoken word recognition and production, and describe how the neighborhood construct (and similarity more generally) has been applied in phonologically based psycholinguistic research. We then examine how the notion of neighbor has been applied to domains such as language acquisition, language impairment, and other modalities of communication, including written language processing and audiovisual speech perception.

1.1 *Neighbors compete for lexical selection*

The term "mental lexicon" is typically used to refer to a network of words representing an individual's lexical knowledge (Oldfield 1966; Forster 1978). Accounts of lexical knowledge posit multiple types of lexical organization, notably including meaning-based organization and form-based organization. The notion of similarity among lexical items varies at these levels; at the level of meaning-based organization, neighbors share semantic features (e.g. *dog* and *cat* are semantic neighbors, sharing several semantic properties, such as "four-legged," "domesticated," and "animal," among others). With respect to form-based organization, neighbors are defined words that share phonological and phonetic detail, such as *cat* [kæt] and *cap* [kæp] (e.g. Luce and Pisoni 1998; see also Greenberg and Jenkins 1964 and

Landauer and Streeter 1973). The phonological neighbors of a word (e.g. *cat*) are the other words that share phonological structure (e.g. *cap, hat, kit*) and become activated when the word is activated in spoken word recognition and in word production. In short, these are the other words that are competing for lexical access. As all psycholinguistic accounts of lexical processing posit separate levels of meaning-based and form-based processing, the neighbors of a word at each of these levels are the other words competing for lexical selection (Dell *et al.* 1997; Vitevitch and Luce 1998, 1999; Levelt *et al.* 1999; Luce *et al.* 2000; Rapp and Goldrick 2000).

The first part of the chapter discusses how the fundamental description of neighbors as competitors in lexical selection processes affects spoken word recognition and spoken word production. We begin with a discussion of some of the seminal results in word recognition that have helped to shape our understanding of neighborhood effects, and explore how neighborhood effects relate to other lexical and sub-lexical properties. This is followed by a review of the spoken word production literature, including both lexical-level processing and phonetic differences that arise due to neighborhood structure.

2 Neighborhood effects in spoken word recognition

While there are many differences among accounts of spoken word recognition, there is widespread agreement that when a word is heard, recognition involves a selection process in which the listener accesses a lexical item among several competing alternatives (Morton 1969, 1979; Marslen-Wilson and Welsh 1978; Elman and McClelland 1986; Norris 1994; Luce and Pisoni 1998; Luce *et al.* 2000; Norris *et al.* 2000; see Jusczyk and Luce 2002 for a review). Luce and Pisoni (1998) formalize the Neighborhood Activation Model (NAM), in which it follows from the nature of lexical competition that *ceteris paribus* words with a lot of active neighbors (i.e. words in *dense* lexical neighborhoods) are harder to access than words with few active neighbors (i.e. in *sparse* lexical neighborhoods), as there is more competition during lexical selection (see also Luce *et al.* 2000 for a computational implementation of NAM). Because lexical selection is a competitive process, factors that strengthen a word's activation (e.g. high frequency) also help in recognizing that word. Therefore, high-frequency words from low-density neighborhoods are easier to access than low-frequency words from high-density neighborhoods (see also CHAPTER 90: FREQUENCY EFFECTS).

Luce and Pisoni (1998) examined the performance of participants on a variety of word recognition tasks with "easy" stimuli (high-frequency words from low-density neighborhoods) and "hard" stimuli (low-frequency words from high-density neighborhoods). The tasks they used included perceptual identification (written response to aural presentation), lexical decision, and word repetition. For each task, they reported that participants responded faster and/or more accurately to words from sparse phonological neighborhoods compared to words from dense phonological neighborhoods. The results indicated that the best predictor of performance was frequency-weighted neighborhood density, a measure indicating the frequency of a word compared to the total frequency of that word and its neighbors (also see Newman *et al.* 1997). These results are consistent with the claim that an increase in the number of neighbors leads to more competition in tasks that involve spoken word recognition and that more competition makes recognition slower and less accurate.

To determine whether these neighborhood properties affect online word recognition errors, Vitevitch (2002b) analyzed a corpus of spoken word recognition errors ("slips of the ear") and determined that words occurring in dense phonological neighborhoods were more prone to recognition errors than words in sparse neighborhoods. Taken together, these results suggest that the density of a word's phonological neighborhood is directly related to the strength of competition in word recognition; words from high-density neighborhoods have stronger, more active competitors, and recognition of these words is slower and more likely to engender spoken word recognition errors than words with low density neighborhoods (cf. Vitevitch and Rodríguez 2005).

Other studies have shown variance in neighborhood effects that are based on the *structure* of a neighborhood, even when the number of neighbors remains constant. Vitevitch (2002a) reported that subjects were slower to perform shadowing and lexical decision tasks for words with high onset density (i.e. a large proportion of neighbors sharing the initial phoneme) compared to words with lower onset density. Vitevitch (2007) reported on an auditory lexical decision task, a repetition task, and an AX (same–different) discrimination task using groups of words matched for neighborhood density but differing in neighborhood structure. In particular, Vitevitch used CVC words with different neighborhood *spread* – that is, the number of segment positions in the word that can be changed to form a new word. For example, *cat* has a spread of 3, as each segment can be changed to form a new word (e.g. *hat, kit, cap*), whereas *mob* has a spread of 2 (e.g. *lob, mop*, but there are no [mVb] words where V is not [ɑ]). Some stimulus words had neighbors that could be formed by substitutions of each segmental position, and others only had neighbors that could be formed from substituting one or two positions. Vitevitch (2007) found that participants were slower at responding to words with a spread of 3 than words with a spread of 2, even when the overall neighborhood size and frequency were controlled. The findings from these studies indicate that both the size and the structure of a lexical neighborhood affect spoken word recognition.

Magnuson *et al.* (2007) reported on the results from an eye-tracking study that provides more insight into the time course of lexical competition based on neighborhood properties. Magnuson *et al.* asked participants to perform an auditory word–picture matching recognition task. Participants heard instructions to click on a picture of an object which was in an array of four pictures. Rather than relying on accuracy or reaction time paradigms which infer the nature of cognitive processes from a single response, Magnuson *et al.* measured the participants' gaze toward different pictures in the array over time; thus, the eye-tracking paradigm allowed them to examine competition over the course of processing. Their findings indicated an early facilitatory effect of neighborhood density (revealed by participants looking toward the target), followed by a later inhibitory effect such as those more typically seen in word recognition studies (discussed above). Thus, by examining processing throughout the course of lexical access, Magnuson *et al.* were able to uncover a more nuanced characterization of competition effects in lexical access.

Thus far, we have discussed neighborhood density without providing a definition of a neighbor. The definition of a neighbor that is most commonly used in the word recognition literature is a word that differs from the target by a single segment deletion, addition, or substitution; in other words, words that differ from each other by an edit distance of 1. Thus, the neighbors of *cat* include *hat* (first segment substitution), *kit* (second segment substitution), *cap* (third segment substitution), *scat, cats* (insertions), and *at* (deletion), among others. This definition of neighbor has been used in large part because many of the seminal studies have focused on either CVC words or monosyllabic words more generally (Luce 1986; Luce and Pisoni 1998; Benkí 2003). For these words, there is a fair amount of variation in neighborhood density. However, when one looks over the entire lexicon rather than at a subset, more than half of the words in the lexicon are "hermits" by this definition; that is, they do not have any neighbors (Vitevitch 2008).

Buchwald *et al.* (2008) and Felty *et al.* (2008) attempt to examine this issue by inferring the most accurate definition of a neighbor from spoken word recognition errors. Felty *et al.* (2008) reported on a large database of spoken word recognition errors obtained by having participants identify words mixed with noise. Their 1428 stimulus words were designed to be representative of the English lexicon, including a range of syllable length, stress patterns, frequency, and familiarity. The words were randomly selected from a larger lexical database. The word recognition errors reported by the participants were taken to be a direct reflection of the words that were highly competitive with the target. They reported that, particularly for longer words, the responses typically differed from the target by an edit distance of greater than 1. Thus, while it remains likely that using words with a phoneme edit distance of 1 to define the lexical neighborhood can reasonably approximate neighborhood density effects in CVC words, it may not be the most appropriate definition of a lexical competitor.

2.1 *Phonotactic probability and lexical neighborhoods*

The nature of the competition effects we have been discussing focuses on the lexical selection process; that is, competition among lexical forms at a level of processing with word-level representations. However, it should be noted that the notion of lexical neighbor rests on some measure of sub-lexical phonetic similarity; for two words to be neighbors, they must share sub-lexical structure (e.g. *cat* and *hat* share [æt]). If a word has a lot of neighbors, then by definition there are many words that share its sub-lexical structure. Thus, it is worth asking whether having a lot of shared sub-lexical structure is directly related to the competition effect; that is, what is the role of *phonotactics*?

"Phonotactics" is the term used to refer to the sequential arrangements of segments in the words of a language (see also CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE). Consider the word *hang* [hæŋ], which is a perfectly well-formed word of English (i.e. it is phonotactically legal), with [h] in word-onset position and [ŋ] in the word-final coda position. However, a non-word such as *ngah* *[ŋæh] is not phonotactically legal, because it does not conform to the phonotactics of English, since English has phonotactic constraints both against [h] in coda and against [ŋ] in onset. The fact that languages have different phonemic inventories and that some languages restrict the segments in certain positions is a direct reflection of categorical phonotactic constraints (see CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS). In addition to these within-language constraints, it has been shown that participants can learn new phonotactic rules that do not exist in their language over the course of an experiment (Dell *et al.* 2000).

Of course, with respect to lexical selection, all lexical items in a language are phonotactically legal. Thus, many investigations in phonotactics have been more concerned with probabilistic phonotactics; that is, the likelihood that segments appear in specific syllabic positions, and the likelihood with which segments may co-occur (Jusczyk *et al.* 1994). Jusczyk *et al.* (1994) showed that 9-month-old infants prefer listening to phonotactic patterns that occur often in their native language to those that are attested, but occur less frequently.

The concept of phonotactic probability is closely tied to that of neighborhood density; phonotactic sequences that have high probability will be shared by many words. Thus, words containing those sequences will be likely to belong to high-density neighborhoods. In a variety of tasks, including wordlikeness judgments and non-word repetition, English speakers have been shown to process non-words with high phonotactic probability faster and rate these non-words as more "English-like" compared to their low phonotactic probability counterparts (Vitevitch *et al.* 1997; Vitevitch *et al.* 1999; Frisch *et al.* 2000). This leads to an apparent contradiction: acoustic stimuli are processed faster and more accurately when they have high phonotactic probability but slower and less accurately when they are words from dense neighborhoods.

Vitevitch and Luce (1998, 1999) explored this apparent contradiction by contrasting words of high and low phonotactic probability (and neighborhood density) with non-words of high and low phonotactic probability. Vitevitch and Luce (1998) had participants perform a repetition task on these four stimulus types. Their results indicated that the competition effects of words from high-density neighborhoods slowed processing – even though those words had high phonotactic probability. In contrast, having a high phonotactic probability facilitated the processing of non-words. Thus, while words with a lot of competitors were processed less efficiently, the processing of non-words was facilitated when they had substantial lexical support for their segments and segmental sequences, indicating that the mechanism for non-word processing is sensitive to phonotactic probability. These results were interpreted as evidence for separate lexical and sub-lexical representations and processing systems. High phonotactic probability facilitates sub-lexical processing in production tasks whereas low neighborhood density facilitates lexical processing in perception tasks.

Bailey and Hahn (2001) raised the issue that the data of Vitevitch and Luce do not specifically show separate lexical effects (i.e. neighborhood density) and sub-lexical effects (i.e. phonotactic probability). To address this issue, Bailey and Hahn presented participants with a wordlikeness judgment task in which they directly contrasted phonotactic probability and neighborhood density for non-words. They reported that the best predictors of wordlikeness judgments incorporated lexical neighborhood influences, phonotactic probability, and the relationship between the two. One important innovation of Bailey and Hahn's work was to incorporate a measure of phonetic similarity that considered similarity among segments (e.g. that treated /k/ and /g/ as more similar than /k/ and /b/), thus using a more linguistically sophisticated notion of phonetic similarity among words. Pylkkänen *et al.* (2002; see also Pylkkänen and Marantz 2003) provided MEG (magneto-encephalography) support for the claim that lexical neighborhood effects and phonotactic probability effects are neurally distinct.

Storkel *et al.* (2006) further separated the effects of phonotactic probability and neighborhood density in a word-learning task. Participants had to learn novel words that varied orthogonally in phonotactic probability and neighborhood density. The results indicated that the participants were better at learning words from high-density neighborhoods than low-density neighborhoods, but worse at learning high-probability words than low-probability words. Storkel *et al.* argued that these findings revealed a facilitatory effect of neighborhood density in encoding and integrating novel lexical representations with previously stored lexical representations.

2.2 *Summary*

Neighborhood effects in word recognition may be described as effects of competition: words with a lot of phonetic neighbors that are strong competitors have a lot of competition for lexical access. Thus, having a large number of neighbors typically increases response latency in word recognition tasks. The precise properties of the neighbors and their relationship to the target can modulate these effects to some extent; in other words, some competitors are stronger than others. More recent work has begun to examine both the changes in lexical competition over the time course of word recognition and how different degrees of sub-phonemic similarity affect the strength of competitors in word recognition. The relationship between lexical competition resulting from neighborhood size and other phonological phenomena driven by similarity effects (e.g. gradient OCP constraint based on consonant similarity: Frisch *et al.* 2004; Coetzee and Pater 2008) remains relatively unexplored.

3 Neighborhood effects in spoken word production

As noted above, the neighbors of a target word are the other words that become activated when the target word is active. In the process of spoken word recognition, these other words compete for lexical selection and can make lexical access slower and more error prone. In spoken word production, however, the opposite pattern is seen; words from dense neighborhoods are produced faster and are less error prone than words from sparse neighborhoods.

In a variety of tasks examining speech production errors, Vitevitch and colleagues have reported that words from dense neighborhoods are produced faster and more accurately than words from sparse neighborhoods in English (Vitevitch 2002c; Vitevitch and Sommers 2003; Vitevitch *et al.* 2004). Vitevitch and Sommers (2003) performed a tip-of-the-tongue (TOT) elicitation task in which the lexical retrieval process is thought to stall before production of a word's form, even when the speaker may be able to access a variety of information about a word (meaning, gender, etc.). Vitevitch and Sommers reported that adults were significantly more likely to achieve a TOT state for words from sparse neighborhoods compared to words from dense neighborhoods. Similarly, Vitevitch (2002b) reported that speech errors were more likely for words from sparse neighborhoods compared to words from dense neighborhoods for two additional speech error-inducing tasks. These results are consistent with data from picture naming tasks as well (Vitevitch 2002c; Vitevitch *et al.* 2004), indicating that having a large number of similar words seems to facilitate the process of lexical retrieval in production.

Here we have another apparent contradiction: in spoken word recognition tasks, words from high-density phonological neighborhoods are recognized more slowly and less accurately, whereas in spoken word production tasks these words appear to be facilitated by their neighbors. The facilitatory effect of neighborhood density in speech production has been argued to follow in a straightforward fashion from interactive theories of speech production in which words are activated at a lexical level for lexical selection, and there is feedback from a "lower" sub-lexical level back up to this lexical level (Dell 1986, 1988; Dell *et al.* 1997; Rapp and Goldrick 2000; Dell and Gordon 2003; cf. Levelt *et al.* 1999, and see Vitevitch *et al.* 2004 for discussion). When a word is activated at the lexical level, it sends activation downstream to its phonemic constituents. Through the interactive process of feedback, the units representing the active phonemes send activation back up to the items at the lexical level that contain them. Thus when the unit(s) representing the word *cat* are activated on the lexical level, the units representing [k], [æ], and [t] on the sub-lexical level receive activation from this lexical unit. In a system with feedback, these sub-lexical units then send activation not only back up to *cat*, but also to the other words they are connected to (e.g. *cap*, *hat*, etc.), i.e. the neighbors of *cat*. This in turn provides more lexical support for the units representing those sounds, and thus makes the target word more likely to be produced than a semantic competitor which is not receiving additional activation from feedback. Thus, it is the interaction of lexical information and sub-lexical information that creates the facilitatory effect of high-density neighborhoods.

3.1 Phonetic effects of neighborhood density in speech production

In addition to examining speed and accuracy of production, researchers have looked at the effect of neighborhood density on a variety of phonetic and acoustic properties of speech production. One phenomenon that has been well documented is the expansion of the vowel space in the production of words from high-density neighborhoods compared to low-density neighborhoods (Munson and Solomon 2004; Wright 2004; Munson 2007). Each of these studies shows that vowels in words from dense neighborhoods are produced closer to the periphery of the vowel space, whereas vowels in words from sparse neighborhoods are produced closer to the center of the vowel space; in other words, the distinctiveness of the vowels in words from high-density neighborhoods is enhanced relative to vowels in words from low-density neighborhoods.

Vowel space expansion such as is documented in words from high-density neighborhoods is typical of what speakers do when they are producing "clear" speech (Bradlow 2002), and is also associated with more intelligible speakers (Bradlow *et al.* 1996). Thus, when speakers are producing words from dense phonological neighborhoods, they adopt the strategies used in clear speech, referred to as hyperarticulation in Lindblom's *Hyperspeech and Hypospeech* (H&H) theory (Lindblom 1990). Additionally, Scarborough (2004) found that vowels in low-frequency words from high-density neighborhoods exhibit more co-articulation (e.g. V-to-V co-articulation) than vowels in high-frequency words from low-density neighborhoods, which she argued (contra Lindblom 1990) is helpful to the process of lexical access for the listener.

In recent work, Baese-Berk and Goldrick (2009) address a specific type of neighborhood effect: the presence or absence of a specific minimal pair lexical item. They examined the productions of words beginning with a voiceless consonant that have a voiced consonant-initial cognate (e.g. *cod* ~ *god*) and compared them to voiceless-initial words without a voiced cognate (e.g. *cop* ~ **gop*). Their data revealed that participants produce more extreme voice onset times (VOT) when producing words with the minimal pair neighbor than when producing words without a minimal pair neighbor. As VOT is a key indicator of the voicing contrast (Lisker and Abramson 1964), the enhanced VOT in the presence of a minimal pair neighbor can be viewed as another type of hyper-articulation due to a lexical item being from a lexical neighborhood with a particular neighborhood structure.

As can be inferred from this limited review of phonetic consequences of neighborhood density and structure, there have been relatively few attempts to understand how lexical phonological properties such as neighborhood density can affect the acoustic details of speech production. Nevertheless, this remains a fruitful area of research for the future and will likely lead to further insights regarding the relationship between lexical representations in the lexicon and the processing systems that allow those representations to be articulated in speech production.

4 Lexical neighborhoods: Language acquisition and language impairment

4.1 *Language acquisition*

If the neighbors of a word are other phonetically similar words, then it stands to reason that, as we learn more words, the structure of our lexical neighborhoods will change. This issue of how lexical neighborhoods develop during language acquisition and how lexical neighborhoods affect children's language ability has been addressed in the literature in several ways. Many of the attempts to study this issue focus on analyses of children's phonological lexicons. In a straightforward analysis of age-appropriate lexicons, Charles-Luce and Luce (1990, 1995) reported that words in the vocabulary of younger children (5-year-olds) have fewer neighbors than the words in the vocabulary of older children (7-year-olds), which in turn have fewer neighbors than those words in the vocabulary of adults. In other words, as children learn more words, their lexical neighborhoods become denser. Charles-Luce and Luce argued that these findings indicate less need for detailed phonetic representations of words in younger children's lexicons, as there are fewer confusable words (cf. Dollaghan 1994).

One influential account of lexical acquisition holds that children's initial phonological representations are more holistic and only become more differentiated by encoding phonological and phonetic structure after their lexicons have grown (Walley 1993; CHAPTER 72: CONSONANT HARMONY IN CHILD LANGUAGE). Consistent with this account, Metsala (1997; see also Garlock *et al.* 2001) reported that during a word recognition task using the gating paradigm, children required less phonetic material to recognize words from sparse neighborhoods as they got older and their vocabularies increased. She argued that this reflects a more differentiated representation as children get older such that words can be recognized from their constituent parts rather than requiring the whole word for recognition.

Storkel (2004b) reported that children learn words from dense neighborhoods earlier than they learn words from sparse neighborhoods. Storkel (2002) argued that in the developing child's lexicon, words from dense neighborhoods have more detailed representations than words from sparse neighborhoods. Coady and Aslin (2003) reported that the early developing lexicon contains more words from high-density neighborhoods than the later lexicon, suggesting that infrequent sound patterns are learned later. Thus, they claimed that this is not consistent with an account of children's lexical representations in which they start impoverished and become more detailed later (as in Walley 1993).

In a direct comparison of neighborhood density effects in typically developing children, Munson *et al.* (2005) found that children at the age of 4;3 did not exhibit effects of neighborhood density on response time in a repetition task, but older children (7;2) did exhibit effects. However, children in both age groups showed an effect of phonotactic probability on onset-to-onset latency in non-word repetition. As with understanding spoken word processing in adults, the covariance of phonotactic probability and neighborhood density frequently makes effects of these properties quite difficult to disentangle (see Storkel 2004a, 2009 for discussion).

4.2 *Lexical neighborhoods and language impairment*

Effects of lexical neighborhood in language impairment have been studied in two broad populations: adults with acquired language impairment and children with developmental language impairment. While neighborhood effects have been reported in the language processing skills of each of these populations, there are differences that follow from differences in the populations.

With respect to adults with acquired language impairment (i.e. aphasia), Gordon (2002) reported that aphasic speakers produced words from high-density phonological neighborhoods more accurately in both spontaneous speech and in controlled picture-naming tasks, with a strong effect of neighborhood frequency in production as well. Goldrick *et al.* (2010) reported on analyses comparing the phonological errors of aphasic speakers with the intended target words. These analyses inferred neighborhood structure by using the word production errors as an index of the other active lexical competitors. They reported that responses and targets were more likely to share position-specific segmental information than is predicted by chance – in other words, neighbors that share the same segments in the same position within a word were more likely to be produced in error than neighbors of the same edit distance in which the segments did not share segmental position. They also reported an independent effect of sharing the first segment, in which responses sharing the first segment as the target were more likely to be selected than other possible words of the same edit distance from the target.

These effects seen in aphasic production errors presumably follow from the nature of the impairment. The individuals in these studies were adult speakers of the language with fully formed lexicons, and their later impairment had affected retrieval and production of words in production tasks. Thus, if the structure of the lexicon encodes similarity neighborhoods, it is unsurprising that these similarity neighborhoods would affect production errors. This contrasts with individuals with developmental language deficits (including phonological impairment as well as hearing impairment which may affect speech input), in which the impairment affects language processing alongside lexical development.

Newman and German (2002) examined the performance of 7- to 12-year-old typically developing children and children with word-finding difficulties on a series of repetition tasks. Their analyses revealed a number of interesting findings. First, children from both groups were more accurate at repeating words from sparse neighborhoods compared to words from dense neighborhoods, a result that likely reflects the facilitation that words from sparse neighborhoods receive from having fewer competitors for word recognition. Second, children in both groups were more accurate at repeating words from neighborhoods with high average neighborhood frequency. This result seemingly contrasts with the first finding, suggesting that word recognition in a neighborhood facilitates other recognition in that neighborhood. Here we have a facilitatory effect of having frequent neighbors, whereas the first finding was an inhibitory effect of competition. To address this apparent contradiction, Newman and German (2002) examined the number of neighbors that had a higher frequency than the target. Words with fewer neighbors of higher frequency were repeated more accurately than words with many neighbors of higher frequency. Thus, both groups of children were more accurate in repeating words from sparse neighborhoods with high-frequency neighbors. This suggests that the word recognition process for both groups was affected by the neighborhood size and structure, indicating that even children with word-finding difficulties appear to be sensitive to acoustic-phonetic similarity among words and organize their lexicons into lexical neighborhoods that are used in lexical processing tasks.

Another group of children who have been examined with respect to neighborhood effects are children with cochlear implants (CIs). The children that have been studied are typically born profoundly deaf and receive their CI – a neural prosthesis – early in life to provide auditory input from which spoken language abilities may be developed (Svirsky *et al.* 2000). Kirk *et al.* (1995) examined accuracy on word recognition tasks for children with CIs, and reported that they were more accurate at identifying "easy" words (high-frequency words from sparse neighborhoods) compared to "hard" words (low-frequency words from dense neighborhoods). Other work with adults with CIs and other hearing-impaired adults has reported similar effects, though sometimes diminished relative to healthy controls (CI: Collison *et al.* 2004; hearing-impaired: Dirks *et al.* 2001).

4.3 Summary

Research with typically developing children and individuals with language impairment demonstrates that the effects of lexical neighborhood are not limited to neurologically intact adults. There is evidence that these populations are also sensitive to acoustic-phonetic similarity and organize their lexical knowledge accordingly. Given the articulatory differences between words from high-density neighborhoods and words from low-density neighborhoods discussed in §3, it remains unexplored whether children acquiring language are able to exploit these differences in some way that facilitates the organization of lexical knowledge into acoustic-phonetically defined similarity neighborhoods.

5 Lexical neighborhood effects in orthographic processing and audiovisual perception

Thus far, we have discussed lexical neighborhood effects in the perception, acoustics, and articulation of spoken language. This work has largely focused on the storage and processing of acoustic-phonetic lexical knowledge. In this section, we briefly discuss findings indicating that lexical knowledge is stored in terms of lexical similarity neighborhoods in other modalities (e.g. orthography and visual speech) as well as in multimodal processing (e.g. audiovisual speech perception). These findings suggest that the effects of lexical organization on phonological processing are part of a broader class of effects from the organization of lexical knowledge.

5.1 *Orthographic similarity neighborhoods*

The literature on the effect of orthographic similarity neighborhoods on written language processing is extremely large, and much of it is outside the scope of the present chapter. We focus here on parallels between findings in orthography and the findings discussed in this chapter on phonological similarity neighborhoods. In a seminal paper on visual word recognition, Coltheart *et al.* (1977) defined a word's orthographic neighbors as those words that can be formed with one letter changed but letter positions remaining unchanged. This measure has become known as Coltheart's N (or just N, as used here), and has been used quite frequently in studies of visual word recognition (i.e. reading) as well as written language production (i.e. spelling).

Coltheart *et al.* reported that in a lexical decision task for visually presented words and non-words, participants were faster to respond to low N non-words than high N non-words. Coltheart *et al.* did not obtain the analogous finding for words. Later studies have been equivocal on the effect of N and neighborhood frequency for words in visual lexical decision tasks. It is common for studies to report a facilitatory effect of N (e.g. Andrews 1989, 1992; Sears *et al.* 1995) but an inhibitory effect of neighborhood frequency, or even a single high-frequency neighbor (e.g. Grainger *et al.* 1989; Grainger 1990; Grainger and Segui 1990; Carreiras *et al.* 1997). Recent attempts to disentangle this issue have shown that some neighbors appear to be stronger competitors than others (Davis and Taft 2005), and that rigid coding of letter position and word length may not be the best predictor of the make-up of orthographic similarity neighborhoods (see Grainger 2008 for a review, and Yarkoni *et al.* 2008 for a proposal).

In research on written language production, there have been relatively few studies, but the results have consistently shown a facilitatory effect of high-neighborhood density. For example, Roux and Bonin (2009) reported that healthy adults demonstrate faster and more accurate spelling of words from dense neighborhoods compared to words from sparse neighborhoods. Sage and Ellis (2006) reported on the spelling performance of BH, a brain-damaged individual with acquired dysgraphia. BH's impairment affected the working memory mechanism in written language production and she showed less impairment producing words from dense neighborhoods compared to words from sparse neighborhoods. The authors reported that there was a beneficial therapeutic effect for neighbors of trained words; if a word was trained, spelling accuracy on the neighbors of that word improved. Taken together, these studies suggest that orthographic similarity neighborhoods affect the cognitive processing involved in producing written language.

5.2 Similarity neighborhoods in visual and audiovisual speech processing

While researchers examining speech perception and spoken word recognition typically focus on the auditory modality, there has been longstanding evidence that shows that visual speech perception actively contributes to word recognition, even in normal hearing adults in clear listening conditions (Sumby and Pollack 1954; McGurk and MacDonald 1976). Over the past two decades, characterizations of speech perception and the sensory processing of speech signals have become more focused on the multimodal nature of speech (Massaro 1987, 1998; Summerfield 1987; Massaro and Cohen 1995; Massaro and Stork 1998; Calvert *et al.* 2004; Kim *et al.* 2004; Bernstein 2005; Rosenblum 2005). One line of work on this topic has explored the processing of linguistic information conveyed in visual speech, including the ability to identify words from visual-only speech signals (e.g. Auer and Bernstein 1997; Lachs *et al.* 2000; Auer 2002; Mattys *et al.* 2002).

Expanding on the notion of the perceptual equivalence class from Miller and Nicely (1955) (see also Shipman and Zue 1982 and Huttenlocher and Zue 1984), Auer and Bernstein (1997) developed the construct of lexical equivalence class (see also Lachs *et al.* 2000 and Mattys *et al.* 2002), which is an equivalence class for words that are indistinguishable from the visual speech stream (e.g. *pin* and *bin*, which differ only in voicing, a feature that is not detectable in visual speech). Mattys *et al.* (2002) and Auer (2002) showed that words with a large lexical equivalence class – the visual speech similarity neighborhood – were less recognizable than words with a small lexical equivalence class for both hearing and deaf observers. This is consistent with the notion of other words in the similarity neighborhood as competitors for word recognition, so words with few competitors are easier to recognize. Further, Tye-Murray *et al.* (2007) demonstrated that both auditory speech-based similarity neighborhoods and visual speech-based similarity neighborhoods are predictive of word recognition in audiovisual speech perception.

5.3 Summary

This section has explored the effects of similarity neighborhoods on other modalities of processing, including orthographic processing, visual speech processing, and audiovisual speech processing. In each of these domains, we have seen effects of similarity neighborhoods on task performance reflecting that the organization of lexical knowledge into neighborhoods of similar words is a general cognitive mechanism that is not specific to acoustic stimuli.

6 Conclusion and future directions

The research on form-based lexical similarity neighborhoods has been influential in revealing the nature of lexical processing as competition among lexical nodes.

This conception helps explain a variety of findings in spoken word recognition, as well as word recognition in other modalities. In language production, two distinct types of effect of lexical neighbors have been reported. First, there is a facilitatory effect of neighborhood size in tasks of lexical access in word production. This is likely due to feedback from a segmental level of processing to a lexical level – when words similar to the target are activated, they activate their constituent segments. These segments in turn send activation back up to the lexical nodes they are connected to, thus providing a facilitatory effect of density. Second, there are reported effects of neighborhood structure on the acoustic details of speech production. Many of the reported effects lead to hyperarticulation of words in dense neighborhoods (e.g. expanded vowel space; more extreme VOT), which presumably helps to keep these words distinct from their neighbors. Many of these effects of neighborhood are also seen in acquisition, in cases of language impairment, and in other modalities of form-based lexical processing.

One issue that arises in each domain we discussed is the relationship between lexical effects of neighborhood structure and sub-lexical effects of phonotactic probability. The widely held view on these related phenomena is that they exert their effects on different levels of processing, with a large neighborhood associated with slower and less accurate word recognition at a lexical processing level but high probability associated with faster and more accurate recognition at a sub-lexical processing level. While drawing the distinction between the two levels is parsimonious and fits with both behavioral and neural evidence, it is worth noting that phonotactic probability is an emergent property of the structure of the lexicon; that is, phonotactic probability is defined over the lexicon. Thus, these two properties that appear to generate opposite effects in word recognition come from different levels of generalization over the same representations, and each can be thought of as an effect of similarity. The relationship between these similarity-based effects and other similarity effects that have been explored in phonology (e.g. Frisch *et al.* 2004; Coetzee and Pater 2008) remains somewhat unexplored in the literature.

Another direction of future research, which builds on some recent work, involves generating a more detailed definition of similarity and of what it means to be the competitor of a target word. While many of the previous investigations into lexical similarity neighborhood effects have used relatively coarse-grained metrics for discussing similarity, a variety of more recent efforts have been made to generate a more phonetically driven notion of similarity (Bailey and Hahn 2001; Hahn and Bailey 2005; Albright 2006; Felty *et al.* 2008). One promising direction involves using graph theory to better understand the network structure of the mental lexicon (Vitevitch 2008; Gruenenfelder and Pisoni 2009). Future attempts to refine these metrics will hopefully provide a more precise, gradient measure of similarity which can better predict effects and perhaps be useful in designing treatment protocols that help people with acquired form-based lexical impairment (as in Sage and Ellis 2006).

Finally, it is critical to consider the possible role of lexical neighborhood density and neighborhood effects in phonological alternations. The notion that lexical frequency may play a role in phonological alternations has been advanced (e.g. Pierrehumbert 2001) with the argument that frequent words may be more likely to undergo sound change, but whether other lexical factors such as neighborhood density affect the likelihood of a word participating in alternations has received less attention. Hall (2005) reported that the incidence of Canadian Raising (i.e. where /aɪ/ → [ʌɪ] before voiceless obstruents, as in *write* → [rʌɪt]; cf. *ride* → [raɪd]) among English speakers was actually affected by the neighbors sharing the CV biphone, even though Canadian Raising is thought to be conditioned by whether the following consonant is voiced. However, whether lexical neighborhoods can be shown to predict the application of a phonological process (as lexical frequency has been shown to) remains largely unexplored.[1]

REFERENCES

Albright, Adam. 2006. Gradient phonotactic effects: Lexical? grammatical? both? neither? Paper presented at the 80th Annual Meeting of the Linguistic Society of America, Albuquerque.

Andrews, Sally. 1989. Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory and Cognition* 15. 802–814.

Andrews, Sally. 1992. Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory and Cognition* 18. 234–254.

Auer, Edward T., Jr. 2002. The influence of the lexicon on speech read word recognition: Contrasting segmental and lexical distinctiveness. *Psychonomic Bulletin and Review* 9. 341–347.

Auer, Edward T., Jr. & Lynne E. Bernstein. 1997. Speechreading and the structure of the lexicon: Computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness. *Journal of the Acoustical Society of America* 102. 3704–3710.

Baese-Berk, Melissa & Matthew Goldrick. 2009. Mechanisms of interaction in speech production. *Language and Cognitive Processes* 24. 527–554.

Bailey, Todd M. & Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44. 568–591.

Benkí, Jose. 2003. Quantitative evaluation of lexical status, word frequency and neighborhood density as context effects in spoken word recognition. *Journal of the Acoustical Society of America* 113. 1689–1705.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Ph.D. dissertation, University of Massachusetts, Amherst.

Bernstein, Lynne E. 2005. Phonetic processing by the speech perceiving brain. In Pisoni & Remez (2005). 79–98.

Bradlow, Ann R. 2002. Confluent talker- and listener-related forces in clear speech production. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology 7*, 241–273. Berlin & New York: Mouton de Gruyter.

Bradlow, Ann R., Gina Torretta & David B. Pisoni. 1996. Intelligibility of normal speech. I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20. 255–272.

[1] With respect to a formal mechanism that could account for lexical neighborhood effects on phonological alternations, one prominent approach within Optimality Theory (Prince & Smolensky 1993) focuses on output–output faithfulness (Burzio 1994; Benua 1997) among morphemes and their related allomorphs. While this formal mechanism is typically applied to the different allomorphs of a single morpheme rather than to words from different morphological paradigms, it provides a blueprint of a formal mechanism that could be adapted to account for effects among similar words that are not from the same morphological paradigm. See also CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS and CHAPTER 83: PARADIGMS.

Buchwald, Adam, Robert A. Felty & David B. Pisoni. 2008. Neighbors as competitors: Phonological analysis of spoken word recognition errors. *Journal of the Acoustical Society of America* 123. 3328.

Burzio, Luigi. 1994. *Principles of English stress*. Cambridge: Cambridge University Press.

Calvert, Gemma A., Charles Spence & Barry E. Stein (eds.) 2004. *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Carreiras, Manuel, Manuel Perea & Jonathan Grainger. 1997. Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory and Cognition* 23. 857–871.

Charles-Luce, Jan & Paul A. Luce. 1990. Similarity neighborhoods of words in young children's lexicons. *Journal of Child Language* 17. 205–215.

Charles-Luce, Jan & Paul A. Luce. 1995. An examination of similarity neighborhoods in young children's receptive vocabulary. *Journal of Child Language* 22. 727–735.

Coady, Jeffry A. & Richard N. Aslin. 2003. Phonological neighborhoods in the developing lexicon. *Journal of Child Language* 30. 441–469.

Coetzee, Andries W. & Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26. 289–337.

Collison, Elizabeth A., Benjamin Munson & Arlene E. Carney. 2004. Relations among linguistic and cognitive skills and spoken word recognition in adults with cochlear implants. *Journal of Speech, Language, and Hearing Research* 47. 496–508.

Coltheart, Max, Eileen Davelaar, Jon Torli Jonasson & Derek Besner. 1977. Access to the internal lexicon. In Stan Dornic (ed.) *Attention and performance*, vol. 6, 535–555. Hillsdale, NJ: Lawrence Erlbaum.

Davis, Colin J. & Marcus Taft. 2005. More words in the neighborhood: Interference in lexical decision due to deletion neighbors. *Psychonomic Bulletin and Review* 12. 904–910.

Dell, Gary S. 1986. A spreading activation theory of retrieval in sentence processing. *Psychological Review* 93. 283–321.

Dell, Gary S. 1988. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language* 27. 124–142.

Dell, Gary S. & Jean K. Gordon. 2003. Neighbors in the lexicon: Friends or foes. In Niels O. Schiller & Antje S. Meyer (eds.) *Phonetics and phonology in language comprehension and production: Differences and similarities*, 9–38. Berlin & New York: Mouton de Gruyter.

Dell, Gary S., Kristopher D. Reed, David R. Adams & Antje S. Meyer. 2000. Speech errors, phonotactic constraints & implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26. 1355–1367.

Dell, Gary S., Myrna F. Schwartz, Nadine Martin, Eleanor M. Saffran & Deborah A. Gagnon. 1997. Lexical access in aphasic and nonaphasic speakers. *Psychological Review* 104. 801–838.

Dirks, Donald D., Sumiko Takayanagi, Moshfegh Anahita, P. Douglas Noffsinger & Stephen A. Fausti. 2001. Examination of the neighborhood activation theory in normal and hearing-impaired listeners. *Ear and Hearing* 22. 1–13.

Dollaghan, Charles A. 1994. Children's phonological neighborhoods: Half empty or half full? *Journal of Child Language* 21. 257–271.

Elman, Jeffrey L. & James L. McClelland. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18. 1–86.

Felty, Robert A., Adam Buchwald & David B. Pisoni. 2008. Lexical analysis of spoken word recognition errors. *Journal of the Acoustical Society of America* 123. 3327.

Forster, Kenneth I. 1978. Accessing the mental lexicon. In Edward Walker (ed.) *Explorations in the biology of language*, 139–174. Montgomery, VT: Bradford.

Frisch, Stefan A., Nathan R. Large & David B. Pisoni. 2000. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42. 481–496.

Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22. 179–228.

Garlock, Victoria M., Amanda C. Walley & Jamie L. Metsala. 2001. Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language* 45. 468–492.

Goldrick, Matthew, Jocelyn R. Folk & Brenda Rapp. 2010. Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language* 62. 113–134.

Gordon, Jean K. 2002. Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language* 82. 113–145.

Grainger, Jonathan. 1990. Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language* 29. 228–244.

Grainger, Jonathan. 2008. Cracking the orthographic code: An introduction. *Language and Cognitive Processes* 23. 1–35.

Grainger, Jonathan & Juan Segui. 1990. Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception and Psychophysics* 47. 191–198.

Grainger, Jonathan, J. Kevin O'Regan, Arthur M. Jacobs & Juan Segui. 1989. On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception and Psychophysics* 45. 189–195.

Greenberg, Joseph H. & James J. Jenkins. 1964. Studies in the psychological correlates of the sound system of American English. *Word* 20. 157–177.

Gruenenfelder, Thomas M. & David B. Pisoni. 2009. The lexical restructuring hypothesis and graph theoretic analyses of networks based on random lexicons. *Journal of Speech, Language, and Hearing Research* 52. 596–609.

Hahn, Ulrike & Todd M. Bailey. 2005. What makes words sound similar? *Cognition* 97. 227–267.

Hall, Kathleen Currie. 2005. Defining phonological rules over lexical neighbourhoods: Evidence from Canadian raising. *Proceedings of the West Coast Conference on Formal Linguistics* 24. 191–199.

Huttenlocher, D. P. & Victor W. Zue. 1984. A model of lexical access from partial phonetic information. Paper presented at *IEEE International Conference on Acoustics, Speech and Signal Processing*, San Diego.

Jusczyk, Peter W. & Paul A. Luce. 2002. Speech perception and spoken word recognition: Past and present. *Ear and Hearing* 23. 2–40.

Jusczyk, Peter W., Paul A. Luce & Jan Charles-Luce. 1994. Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33. 630–645.

Kim, Jeesun, Chris Davis & Phill Krins. 2004. Amodal processing of visual speech as revealed by priming. *Cognition* 93. B39–B47.

Kirk, Karen Iler, David B. Pisoni & Mary Joe Osberger. 1995. Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear and Hearing* 16. 470–481.

Lachs, Lorin, Jonathan W. Weiss & David B. Pisoni. 2000. Use of partial stimulus information by cochlear implant users and listeners with normal hearing in identifying spoken words: Some preliminary analyses. *The Volta Review* 102. 303–320.

Landauer, Thomas K. & Lynn A. Streeter. 1973. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior* 12. 119–131.

Levelt, Willem J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22. 1–38.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds.) *Speech production and speech modeling*, 403–439. Dordrecht: Kluwer.

Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384–422.

Luce, Paul A. 1986. Neighborhoods of words in the mental lexicon. Ph.D. dissertation, Indiana University.

Luce, Paul A. & David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19. 1–36.

Luce, Paul A., Stephen D. Goldinger, Edward T. Auer, Jr. & Michael S. Vitevitch. 2000. Phonetic priming, neighborhood activation and PARSYN. *Perception and Psychophysics* 62. 615–625.

Magnuson, James S., James A. Dixon, Michael K. Tanenhaus & Richard N. Aslin. 2007. The dynamics of lexical competition during spoken word recognition. *Cognitive Science* 31. 133–156.

Marslen-Wilson, William & A. Welsh. 1978. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10. 29–63.

Massaro, Dominic W. 1987. Speech perception by ear and eye. In Barbara Dodd & Ruth Campbell (eds.) *Hearing by eye: The psychology of lip-reading*, 53–84. Mahwah, NJ: Lawrence Erlbaum.

Massaro, Dominic W. 1998. *Perceiving talking faces: From speech perception to a behavioral principle.* Cambridge, MA: MIT Press.

Massaro, Dominic W. & Michael M. Cohen. 1995. Perceiving talking faces. *Current Directions in Psychological Science* 4. 104–109.

Massaro, Dominic W. & David G. Stork. 1998. Speech recognition and sensory integration: A 240-year-old theorem helps explain how people and machines can integrate auditory and visual information to understand speech. *American Scientist* 86. 236–244.

Mattys, Sven L., Lynne E. Bernstein & Edward T. Auer, Jr. 2002. Stimulus-based lexical distinctiveness as a general word-recognition mechanism. *Perception and Psychophysics* 64. 667–679.

McGurk, Harry & J. MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264. 746–748.

Metsala, Jamie L. 1997. An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory and Cognition* 25. 47–56.

Miller, George A. & Patricia Nicely. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27. 338–352.

Morton, John. 1969. Interaction of information in word recognition. *Psychological Review* 76. 165–178.

Morton, John. 1979. Word recognition. In John Morton & J. C. Marshall (eds.) *Structures and processes*, 109–156. Cambridge, MA: MIT Press.

Munson, Benjamin. 2007. Lexical access, lexical representation & vowel articulation. In Jennifer Cole & José Ignacio Hualde (eds.) *Laboratory phonology 9*, 201–228. Berlin & New York: Mouton de Gruyter.

Munson, Benjamin & Nancy P. Solomon. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47. 1048–1058.

Munson, Benjamin, Cyndie L. Swenson & Shayla C. Manthei. 2005. Lexical and phonological organization in children: Evidence from repetition tasks. *Journal of Speech, Language, and Hearing Research* 48. 108–124.

Newman, Rochelle S. & Diane J. German. 2002. Effects of lexical factors on word naming among normal-learning children and children with word-finding disorders. *Language and Speech* 43. 285–317.

Newman, Rochelle S., James R. Sawusch & Paul A. Luce. 1997. Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance* 23. 873–889.

Norris, Dennis. 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52. 189–234.

Norris, Dennis, James M. McQueen & Anne Cutler. 2000. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23. 299–370.

Oldfield, R. C. 1966. Things, words and the brain. *Quarterly Journal of Experimental Psychology* 18. 340–353.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam & Philadelphia: John Benjamins.

Pisoni, David B. & Robert E. Remez (eds.) 2005. *The handbook of speech perception.* Malden, MA & Oxford: Blackwell.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Pylkkänen, Liina & Alec Marantz. 2003. Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences* 7. 187–189.

Pylkkänen, Liina, Andrew Stringfellow & Alec Marantz. 2002. Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language* 81. 666–678.

Rapp, Brenda & Matthew Goldrick. 2000. Discreteness and interactivity in spoken word production. *Psychological Review* 107. 460–499.

Rosenblum, Lawrence D. 2005. Primacy of multimodal speech perception. In Pisoni & Remez (2005), 51–78.

Roux, Sébastien & Patrick Bonin. 2009. Neighborhood effects in spelling in adults. *Psychonomic Bulletin and Review* 16. 369–373.

Sage, Karen & Andrew W. Ellis. 2006. Using orthographic neighbours to treat a case of graphemic buffer disorder. *Aphasiology* 20. 851–870.

Scarborough, Rebecca A. 2004. Coarticulation and the structure of the lexicon. Ph.D. dissertation, University of California, Los Angeles.

Sears, Chris R., Yasushi Hino & Stephen J. Lupker. 1995. Neighborhood frequency and neighborhood size effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 21. 876–900.

Shipman, David W. & Victor W. Zue. 1982. Properties of large lexicons: Implications for advanced isolated word recognition systems. Paper presented at *IEEE 1982 International Conference on Acoustics, Speech and Signal Processing.*

Storkel, Holly L. 2002. Restructuring of similarity neighborhoods in the developing mental lexicon. *Journal of Child Language* 29. 251–274.

Storkel, Holly L. 2004a. Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, and Hearing Research* 47. 1454–1468.

Storkel, Holly L. 2004b. Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Journal of Applied Psycholinguistics* 25. 201–221.

Storkel, Holly L. 2009. Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language* 36. 291–321.

Storkel, Holly L., Jonna Armbrüster & Tiffany P. Hogan. 2006. Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research* 49. 1175–1192.

Sumby, William H. & Irwin Pollack. 1954. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* 26. 212–215.

Summerfield, A. Quentin. 1987. Some preliminaries to a comprehensive account of audio-visual speech perception. In Barbara Dodd & Ruth Campbell (eds.) *Hearing by eye: The psychology of lip-reading*, 3–52. Mahwah, NJ: Lawrence Erlbaum.

Svirsky, Mario A., Amy M. Robbins, Karen Iler Kirk, David B. Pisoni & Richard T. Miyamoto. 2000. Language development in profoundly deaf children with cochlear implants. *Psychological Science* 11. 153–158.

Tye-Murray, Nancy, Mitchell S. Sommers & Brent Spehar. 2007. Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification* 11. 233–241.

Vitevitch, Michael S. 2002a. Influence of onset density on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28. 270–278.

Vitevitch, Michael S. 2002b. Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of the ear. *Language and Speech* 45. 407–434.

Vitevitch, Michael S. 2002c. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 28. 735–747.

Vitevitch, Michael S. 2007. The spread of the phonological neighborhood influences spoken word recognition. *Memory and Cognition* 35. 166–175.

Vitevitch, Michael S. 2008. What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research* 51. 408–422.

Vitevitch, Michael S. & Paul A. Luce. 1998. When words compete: Levels of processing in perception of spoken words. *Psychological Science* 9. 325–329.

Vitevitch, Michael S. & Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40. 374–408.

Vitevitch, Michael S. & Eva Rodríguez. 2005. Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders* 3. 64–73.

Vitevitch, Michael S. & Mitchell S. Sommers. 2003. The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory and Cognition* 31. 491–504.

Vitevitch, Michael S., Paul A. Luce, Jan Charles-Luce & David Kemmerer. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40. 47–62.

Vitevitch, Michael S., Paul A. Luce, David B. Pisoni & Edward T. Auer, Jr. 1999. Phonotactics, neighborhood activation and lexical access for spoken words. *Brain and Language* 68. 306–311.

Vitevitch, Michael S., Jonna Armbrüster & Shinyung Chu. 2004. Sub-lexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory and Cognition* 30. 514–529.

Walley, Amanda C. 1993. The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review* 13. 286–350.

Wright, Richard A. 2004. Factors of lexical competition in vowel articulation. In John Local, Richard Ogden & Rosalind Temple (eds.) *Phonetic interpretation: Papers in laboratory phonology VI*, 26–50. Cambridge: Cambridge University Press.

Yarkoni, Tal, David Balota & Melvin Yap. 2008. Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin and Review* 15. 971–979.

66 Lenition

Naomi Gurevich

Lenition (German *Lenierung*, from Latin *lenire* 'weaken') is most commonly defined as "a 'relaxation' or 'weakening' of articulatory effort" (Hock 1991: 80). The term was coined by Thurneysen as one "used to describe a mutation of consonants which normally originated in a reduction of the energy employed in their articulation," and affects mostly consonants in intervocalic position (Thurneysen 1946: 74).

Below I present the processes that most commonly fall under the label of *lenition*, and make some observations that emerge from this list. The similarities between these processes and the way in which they pattern support the mostly uncontroversial view that lenition is indeed a group of similar phenomena. Kirchner writes that to his knowledge "no linguist has ever explicitly maintained the contrary view, that 'lenition' is merely an arbitrary collection of unrelated processes" (1998: 5). But while most acknowledge that the processes considered leniting are indeed related and form a coherent group, defining the exact criteria for group membership remains largely controversial and debates about the formalization, motivation, and even goal of lenition abound.

Three main approaches to lenition are presented here: formal, phonetic, and functional. The formal approach is the generative theory-based search to formalize synchronic rules of lenition processes. The phonetic approach seeks to determine what motivates the sound changes in question and what exactly constitutes "articulatory effort." The third, functional, take on lenition looks to matters of contrast maintenance, which exhibit an additional way in which lenitions pattern with great similarity, in its search for what constrains and sometimes triggers lenition phenomena.

1 Leniting processes

In this section I summarize the processes most commonly considered to fall under the cover term *lenition*. There is general agreement in the literature that all the processes listed below may be considered lenitions, but there is little agreement regarding the exact criteria required for membership in the lenition group.

In coining the term, Thurneysen suggests that leniting processes are characterized by some reduction in articulatory effort, but to date there has been no agreement on exactly what this entails. Among linguists there are intuitions regarding articulatory effort, some more accepted than others. Voicing, for example, has an explanation rooted in the laws of physics, specifically aerodynamics: intervocalically the vocal cords may continue to vibrate after the first vowel, through the consonant, and into the second vowel; but in final position – where final devoicing is encountered – aerodynamic conditions are not conducive to voicing (Westbury and Keating 1986). Other processes listed below have explanations that are more difficult to quantify, such as "some reduction in constriction degree or duration" (Kirchner 1998: 1).

Both open questions, criteria for group membership, and how to formalize effort reduction, are reiterated in the discussion of the formal and phonetic approaches to lenition (§3.1 and §3.2).

1.1 *Degemination*

Degemination is the shortening of a CC cluster, where both consonants are the same, to a singleton C (see also CHAPTER 37: GEMINATES). Two examples are diachronic degemination in Numic *kk > k / V __ V (Manaster Ramer 1993) and word-final degemination in Tiberian Hebrew.

(1) *Word-final degemination in Tiberian Hebrew* (Malone 1993: 73)

[qal] 'light (MASC)' [qallɔɔ] 'light (FEM)'

1.2 *Deaspiration*

Deaspiration is the loss or reduction of aspiration. For example, in the Pattani dialect of the Sino-Tibetan language Lahaul, aspiration of the bilabial voiceless stop is reduced in pre-accented syllables. In medial and final contexts, [p^h] is in free variation with [p].

(2) *Deaspiration in the Pattani dialect of Lahaul* (Sharma 1982: 48)

a. Reduction of aspiration
   i. aspiration in accented syllable p^{h}uka 'body'
   ii. reduced aspiration in pre-accented syllable p^{h}ukan 'four'

b. [p^h] ~ [p]
   i. ɖəgegp^{h}i ~ ɖəgegpi 'to tremble, shiver'
   ii. hjup^{h}ʧi ~ hjupʧi 'to open'

1.3 *Voicing*

Voicing involves a change from a voiceless sound to a voiced one and is a very common lenition process, second in prevalence only to spirantization (Gurevich 2004). Voicing usually affects whole series of sounds in the language where it applies. And although voicing can affect fricatives as in Sekani, a Na-Dene language spoken in Canada (3), where voiceless initials of noun and postposition stems voice when prefixed or preceded by a nominal possessor or object, it is much more common with stops, such as the intervocalic voicing in the Yanomam language Sanuma (4).

(3) /s ɬ ç x ʍ/ → [z l ʝ ɣ w] *in Sekani* (Hargus 1985: 270–271)

a.	xàs	'planning tool'	səɣàsè	'my planning tool'
b.	çən	'song'	səʝənè	'my song'
c.	xàz	'windfall roots'	tse ɣàz-e	'Old Friend Mt. (*roots* stem)'

(4) /p t ts k/ → [b d dz g] / V __ V *in Sanuma* (Borgman 1990: 220)

a.	ipa	[ipa] or [iba]	'my'
b.	hute	[hute] or [hude]	'heavy'
c.	hatsa	[hatsa] or [hadza]	'deer'
d.	ãka	[ãka] or [ãga]	'tongue'

1.4 *Spirantization*

Cross-linguistically, this is by far the most common lenition process. Spirantization involves the change of a stop to a fricative, most commonly in intervocalic position. This quite often affects whole series of sounds in a language's inventory, as for example in the Paya dialect of the Chibchan language Kuna (5), or in the Tümpisa dialect of the Uto-Aztecan language Shoshone (6).

(5) b d g → β ð ɣ / V __ V *following a stressed syllable in Paya Kuna* (Pike *et al.* 1986: 459)

a.	paβa	'father'
b.	peðe	'you'
c.	naɣa	'foot'

(6) p t c k kʷ → ɸ θ ç x xʷ / V __ [voiceless]V *in Tümpisa Shoshone* (Dayley 1989: xxviii–xxix)

a.	wisipin	[wiʃiɸi̥]	'thread'
b.	tapettsi	[taβetʧi̥]	'sun'
c.	citoohin	[ciðo:hi̥]	'push'
d.	petɨ$^{\text{-gem}}$	[peðɨ] ~ [peθɨ̥]	'daughter' (ends with a geminating segment)
e.	katɨ$^{\text{-gem}}$	[kaɾɨ] ~ [kaɾ̥ɨ̥]	'sit'
f.	miʔakwa	[mĩʔaɣʷa] ~ [mĩaɣʷa]	'go away!'

1.5 *Flapping*

Flapping is a process whereby a sound is replaced by a flap (usually either alveolar [ɾ] or retroflex [ɽ]; see CHAPTER 113: FLAPPING IN AMERICAN ENGLISH). Two examples are intervocalic flapping of the trill in the Sino-Tibetan language Kagate spoken in the village of Phedi (7) and final flapping of the retroflex stop in the Indo-European language Gujarati (8).

(7) *Flapping in Kagate:* /r/ → [ɾ] / V __ V (Hoehlig and Hari 1976: 19)

a.	/tari/	[târi]	'axe'
b.	/tihriŋ/	[tĭɾiŋ]	'today'
c.	/guhri/	[guɾi]	'cat'

(8) *Free variation of* [ɖ] *and* [ɽ] *in final position in Gujarati* (Cardona 1965: 24)

[ɟʰaɖ] ~ [ɟʰaɽ] 'tree'

Quite often the alveolar and retroflex stops undergo flapping as part of a more general spirantization process that affects other stops. For example, in the Calabar-Creek Town dialect of Efik (a Niger-Congo language) /b d k/ → [β ɾ ɣ] non-initially in a word stem and before a vowel: where /b/ and /k/ spirantize, the alveolar /d/ flaps (9). In the case of the retroflex, the [ɽ] appears to be the usual output of a process where the rest of the stop series spirantize. For example, in the Afro-Asiatic language of Somali, /b d ɖ g/ spirantize to [β ð ɽ ɣ] intervocalically, especially after a stressed syllable: where /b d g/ spirantize, the /ɖ/ flaps (10). In the case of the alveolar stops, however, there is a discernible pattern: spirantization of the stop series in languages whose inventories include a phonemic trill usually results in a /d/ → [ð] substitution, while in trill-less languages the alveolar stop flaps (/d/ → [ɾ]) where the rest of the stop series spirantizes. This happens with regularity in all 33 languages where alveolar stops are affected (Gurevich 2004).

(9) *Non-initial in a word stem and before a vowel in the Calabar-Creek Town dialect of Efik* (Dunstan 1969: 38)

a.	/b/	→	[β]	dwɔ̀β-èbà	'twelve'
b.	/d/	→	[ɾ]	íkɔ̥́ɾ-ẹ̀kpẹ̀nẹ̀	'name of a town'
c.	/k/	→	[ɣ]	úfɔ̥̀ɣ-údwà	'market store'

(10) *Spirantization of stops in Somali* (Armstrong 1964)

a.	'laba	['laβa]	'two'
b.	'badag	['baðag]	'goose'
c.	'tiɖi	['tiɽi]	'she said'
d.	'sagaal	['saɣaal]	'nine'

1.6 Debuccalization

Debuccalization is the loss of place of articulation, preserving only glottal constriction, resulting most commonly in either [h] or [ʔ] (see also CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION). For example, in the Cuisnahuat dialect of the Uto-Aztecan language Pipil, a syllable-final /w/ becomes [h] in word-final or preconsonantal position (11). In the Austronesian language Toba Batak, preconsonantal voiceless stops surface as glottal stops (12).

(11) *Debuccalization in the Cuisnahuat dialect of Pipil:* /w/ – /h/ (Campbell 1985: 34)

a.	kuwa	'to buy'	kuhki	'bought'
b.	puwa	'to count'	puhki	'counted'

(12) *Debuccalization in Toba Batak:* /p t k/ → [ʔ] / __ C (Hayes 1986: 341)

halak 'person' halaʔ batak 'Batak person'

1.7 Gliding

Gliding is the replacement of stops or spirants with a homorganic glide (see CHAPTER 15: GLIDES). For example, in the Djapu dialect of the Australian language Yolngu, [ḓ ɟ] ~ [j], [b g] ~ [w] in word-medial position following a vowel, liquid, or semivowel (13).

(13) *Gliding in the Djapu dialect of Yolngu* (Morphy 1983: 29)

a. minj'ci 'colour, paint' + ḓarpu-NG 'pierce'
minj'ci-jarpu-NG 'paint'
b. ɟarakaḻa'ju-N 'move in an uncontrolled way'
ɟarakaḻa'-jarakaḻaju-N 'keep moving in an uncontrolled way'
c. ḓaː 'mouth' + birka'ju-N 'try'
ḓaː-wirka'ju-N 'ask'
d. ḓawaḻ 'country' + gujaNi-∅ 'I think'
ḓawaḻ-wujaNi-∅ 'be born'

1.8 Loss

Loss is the deletion of a sound (most commonly a glide or a glottal) in certain contexts (see CHAPTER 68: DELETION), for example, the occasional loss of the intervocalic glottals [ʔ] and [h] in the Uto-Aztecan language Tümpisa Shoshone (14), where the presence and absence of these sounds is in free variation.

(14) *[ʔ h] ~ ∅ in Tümpisa Shoshone* (Dayley 1989: xxix)

a.	miʔakwa	[mĩʔaɣʷa] ~ [mĩaɣʷa]	'go away!'
b.	soʔoppɨtin	[soʔɔpːiɾɨ] ~ [sɔːpːiɾɨ]	'much, many'

1.9 Devoicing

Devoicing is the loss of voicing, usually in final positions (see CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). For example, the final devoicing of obstruents in Standard Bulgarian (15):

(15) *Final devoicing in Bulgarian* (Scatton 1984: 73)

a.	grad-ove	'cities'
b.	grat	'city'

2 Patterns of lenition

The examples in §1 touch on the cross-linguistic prevalence of some leniting phenomena, how common they are among the world's languages, and how widespread their effect may be within the languages where they apply. In this section I elaborate on a few other patterns that emerge from the list of leniting processes. The degree to which each theory of lenition discussed in the following section accounts for these patterns provides added perspective into the differing points of view.

The processes described above operate in two main contexts: syllable/word-finally and intervocalically. The bulk of the processes, those that apply in intervocalic context, line up in a discernible sequence: the products of degemination and deaspiration (/tt tʰ/ → [t]) are the sounds that undergo voicing (/t/ → [d]), resulting in the sound that most commonly undergoes spirantization, flapping, debuccalization, or gliding (/d/ → [ð ʔ/h ɾ j]), and glides and glottals are the sounds most commonly lost (/j ʔ h/ → Ø). This chain shift alignment of intervocalic lenition processes, where the output of some is the exact input of others, is illustrated in (16). In (17), the processes are listed in the order they could apply if they were to affect the same phoneme or its correspondent diachronically, although in some cases like spirantization and flapping or debuccalization and gliding, this order is arbitrary.

(16) *Hierarchy of input/output sounds in intervocalic lenition processes*

tt tʰ > t > d > (θ) ð ɾ > ʔ h j > Ø

(17) *The general order in which intervocalic lenition processes might apply*

Degemination	tt → t
Deaspiration	tʰ → t
Voicing	t → d
Spirantization	t d → (θ) ð
Flapping	t d → ɾ
Debuccalization	t → ʔ h
Gliding	t → j
Loss	ʔ h j → Ø

If the oft-cited observation that "a segment X is said to be weaker than a segment Y if Y goes through an X stage on its way to zero" (Venneman, cited in Hyman 1975: 165) is an accurate diagnostic of intervocalic consonant strength, then (16) lists consonants in order of their relative strength, from strongest to weakest (where [tt] is stronger than [t], which is stronger than [d], etc., to the weakest possible outcome of lenition, which is Ø). The resulting "weakening hierarchy" gives birth to the notion of lenition as gradation toward loss.

There are two ways in which the patterns illustrated in (16) and (17) manifest themselves in language data: as an outline of attested diachronic sound changes of the same phoneme or its correspondent, and as a list of synchronic sound substitutions that occur, often simultaneously, in any given language. Two examples of attested diachronic sound changes where some of the lenition processes listed in (17) sequentially affect the same phoneme are French, where intervocalic stops were voiced then spirantized before eventual gliding (not shown in the example) and deletion (18), and Latin, where there was a similar intervocalic voicing, spirantization/gliding, then loss (19).

(18) *French lenition:* t → d → ð → Ø (Jacobs 1994: 2)

fratrem > *[fradre] > [fraðre] > *frère* 'brother'

(19) *Latin lenition* (Hock 1991: 81)

pacatum > (*)*pagado* > Sp. [paɣaðo] > dialectal [paɣao] 'pacified, pleased'

Lenition processes that operate simultaneously in a given language may also exhibit the pattern in which the output of one process is the exact input of another, but in some cases these actually affect different phonemes. For example, in a given language there could be synchronic voicing of voiceless stops and spirantization of voiced ones. In such cases the chain-shift pattern of [t] → [d] and [d] → [ð] is maintained, but the phone [d] that is the output of voicing and the phonetically comparable [d] that undergoes spirantization do not represent the same phoneme. Two examples of such *phonemic overlap*, an intersection of phonemes where "a given sound [. . .] may belong to two or more different phonemes in the same dialect" (Bloch 1941: 93), are the interactions between intervocalic voicing and spirantization in Northern Corsican, and between debuccalization and loss in Nepali. In Northern Corsican, intervocalic voiceless stops are voiced (20), while existing voiced stops in the language are spirantized in the same context (21). The phonetic contrast between intervocalic voiceless and voiced stops shifts to spirantization and is maintained. In the Indo-European language Nepali there is intervocalic debuccalization of [ts^h] to [h] (Bandhu and Dahal 1971) and in the same context loss of the existing [h] in normal speech (22). So while the chain-shift pattern of [ts^h] → [h] and [h] → Ø is maintained, the output of the first process and the input of the second are not the same phoneme, and the previous contrast between the phonemes /ts^h/ and /h/ is preserved, and shifts to /h/ and Ø.

(20) [p t k] → [b d g] / V __ V *in Northern Corsican* (Dinnsen and Eckman 1977: 6)

a.	[peðe]	'foot'	[u beðe]	'the foot'
b.	[tengu]	'I have'	[u dengu]	'I have it'
c.	[kaza]	'house'	[a gaza]	'the house'

(21) [b d g] → [β ð ɣ] / V __ V *in Northern Corsican* (Dinnsen and Eckman 1977: 6)

a.	[bokka]	'mouth'	[a βokka]	'the mouth'
b.	[dente]	'tooth'	[u ðente]	'the tooth'
c.	[gola]	'throat'	[di ɣola]	'of throat'

(22) *Intervocalic loss of* /h/ *in Nepali* (Bandu and Dahal 1971: 26)

a.	/bəhiro/	[bəiro]	'deaf'
b.	/məhə/	[məə]	'honey'

The discussion of phonemic overlap brings up two additional features common to most lenition processes. First, Nepali intervocalic /h/-loss is reported to occur in *normal* speech. This added dimension to what we know about the context of many lenition processes is rather common cross-linguistically, and many leniting sound substitutions are reported mostly to occur in *relaxed*, *fast*, and *normal* speech (see CHAPTER 79: REDUCTION).

Second, the matter of contrast maintenance is raised (see CHAPTER 2: CONTRAST). In both the Northern Corsican and the Nepali examples the phonetic distinction between two phonemes is threatened by a leniting sound substitution (in Corsican the /p/:/b/, /t/:/d/, /k/:/g/ oppositions are threatened by the voicing of /p t k/, and in Nepali the /ts^h/:/h/ opposition is threatened with the debuccalization of /ts^h/ to [h]). In both cases, however, the distinctions are maintained by an additional leniting sound substitution in the same context (in Corsican the intervocalic spirantization of the existing voiced stops shifts the /p/:/b/, /t/:/d/, /k/:/g/ contrast to /b/:/β/, /d/:/ð/, /g/:/ɣ/, and in Nepali the loss of intervocalic /h/ shifts the /tsʰ/:/h/ contrast to /h/: Ø). As it turns out, lenition phenomena in general very rarely lead to neutralization, and almost never result in homophonic forms. A survey of 230 mostly leniting processes in 153 languages found that 92 percent avoid neutralization, while only 8 percent could potentially result in the kind of homophony that leads to the loss of lexical distinction that may interfere with communication (Gurevich 2004).

In summary, several ways in which lenition behaviors pattern together have been outlined in this section. These are listed in (23).

(23) *Patterns of lenitions*

a. The prevalence of certain processes (that is, how common some lenition processes are cross-linguistically and whether they affect a single sound or an entire series).
b. The fact that many lenition processes are reported to occur in *natural* or *fast* speech.
c. The gradation pattern of intervocalic lenition processes.
d. Phonemic overlap.
e. The strong tendency of lenition phenomena to avoid neutralization.

Current debates on lenition focus on formulating a unified description of all leniting processes, isolating what exactly motivates them and in some cases what constrains them. Exploring how and the degree to which each approach accounts for the patterns presented here sheds light on the foundation of each theory and its capacity to accommodate empirical data.

3 Theoretical approaches to lenition

There is somewhat of a general agreement among phonologists about the main processes that can be considered leniting, but formalizing this agreement has proved controversial. In this section, three main approaches to the question are explored. The formal approach (§3.1) seeks to define unified rules that would encode all vital information about lenition processes. These rules should form a model that can be used to determine unambiguously which processes are leniting, including all those that are and excluding all those that aren't. The phonetic approach (§3.2) seeks to isolate the underlying physical causes of all lenition processes. The third approach (§3.3) builds on the contrast-maintaining behavior of most lenition processes to identify the forces that may constrain the progress and outcome of such sound changes.

3.1 The formal approach

The formal approach to lenition is taken by generative grammarians. Its goal is to formalize a set of purely synchronic rules that would model all cases of lenition. Three notable formalizations of lenition under this approach are *feature spreading*, *sonority promotion*, and *simplification*. Additional formalizations exist, but they are mostly variations on these three models. All three models attempt to subsume the various sound changes that can be considered leniting, while excluding all other processes, under one formal expression. Success in this endeavor would result in a rule to be included in Universal Grammar (UG).

Lenition as autosegmental *feature spreading* (e.g. Jacobs and Wetzels 1988) involves the spreading of some feature of the surrounding sounds to the element undergoing lenition (see CHAPTER 81: LOCAL ASSIMILATION), for example the spreading of either the [+voiced] or [+continuant] feature of the vowels surrounding an intervocalic stop to that stop, causing it to either voice or spirantize. This formalization works well for voicing and spirantization, but not for the leniting process of debuccalization which, if anything, involves the delinking of features rather than the acquisition of new ones. Additional rules would be required to predict when a stop is voiced and when it is spirantized, since the surrounding vowels possess both features.

Lenition as *sonority promotion* (e.g. Hock 1991; Lavoie 1996) formalizes lenition rules as replacing a sound by a more sonorous version of itself in certain contexts. Sonority is determined based on the principle that "requires onsets to rise in sonority toward the nucleus and codas to fall in sonority from the nucleus" (Kenstowicz 1994: 254; see also CHAPTER 49: SONORITY). On the scale of sonority, stops are least sonorous, followed by fricatives, nasals, liquids, glides, and finally vowels, which are most sonorous. Lenition as sonority promotion is descriptively accurate for some of the leniting processes such as spirantization and gliding, but as a unified formalization of lenition it fails to include other processes commonly considered leniting such as deaspiration and degemination, neither of which has a more sonorous output than input.

A third formal view of lenition is one of *simplification*, where segmental complexity is measured by the number of features required to describe a consonant, and lenition is a process that simplifies this complexity by delinking some of the features. For example, deaspiration would involve the delinking of laryngeal features and debuccalization would delink place of articulation features. When this formalization is faced with a leniting sound change that does not appear to reduce the number of basic features of a given element, it turns to *markedness* for help: *markedness* is used as a measure of some degree of "naturalness," meant to make phonological features less abstract in terms of their intrinsic content (e.g. Chomsky and Halle 1968; Guitart 1976; McMahon 1994; Rice 1999). An element is considered *marked* by definition if it is less natural or more complex than another. In this way, every case that does not immediately conform to the view of lenition as simplification is solved, because within this approach the input of any lenition process, by definition, is more *marked* than the output, hence every lenition process is one of a move to the *unmarked*, or less complex state (see CHAPTER 4: MARKEDNESS).

Let us examine how the formal models presented here account for the patterns of lenition discussed in §2. Although there is not an explicit concern with the prevalence of some lenition processes over others (23a), the frequency of any element in comparison with another can be accommodated within generative theory as part of the UG principles of markedness: the less marked elements are expected to be more frequent. The view of lenition as gradation (23c) is relevant to both the *sonority promotion* and the *simplification* models: both formalize lenition as a move along a graded scale, either of sonority or of segmental complexity.

The sonority scale emulates to a large extent the lenition hierarchy mapped out in (16), and the case of loss especially provides compelling support for lenition as *simplification*. Due to its superlative nature, the output of loss is arbitrarily the least complex, least marked, and most natural segment, which, as the ultimate step in the gradation pattern, gives the impression that it is the goal. The remaining three patterns, fast speech (23b), phonemic overlap (23d), and contrast maintaining behavior of lenitions (23e), are not addressed.

The advent of Optimality Theory (Prince and Smolensky 1993) allows for the incorporation of elements of phonetic and functional detail into the formal grammatical expression of lenition patterns. Kirchner (1998), for example, incorporates notions of "articulatory ease" directly into his formal statement, in the form of so-called "lazy" constraints that are ranked with respect to (presumably "non-lazy") faithfulness constraints. Such a research program provides a very promising link to the phonetic and functional pressures that are demonstrably acting on patterns of lenition. We hold off our investigation of such approaches – despite their formal rigor – until the next section.

To summarize, the formal approach to lenition arises from the generative tradition. Its main goal is to find a unified formal rule that would subsume all the various processes that can be considered leniting. If such a rule is defined, it would be recognized as a linguistic universal, and the question of lenition would be solved. If a model is posited that does not apply to all processes generally considered leniting, then the offending processes are either amended (as in the case of the simplification model's use of *markedness*) or removed from the list of lenitions, using the argument that the model is sound, and if it is sound and does not include X, X must not be relevant. For example, degemination does not fit the view of lenition as "delinking of privative features" (a variation on the *simplification* model), so it is argued that this process should be excluded from the list of unambiguously agreed-on lenition phenomena (Szigetvári 2008).

3.2 The phonetic approach

The second approach is more in the spirit of how the term *lenition* came to be coined: an observation of how common certain processes are in certain environments and that there is some "reduction of articulatory energy" associated with these sound changes. Research in this direction is especially concerned with the physical motivation behind lenition, be it *articulatory effort reduction*, *prosody maintenance*, related to *acoustics* and *perception*, or possibly some combination of these factors.

Ohala (1981) notes that sound changes which are attested in diverse and unrelated languages are likely to have a phonetic origin. The degree to which lenition processes are common cross-linguistically suggests that they are motivated by what is common among speakers, i.e. biological factors which include the physical shape of the vocal organs, their movement, and their acoustic correlates. However, while it is widely assumed that lenition is conditioned by phonetic properties such as ease of production and perception (e.g. Flemming 1995; Jun 1995; Kirchner 1998; Steriade 2000), how exactly these properties should be described and measured is far from settled.

The concept of *effort minimization* precedes formal generative theories: Zipf, as early as 1935, suggests that the frequency of sounds depends on their degree of articulatory complexity. Rejection of this concept, based on the fact that effort is difficult to measure, was swift: Trubetzkoy (1939) argues that it is difficult to pinpoint the degree of complexity of sounds (e.g. which is more complex: tense vocal cords and relaxed mouth organs, or lax vocal cords and tense mouth organs?). Trubetzkoy's objection to describing sounds in terms of phonetic complexity leads to the introduction of *markedness* values and the eventual move toward formalizing phonology rules completely removed from phonetic information. But interest in characterizing lenition in terms of effort minimization has not waned.

Kirchner's (1998) phonetically based approach posits an effort minimization model of lenition where greater articulatory movement constitutes greater effort, and the push to reduce this effort results in the reduction of constriction degree or duration of an affected sound. Kirchner incorporates phonetic theory into a formal approach and models it within the framework of Optimality Theory (Prince and Smolensky 1993), where conflicting universal constraints are ranked with respect to each other. Articulatory effort minimization is therefore identified as a constraint (LAZY) that is ranked with respect to the counter-force constraints of *faithfulness* and *fortition*. Lenition is thus viewed as a force, encoded in universal grammar, to reduce articulatory effort by reducing articulatory movement and timing which results in falling short of articulatory targets.

Kingston argues that:

> the differences in effort between the lenited and unlenited pronunciations are so miniscule that they can hardly be what motivates a speaker to lenite. Both the differences in the distance the articulators travel (mere millimeters) and the time scales (at most tens of milliseconds) are much too small for effort to differ detectably between the two pronunciations. (Kingston 2008: 1)

He suggests that lenition's purpose is to *maintain prosodic structure*. He shows that lenition occurs most commonly inside prosodic constituents and argues that it is meant to communicate a continuing constituent, thereby reducing a sound's interruption of the stream of speech. Within this view, lenited pronunciation is the result of achieving a specific target that produces the desired acoustic consequences, such as greater intensity, rather than falling short of the desired target. Kingston also notes that "[l]enition is more likely in more frequent words than less frequent ones, because the listener needs less information to recognize more frequent words" (2008: 17); see also CHAPTER 90: FREQUENCY EFFECTS.

Acoustics and *perception* must also play some role in lenition, even if only as part of the natural interaction between speaker and hearer: "Speaker and hearer are interested in communicating and will pronounce words only as they have heard them (or think they have heard them) pronounced by others" (Ohala 1981: 197); see also CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY. Acoustic considerations may affect the perceived differences between certain sounds in certain contexts, which may facilitate or otherwise influence lenition phenomena. For example, acoustic theory suggests that prevocalic distinctions are more perceptible than preconsonantal ones (e.g. Silverman 1995; Ségéral and Scheer 1999; Steriade 1999). Since lenition processes are most frequent in intervocalic contexts, it is possible that lenition may proceed more easily in contexts where a sound is easier to perceive even when it is lenited. Kaplan (2008) suggests another angle: that in certain contexts the perceptual difference between a sound and its lenited counterpart plays a role in the prevalence of some lenition processes. She has had some success in showing that the perceptual difference between intervocalic voiced stops and spirants is smaller than between voiced and voiceless stops, which is somewhat consistent with the frequency of spirantization, although it is too soon to draw any broad conclusions based on such a limited study. The interaction between articulation and perception, speaker and hearer, also plays a role in the functional approach discussed in the following section.

Turning to the patterns listed in §2, by attributing the sound changes to physical properties common to all speakers, a unified approach to lenition that is based on phonetic considerations accounts for the similarities between, and to some extent the frequency of, cross-linguistic lenition processes (23a) and the fact that many leniting changes are reported to occur in *relaxed* or *fast* speech (23b). In fact, it is these empirical observations that suggest lenition phenomena may be phonetically driven in the first place. The view of lenition as gradation (23c), and especially the weakening hierarchy in (16), are central to effort-reduction theories. As with some formal models discussed above, the case of loss provides compelling support for the view of lenition as a graded move along a scale of effort minimization: loss is the ultimate step in lenition and it results in an element that unambiguously requires the least effort to articulate. Phonemic overlap (23d) is not addressed, but the strong tendency of lenition phenomena to avoid neutralization (23e) is inherent in Kingston's observation that lenition is more likely to affect more common words which rely less on acoustic cues to be recognized. Lenition is less likely to proceed unhindered and have widespread consequences if it obliterates meaning distinctions to the point where it interferes with communication (Gurevich 2004). This is also supported by the acoustic studies that suggest distinctions are more perceptible in prevocalic contexts, where most lenition phenomena occur. I return to this point in the following section.

3.3 *The functional approach*

The functional approach focuses on the effect that leniting sound substitutions have on contrasts in the languages where they apply, that is, the fact that lenition phenomena very rarely result in obliteration of contrast (neutralization). This is the goal of Gurevich (2004), who investigates 230 mostly leniting processes in 153 languages, and finds that 92 percent of these avoid neutralization. This approach is termed *functional* because the meaning distinction of words depends on the system of contrasts in a given language, and these distinctions affect communication, which is the primary *function* of language.

This approach to lenition is not independent of phonetically based motivation, and does not explicitly contradict any of the models presented in the previous section. It does, however, influence the extent to which physical properties can drive sound changes. If there is a physical "push" – whatever that push may be – to lenite, the degree to which this substitution may affect contrast in a given language may hinder the progress of the sound change. Lenition processes that threaten contrast may lead to loss of lexical distinctions which, in turn, could induce a significant amount of homophony that would result in confusion. And since confusing signals are less likely to be reproduced as listeners become speakers (Silverman 2006), the neutralizing sound substitutions, regardless of the physical "push" to invoke them, are less likely to become widespread.

Gurevich (2004) clearly shows that the progress and outcome of lenition processes are constrained by the functional considerations of contrast maintenance. The systems of contrast in languages appear to exert a gradual diachronic force over phonetic processes, affecting the progress and outcome of such processes depending on the degree to which they threaten contrast. Lenition processes that do not threaten contrast are far more likely to proceed unhindered, with the widespread consequences of affecting entire series of sounds in a language (this most commonly happens with voicing and spirantization). Changes that somewhat threaten contrast pattern according to the shapes of the phonemic inventory of a language. An example of this is the flapping that patterns with spirantization discussed in §1.5: in languages where the entire stop series is spirantized, the retroflex stop [ɖ] always results in a flap [ɽ], while the alveolar stop [d] results in a flap [ɾ] only in languages that do not have a phonemic trill; otherwise it spirantizes to [ð]. This suggests that flapping may be the preferred outcome of alveolar spirantization, as in the case of the retroflex stops, but is avoided when contrast is threatened. This threat comes in the form of a phonemic trill, a sound that is phonetically similar to a flap (the /r/:/ɾ/ contrast exists, but is rare; it is found in only three of the 153 languages investigated). Finally, changes that clearly threaten contrast, such as sound mergers and loss, often induce further changes that reshape phonological systems, thereby avoiding contrast obliteration: for example, cases of phonemic overlap discussed in §2 (e.g. in (20), where [p t k] → [b d g], and (21), where [b d g] → [β ð ɣ] intervocalically in Northern Corsican, which results in the avoidance of neutralization between the output of the voicing process and existing voiced stops in the language), or contrast shifts, where a sound may be deleted but its absence maintains contrast with Ø, an example of which is provided below in (24).

Phonetically conditioned sound changes that are found to be neutralizing (18 of the 230 processes, or 8 percent) are more common in preconsonantal positions where contrast is less perceptible and harder to maintain (see again Silverman 1995, Ségéral and Scheer 1999, and Steriade 1999). That is, the potential obliteration of contrast to the extent where it could interfere with lexical distinctions occurs more frequently in contexts where contrast is already less perceptible, in which case fewer words should depend exclusively on such contrasts for their distinctions. Hence the potential of these few neutralizing sound substitutions to hinder communication by inducing homophony is reduced, which further suggests that the relationship between leniting sound substitutions and contrast is not arbitrary.

Turning again to the patterns listed in §2, the cross-linguistic similarities of lenition processes are accounted for by relying on phonetically based motivation, and, in addition, the prevalence of certain processes, and possible wide-reaching consequences of affecting entire series of sounds within a language's inventory (23a), are accounted for directly by the degree to which these sound substitutions affect contrasts in a given language. That is, the less likely a sound substitution is to induce homophony, the more likely it is to proceed unhindered and have a widespread effect on the sound system of a language, as is in fact the case for both voicing and spirantization. Voicing, which comprises 39 cases of the 230 investigated, never results in neutralization and is not only common cross-linguistically, but is also most likely to affect whole series of sounds in the languages where it applies. Spirantization, of which there are 76 cases, is 95 percent non-neutralizing, is the most common cross-linguistic form of lenition, and is also likely to affect entire series of sounds.

The fact that lenitions commonly occur in *natural* or *fast* speech (23b) is not explicitly addressed within the functional approach, except as a natural consequence of its reliance on phonetic motivation as the force that drives most lenitions. The gradation pattern of intervocalic lenition processes (23c) has no implications for the functional approach. Interestingly, although the effect of contrast considerations on lenition is not related to the view of such processes as a trajectory of consonants toward their weakest state, the case of loss – which provides compelling support for the gradation view of several models discussed above – also plays a key role here. This final step in the view of lenition as erosion also has a superlative consequence for contrast: loss always results in phonetic neutralization because the elimination of a phoneme in some context always obliterates the phonetic distinction between that phoneme and Ø. But meaning distinctions are actually preserved in 71 percent of the cases of loss in the corpus of 230 processes. The most common ways in which contrast is maintained in such cases are phonemic overlap, like the examples in (20)–(22) above, and contrast shifts such as the one in Bulgarian (24), where nasals are lost between vowels and fricatives, followed by nasalization of the vowel. Here the /n/ is lost, but the nasalization of the preceding vowel maintains the distinction between /n/ and Ø.

(24) *Loss of nasal and nasalization of vowel in Bulgarian* (Scatton 1984: 57)

/onzi/ [õzi] 'that'

Phonemic overlap (23d) is central to the functional approach. It shows how leniting processes that threaten contrast may induce further changes, and suggests that some leniting changes may actually be, at least partially, triggered by contrast maintenance. Finally, the strong tendency of lenition to avoid neutralization (23e) is, of course, the basis of the functional approach, which posits that this tendency is what constrains the progress and outcome of lenitions. One could question the significance of a pattern that is so important to one approach but is not addressed by others. However, it is a pattern that is extremely prevalent and something that lenitions have more in common than almost any other characteristic. Its omission from all formal and most phonetically based models stems in part from the fact that this pattern had not previously been investigated, and in part from the general belief that, because contrasts are language specific, they have no place in universal grammar.

4 Conclusion

Leniting sound changes are common and exhibit similar cross-linguistic behaviors – this much we know. How to formalize this information in a unified model of lenition that all phonologists can agree on has so far eluded us. Depending on one's approach, such strong cross-linguistic similarities must be encoded in the UG and/or grounded in the physical properties of both speakers and hearers. Current debates range from the generative approach of how to encode lenition as purely phonological synchronic universal rules, to isolating the physical driving force that motivates these sound changes, to how the systems of contrasts in a given language play a role in constraining the progress and outcome of lenition processes.

The various models of lenition presented in this chapter have differing approaches to the questions of how to delimit the collection of processes that almost everyone agrees are related. Five general patterns of lenitions – all based to some extent on empirical data – are identified. The relative significance of each tendency depends on the theory one supports. Below is the summary of each pattern's role within the differing approaches to lenition.

(a) The *prevalence* of certain lenition processes as well as their overall cross-linguistic similarity: cross-linguistic frequency of certain lenitions is implicit in the *markedness* constraints of formal models, where the less marked element is more natural and therefore more frequent, and explicit in the phonetic approach's reliance on physical properties which are common to all speakers and hearers. The prevalence of some processes over others, as well as how widespread their consequence may be, is predicted by the contrast-maintenance considerations of the functional approach, where the less a process threatens contrast the more prevalent it is.

(b) The fact that many lenitions are reported to occur in *natural* or *fast* speech: this pattern is not addressed by the formal models but is inherent in the view of lenition as motivated by physical properties and therefore works with both the phonetic and the functional approaches.

(c) The *gradation pattern* of intervocalic lenition processes: this pattern emerges from the juxtaposition of diachronic lenition processes, and is critical to the sonority scale and simplification models of lenition under the formal approach, and to the view of lenition as consonant erosion within the effort-minimization model. This view of lenition is crucial to the characterization of lenitions as all part of one general process, a notion that is most advantageous to theories concerned with building a UG, whose hypothesized existence is directly tied to encoding rules in their most general and simple manner.

(d) The *phonemic overlap* pattern of some lenitions: this is a pattern that often emerges in languages where two synchronic lenition processes interact in the sense that the output of one is phonetically similar to the input of another. It is only addressed by the functional approach, which views such cases as an indication that the threat of contrast obliteration may trigger additional lenitions.

(e) The strong tendency of lenition phenomena to *avoid neutralization*: this recently identified tendency is the foundation on which the functional approach is built. Since this pattern is based on contrast considerations, which are language-specific, formal models do not address it in their search for universals.

REFERENCES

Armstrong, Lilias E. 1964. *The phonetic structure of Somali*. Ridgewood, NJ: Gregg Press.

Bandhu, C. M. & B. M. Dahal. 1971. *Nepali segmental phonology*. Kirtipur: Tribhuvan University.

Bloch, Bernard. 1941. Phonemic overlapping. *American Speech* 16. 278–284. Reprinted in Martin Joos (ed.) 1957. *Readings in linguistics: The development of descriptive linguistics in America since 1925*, 93–96. Chicago: Chicago University Press.

Borgman, Donald M. 1990. Sanuma. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.) *Handbook of Amazonian languages*, vol. 2, 15–248. Berlin & New York: Mouton de Gruyter.

Campbell, Lyle. 1985. *The Pipil language of El Salvador.* Berlin: Mouton.
Cardona, George. 1965. *A Gujarati reference grammar.* Philadelphia: University of Pennsylvania Press.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.
Dayley, Jon P. 1989. *Tümpisa (Panamint) Shoshone dictionary.* Berkeley: University of California Press.
Dinnsen, Daniel A. & Fred R. Eckman. 1977. Some substantive universals in atomic phonology. *Lingua* 45. 1–14.
Dunstan, Elizabeth (ed.) 1969. *Twelve Nigerian languages.* New York: Africana Publishing Corporation.
Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles.
Guitart, Jorge M. 1976. *Markedness and a Cuban dialect of Spanish.* Washington, DC: Georgetown University Press.
Gurevich, Naomi. 2004. *Lenition and contrast: The functional consequences of certain phonetically conditioned sound changes.* New York & London: Routledge.
Hargus, Sharon. 1985. The lexical phonology of Sekani. Ph.D. dissertation, University of California, Los Angeles. Published 1988, New York: Garland.
Hayes, Bruce. 1986. Inalterability in CV phonology. *Language* 62. 321–351.
Hock, Hans Henrich. 1991. *Principles of historical linguistics.* 2nd edn. Berlin & New York: Mouton de Gruyter.
Hoehlig, Monika & Maria Hari. 1976. *Kagate phonemic summary.* Kathmandu: Institute of Nepal and Asian Studies, Tribhuvan University.
Hyman, Larry M. 1975. *Phonology: Theory and analysis.* New York: Holt, Rinehart & Winston.
Jacobs, Haike. 1994. Lenition and Optimality Theory. Unpublished ms., University of Nijmegen (ROA-127).
Jacobs, Haike & Leo Wetzels. 1988. Early French lenition: A formal account of an integrated sound change. In Harry van der Hulst & Norval Smith (eds.) *Features, segmental structure and harmony processes,* part I, 105–129. Dordrecht: Foris.
Jun, Jongho. 1995. Perceptual and articulatory factors in place assimilation: An Optimality Theoretic approach. Ph.D. dissertation, University of California, Los Angeles.
Kaplan, Abby. 2008. Perceptual, articulatory, and systemic influences on lenition. Unpublished ms., University of California, Santa Cruz.
Kenstowicz, Michael. 1994. *Phonology in generative grammar.* Cambridge, MA & Oxford: Blackwell.
Kingston, John. 2008. Lenition. In Laura Colantoni & Jeffrey Steele (eds.) *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology,* 1–31. Somerville, MA: Cascadilla Press.
Kirchner, Robert. 1998. An effort-based approach to consonant lenition. Ph.D. dissertation, University of California, Los Angeles.
Lavoie, Lisa. 1996. Consonant strength: Results of a data base development project. *Working Papers of the Cornell Phonetics Laboratory* 11. 269–316.
Malone, Joseph L. 1993. *Tiberian Hebrew phonology.* Winona Lake, IN: Eisenbrauns.
Manaster Ramer, Alexis. 1993. On lenition in some Northern Uto-Aztecan languages. *International Journal of American Linguistics* 59. 334–341.
McMahon, April. 1994. *Understanding language change.* Cambridge: Cambridge University Press.
Morphy, Frances. 1983. Djapu, a Yolngu dialect. In R. M. W. Dixon & Barry J. Blake (eds.) *The handbook of Australian languages,* vol. 3, 1–188. Amsterdam: John Benjamins.
Ohala, John J. 1981. The listener as a source of sound change. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 17(2). 178–203.

Pike, Eunice V., Keith Forester & Wilma J. Forester. 1986. Fortis versus lenis consonants in the Paya dialect of Kuna. In Benjamin F. Elson (ed.) *Language in global perspective: Papers in honor of the 50th anniversary of the Summer Institute of Linguistics 1935–1985*, 451–464. Dallas: Summer Institute of Linguistics.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Rice, Keren. 1999. Featural markedness in phonology: Variation. *GLOT International* 4.7: 3–6, 4.8: 3–7.

Scatton, Ernest A. 1984. *A reference grammar of modern Bulgarian*. Columbus: Slavica.

Ségéral, Philippe & Tobias Scheer. 1999. The coda mirror. Unpublished ms., Université Paris 7 & Université de Nice.

Sharma, D. D. 1982. *Studies in Tibeto-Himalayan linguistics: A descriptive analysis of Pattani (a dialect of Lahaul)*. Hoshiarpur: Panjab University.

Silverman, Daniel. 1995. Phasing and recoverability. Ph.D. dissertation, University of California, Los Angeles.

Silverman, Daniel. 2006. *A critical introduction to phonology: Of sound, mind, and body*. London & New York: Continuum.

Steriade, Donca. 1999. Alternatives to syllable-based accounts of consonantal phonotactics. In Osamu Fujimura, Brian Joseph & Bohumil Palek (eds.) *Item order in language and speech*, 205–242. Prague: Karolinum Press.

Steriade, Donca. 2000. Paradigm uniformity and the phonetics–phonology boundary. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, 313–334. Cambridge: Cambridge University Press.

Szigetvári, Péter. 2008. What and where. In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.) *Lenition and fortition*, 93–130. Berlin & New York: Mouton de Gruyter.

Thurneysen, Rudolf. 1946. *A grammar of Old Irish*. Dublin: Dublin Institute for Advanced Studies.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.

Westbury, John R. & Patricia Keating. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22. 145–166.

Zipf, George K. 1935. *The psycho-biology of language: An introduction to dynamic philology*. Boston, MA: Houghton Mifflin.

67 Vowel Epenthesis

Nancy Hall

1 Introduction

The term "vowel epenthesis" can refer to any process in which a vowel is added to an utterance. Beyond this simple description, however, vowel epenthesis processes vary enormously in their characteristics, and many aspects of their typology are still not well understood. Accordingly, the empirical focus of this chapter is on the heterogeneity of vowel epenthesis processes.

This chapter is organized around several empirical questions, namely: What is the function or cause of vowel epenthesis (§2)? What determines the location (§3) and quality (§4) of an epenthetic vowel? Do epenthetic vowels differ phonetically or psycholinguistically from lexical vowels (§5)? What distinguishes an excrescent vowel (§6)? How does vowel epenthesis interact with other phonological processes (§7)? Finally, §8 reviews research on epenthetic vowels in loanwords, and revisits some of the previous questions to discuss how the answers may differ in the case of loanwords.

Throughout this chapter, epenthetic vowels are underlined for visual clarity.

2 What is the function/cause of vowel epenthesis?

In most cases, the function of vowel epenthesis is to repair an input that does not meet a language's structural requirements. In particular, vowel epenthesis allows the surfacing of consonants that underlyingly appear in phonotactically illegal contexts. For example, Lebanese Arabic epenthesizes vowels into many CC codas to break up undesirable coda clusters. Epenthesis is more or less obligatory in coda clusters of an obstruent followed by a sonorant, as in (1a), and optional in most other clusters as in (1b) (see Haddad 1984a for a detailed breakdown of coda types).

(1) *Epenthesis in Lebanese Arabic* (Abdul-Karim 1980: 32–33)

a.	/ʔism/	ʔisi̲m	'name'	b.	/kibʃ/	kibʃ ~ kibi̲ʃ	'ram'
	/ʔibn/	ʔibi̲n	'son'		/sabt/	sabt ~ sabi̲t	'Saturday'
	/ʃiɣl/	ʃiɣi̲l	'work'		/nafs/	nafs ~ nafi̲s	'self'

There is controversy over exactly how to analyze the phonotactic requirements that motivate epenthesis. Probably the most popular approach is to assume that epenthesis allows the syllabification of stray consonants (Itô 1989), but Broselow (1982) explores the idea that some epenthesis is simply triggered by particular sequences of consonants, irrespective of syllable structure requirements. Côté (2000) argues that epenthesis is motivated primarily by the need to make consonants perceptible, based on the Licensing by Cue approach of Steriade (1994). For example, one of the main cues that listeners rely on to identify place features of consonants is the formant transitions on neighboring vowels. Hence, a consonant that is not adjacent to a vowel is less easy to identify (see CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS).

In a case like Lebanese, it might be argued that claiming a structural motivation for vowel epenthesis is circular, given that this optional vowel epenthesis is the only evidence that such clusters are marked in this language. But in some languages, vowel epenthesis is only one of a "conspiracy" of processes removing a particular cluster type. In Welsh, for example, codas with rising sonority are repaired through deletion, as in (2a), lenition (2b), metathesis (2c), or vowel epenthesis (2d), while codas with falling sonority are left intact.

(2) *Welsh repair of obstruent–sonorant codas* (Awbery 1984)

a.	/fenestr/	→	feːnest	'window' (southern dialect)
b.	kevn	>	kɛwn	'back' (Pembrokeshire dialect)
c.	sɔvl	>	sɔlv	'stubble' (north-east dialect)
d.	/kevn/	→	keːve̱n	'back' (southern dialect)

The fact that all four processes target the same cluster type supports the idea that this cluster type is marked, and that vowel epenthesis is one of the repairs for the marked structure.

A second common reason for epenthesis is to bring a word up to a certain minimal size. Some languages require each lexical word to have a minimum of two moras or two syllables. Often, roots of smaller size are augmented with an epenthetic vowel, as shown in (3a) for Mono (Banda, spoken in Congo). The epenthetic vowels do not appear when the same roots appear in longer compounds, as in (3b).

(3) *Mono vowel epenthesis* (Olson 2003)

a.	/ʒī/	→	ī̱ʒī	'tooth'	
	/bè/	→	è̱bè	'liver'	
	/mà/	→	à̱mà	'mouth'	
	/ndà/	→	à̱ndà	'house'	
b.	/mà+ndà/	→	màndà	'door'	*à̱màà̱ndà

Metrical structure above the word level can also affect epenthesis. In Galician, vowels are optionally added at the end of an intonational phrase (Martínez-Gil 1997). This is illustrated in (4), where the word *pan* 'bread' can be pronounced with final [i] only if it directly precedes a prosodic break (a–c), not within an intonational phrase (d).

(4) *Epenthesis at intonational phrase boundaries in Galician* (Martínez-Gil 1997)

a. *Ela vai trael-o pan* (~ *pan*[i̲]).
'She's going to bring the bread.'
b. *O pan* (~ *pan*[i̲]), *fixo-no onte.*
'(As for) the bread, (s)he made it yesterday.'
c. *Dille que traia pan* (~ *pan*[i̲]), *non viño.*
'Tell him/her to bring bread, not wine.'
d. *Ela vai trae-lo pan* (**pan*[i̲]) *que comprou.*
'She's going to bring the bread that she bought.'

This epenthesis occurs only with words whose final syllable is stressed: words like ['bo] 'good' and [ka.'fɛ] 'coffee' have the variants ['bo.i̲] and [ka.'fɛ.i̲], but words with non-final stress like ['la.pis] 'pencil' cannot be pronounced *['la.pis.i̲]. Martínez-Gil proposes that the function of this epenthesis is to create a well-formed bimoraic trochee at the edge of each intonational phrase. A similar pattern occurs with optional [ə]-insertion in Parisian French (Fagyal 2000).

A different aspect of phrasal metrical structure affects epenthesis in Dutch. As shown in (5), Dutch has optional schwa epenthesis in coda clusters that consist of a liquid followed by a non-coronal consonant, as well as coda /rn/.

(5) *Dutch* [ə]-*epenthesis* (Booij 1995)

tʏlə̲p	~ tʏlp	'tulip'
hɛlə̲p	~ hɛlp	'help'
hɛrə̲fst	~ hɛrfst	'autumn'
kalə̲m	~ kalm	'quiet'

Kuijpers and van Donselaar (1997) find that speakers are more likely to insert the schwa if this will create a rhythmic alternation of stressed and unstressed vowels. Epenthesizing a schwa in /tʏlp/ changes the word from a single stressed syllable ('σ) to a stressed–unstressed sequence ('σσ) (see also CHAPTER 40: THE FOOT). This happens significantly more often when the first syllable of the following word is stressed than when it is unstressed, as shown in (6).

(6) *Effects of sentence rhythm on epenthesis in monosyllabic words*

context	[ə]-epenthesis	
σσ __ 'σ	50%	['tʏlp] and ['tʏlə̲p] equally preferred
'σσ __ σ	35%	['tʏlp] preferred over ['tʏlə̲p]

Metrical structure above the word level only has gradient effects on vowel epenthesis; there do not seem to be cases of obligatory vowel epenthesis for rhythmic purposes, aside from the minimal word requirement discussed above. Perhaps this is because phrase-level metrical structures themselves tend to show much optionality.

While most analyses of vowel epenthesis focus on structural motivations, there is a little research examining the effects of epenthesis on perception. Van Donselaar *et al.* (1999) bring evidence that vowel epenthesis in Dutch enhances the perceptibility of the consonants adjacent to the epenthetic vowel, particularly the preceding liquid. In lexical decision and phoneme identification tasks, subjects react faster to forms with epenthesis, like [tʏləp], than to forms without epenthesis, like [tʏlp], even though the form without epenthesis is more canonical and closer to the spelling. The authors suggest that speakers epenthesize the vowels to help the listener.

Finally, there are some cases where epenthetic vowels (or, at least, vowels widely described as epenthetic) have no apparent function in terms of phonotactics, metrics, or any other structural requirements. This is seen in Scots Gaelic, where epenthetic copy vowels historically arose in sonorant-obstruent coda clusters following short stressed vowels, as in (7). These vowels are widely analyzed as being still epenthetic today. As discussed further in §5, these vowels are phonetically marked by a special pitch and duration pattern, and they have a number of distinguishing phonological characteristics. Speakers are reported to consider these VRVC sequences monosyllabic, in contrast to other VRVC sequences.

(7) *Scots Gaelic* (Borgstrøm 1937, 1940; Oftedal 1956)

ʃalak	'hunting'
k^{h}ɛn^{yj}ɛ.p	'hemp'

Interestingly, there are many words where one of the consonants that originally triggered the epenthetic vowel has deleted historically, yet the epenthetic vowel has remained – and retained its unique phonetic and phonological characteristics. In the words in (8), the underlined vowel is one that sounds like an epenthetic vowel in terms of pitch and duration, yet synchronically, there is no consonant cluster present to trigger epenthesis. The epenthetic vowel now precedes a word boundary or another vowel, and hence plays no role in terms of improving phonotactics. In fact, it often creates a V.V sequence, which is cross-linguistically dispreferred.

(8) *Unpredictable vowel epenthesis in Scots Gaelic*

mara.i	*marbhaidh*	'will kill'
dɯr^{j}i	*duirgh*	'fishing lines'
ɛn^{yj}ɛ.i	*aithnichidh*	'will recognize'

There are many possible interpretations of such facts. One theory might be that the triggering consonants are present underlyingly and removed through a separate process; another theory is that vowels originally introduced through epenthesis have been reanalyzed as something else (see Hall 2003 for an argument that all "epenthetic" vowels in Scots Gaelic actually reflect a diphthong-like structure in which a vowel and sonorant are phonologically adjoined, and where their articulations overlap so that the same vowel is heard in two pieces). While cases like Scots Gaelic are unusual, they are a reminder that some vowel epenthesis patterns do not seem to have clear structural motivations.

3 What determines the location of an epenthetic vowel?

When vowel epenthesis is used to break up a consonant cluster, there is often more than one location where the vowel could be placed to produce a phonotactically acceptable output. For example, if a language has the syllable structure (C)V(C), hence disallowing CC clusters at the beginning of a word, an initial CCV could be broken up by putting a vowel before the consonants (VC.CV) or between the consonants (CV.CV). In a medial CCC cluster, the vowel could occur before the second or third consonant. The choice of epenthesis locations is language-specific. Arabic dialects, for example, systematically differ in this regard. As shown in (9), "onset" dialects like Egyptian syllabify the second consonant as an onset, meaning that the epenthetic vowel follows the second consonant, while "coda" dialects like Iraqi syllabify the second consonant as a coda, meaning that the epenthetic vowel follows the first consonant (Broselow 1992; Kiparsky 2003; Watson 2007).

(9) *Treatment of /CCC/ in Arabic dialects* (Itô 1989)

Cairene	/ˀul-t-l-u/	ˀul.ti.lu	'I said to him'
Iraqi	/gil-t-l-a/	gi.lit.la	'I said to him'

Temiar (Mon-Khmer, Malaysia) has a much-studied pattern of epenthetic vowel placement in long consonant clusters. Temiar allows only CV and CVC syllables. Given an onset of three or four consonants, Temiar inserts epenthetic vowels to form a string of open syllables terminated by a closed syllable. The epenthetic vowel is a schwa in open syllables; [e] in closed syllables.

(10) *Temiar syllabification* (Itô 1989)

/slɔg/	səlɔg	'sleep, marry (ACT PERF)'
/snlɔg/	senlɔg	'sleep, marry (ACT PERF NOMINALIZED)'
/snglɔg/	sənegl ɔg	'sleep, marry (ACT CONT NOMINALIZED)'

Itô (1989: 241) argues that these patterns of vowel placement can be explained if syllabification is directional. Abstracting away from certain theoretical details, the insight is that languages like Temiar and Iraqi compute maximal syllables starting from the end of the word, while languages like Egyptian compute maximal syllables from the beginning of the word. A stray consonant that could be syllabified more than one way becomes an onset of a following syllable in right-to-left languages, but the coda of a preceding syllable in left-to-right languages, and the placement of the epenthetic vowel varies accordingly.

(11) *Directionality in syllabification*

Left-to-right syllabification	Right-to-left syllabification	
Cairene	*Iraqi*	*Temiar*
/ˀultlu/	/giltla/	/snglɔg/
ˀul.	.la	.lɔg
ˀul.ti.	.lit.la	.neg.lɔg
ˀul.ti.lu	gi.lit.la	sə.neg.lɔg

While directional syllabification works well to explain epenthetic vowel placement in many languages, I will discuss in §8 some cases of loanword adaptation where directional syllabification cannot explain epenthetic vowel placement.

4 What determines the quality of an epenthetic vowel?

The quality of an epenthetic vowel may be determined in one of two ways: it is either a fixed, default quality (which may, of course, be subject to normal allophonic variation according to the language's phonology), or else the quality is determined by some part of the phonological context.

Lebanese Arabic is an example of a language with fixed-quality epenthetic vowels: the epenthetic vowel is always [i]. Different languages have different qualities for their epenthetic vowels, and some qualities are found more commonly than others. Epenthetic [i] and [ə] are especially frequent, but de Lacy (2006: 289) also lists examples of epenthetic [ɨ], [e], and [a]. It is rare for fixed-quality vowels to be [+round], but examples do occur in Quebec French (Martin 1998) and in the Austronesian languages Buol and Kambera (Rice 2008). (There are, of course, also many cases where a basically fixed-quality vowel becomes predictably rounded in some contexts through additional processes such as vowel harmony.)

In "copy vowel epenthesis," the epenthetic vowel must have the same quality as a nearby vowel. In Welsh, for example, final CC clusters are broken with a vowel that is a copy of the preceding vowel. The forms in the left column of (12) illustrate how the epenthetic vowel is absent when a suffix renders the CC cluster non-final.

(12) *Copy vowel epenthesis in Welsh* (Awbery 1984: 88)

gwadne	gwaːdan	'soles, sole'
kɛvne	keːven	'backs, back'
padri	puːdur	'to rot, rotten'
ɔχri	oːχor	'to side, side'

The direction of copying varies by language; both right-to-left and left-to-right copying are well attested.

In rare cases, the quality may relate to more than one nearby segment. In Scots Gaelic, the quality of the epenthetic vowel depends on both the preceding vowel and the preceding consonant. Sonorants in Scots Gaelic contrast for backness. When epenthesis occurs in a /VRC/ sequence where the vowel and sonorant disagree in backness, the epenthetic vowel shares the backness specification of the sonorant (Clements 1986; Ní Chiosáin 1995; Bosch and de Jong 1998; CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS).

(13) *Incomplete vowel copy in Scots Gaelic* (vowel transcription following Ní Chiosáin 1995)

fæɾak	'anger'
inɨxinɣʲa	'brain'
bulʲik	'bellows'
dɨlʲikʲ	'sorry'
mʌrʲev	'dead'

There has been controversy over whether the grammatical mechanisms that allow epenthetic vowels to copy other vowels' quality might be similar to the mechanisms involved in reduplication (where a morpheme copies its segmental content from other segments in the base word; CHAPTER 100: REDUPLICATION). Kitto and de Lacy (1999) argue for a unified theory of the two processes, in which segments in reduplicants and epenthesized segments both have a "correspondence" relation with another segment elsewhere in the word. Kawahara (2007), however, points out a couple of basic differences between these kinds of copying. First, epenthetic copy vowels always copy a vowel in an adjacent syllable, whereas reduplicants may skip adjacent syllables to copy more distant material. For example, in Nakanai (Oceanic; Johnston 1980), a vowel in a reduplicant copies the most sonorous vowel in the base, regardless of its location. Kawahara finds no cases of epenthetic vowels copying distant vowels in this manner. Secondly, copying in epenthetic vowels (especially in loanwords; see §8) is sometimes blocked when particular kinds of consonants intervene, but blocking effects like this are not found in reduplication, where copying can occur over any type of intervening segment. Kawahara proposes that long-distance, correspondence-based copying is available only for morphological operations like reduplication, and that copying of quality in epenthesis always reflects local feature spreading.

5 Do epenthetic vowels differ phonetically or psycholinguistically from lexical vowels?

5.1 *Phonetic characteristics of epenthetic vowels*

There is evidence that in some languages, epenthetic vowels differ articulatorily and acoustically from lexical vowels, and tests that probe speaker intuitions may also find differences. Since these phonetic or psycholinguistic differences may have implications for phonological questions, I will briefly review the evidence.

As shown in (1), Lebanese Arabic optionally inserts an epenthetic vowel in certain CCC or CC# clusters (/mitl/ → [mitil] 'like'). The epenthetic vowel is normally transcribed as [i], but Haddad (1984b: 61) impressionistically notes that "this representation is rather inadequate since an inserted vowel is more prone to suprasegmental features such as 'guttural' and 'emphatic' [pharyngealized] than an underlying vowel is." An acoustic phonetic study by Gouskova and Hall (2009) finds that for some speakers, epenthetic "[i]" is significantly shorter in duration than a lexical [i], and has a lower second formant value. The low F2 indicates that the articulation is relatively back, so that a more appropriate transcription might be [ɨ].

Sometimes the phonetic differences involved in vowel epenthesis are reported to extend over a longer string of the word. The Siouan language Hocank has epenthesis in certain CCV sequences, as in /kre/ → [kere] 'depart returning'. Although no instrumental study has been done, Susman (1943) and Miner (1979) agree that CVCV sequences resulting from epenthesis are audibly shorter in duration than lexical CVCV. The duration difference appears to involve not only the epenthetic vowel, but also the lexical vowel next to it.

Another kind of phonetic difference is reported in Scots Gaelic, where, as shown in (7), epenthesis occurs in certain CC sequences following a short stressed vowel (/tarv/ → [tarav] 'bull'). These epenthetic vowels are often *longer* than lexical vowels in the same position (Bosch and de Jong 1997). The pitch of the resulting

CVCV̲C sequence is distinctive: although a normal CVCVC disyllable has a rise and fall in pitch, Ladefoged *et al.* (1998) show that epenthetic CVCV̲C has only a pitch rise, confirming Oftedal's (1956) description. Speakers are reported to consider such sequences monosyllabic (Oftedal 1956: 29) or "nearly monosyllabic" (Borgstrøm 1940: 153).

Several studies couched in Articulatory Phonology have offered evidence that epenthetic schwa in English differs articulatorily from lexical schwa (see CHAPTER 26: SCHWA). Davidson and Stone (2003) present an ultrasound study of English speakers pronouncing pseudo-Slavic words that began with consonant clusters that are illegal in English, such as /zgomu/. Subjects frequently inserted an audible epenthetic schwa, producing [zə̲gomu]. However, when the articulation of schwa was compared to the lexical schwa of similar words like *succumb* [səkʌm], the tongue position differed significantly. Davidson and Stone suggest that the acoustic schwa does not correspond to a distinct articulatory gesture, but is essentially a transitional sound, the result of a low degree of overlap between the articulatory gestures comprising /z/ and /g/. Smorodinsky (2002) uses EMA to study the epenthetic schwas in English inflectional morphology, and reports differences (though not very robust ones) in tongue position between the epenthetic schwa in *cheate̲d* [ˈʧirə̲d] and the lexical schwa in *cheetah'd* [ˈʧirəd].

Gick and Wilson (2006) give a related analysis of the schwa that many English speakers insert between a high tense vowel and a liquid, as in *fire* (faɪr ~ faɪə̲r). They argue that the schwa sound is not an inserted phonological unit, but an incidental result of the tongue passing through a schwa-like configuration as it transitions between the opposing tongue root positions of the high front vowel and the liquid.

As of yet, few examples of epenthetic vowels have been instrumentally studied, so it is not clear whether epenthetic vowels differ phonetically from lexical vowels in every language. There are plenty of cases where epenthetic vowels are impressionistically described as being acoustically identical to lexical vowels (e.g. Mohawk; Michelson 1989: 40, 48). It is also unknown whether the vowels' phonetic nature correlates with any aspect of their phonological behavior (such as whether the vowel is obligatory or optional, or whether the vowel interacts opaquely with processes like stress assignment). This is likely a rich area for future research.

5.2 *Speaker intuitions about epenthetic vowels*

There are indications that speakers are not always conscious of epenthetic vowels in the same way as lexical vowels. One type of evidence comes from situations where speakers are asked to write their pronunciations phonetically. Pearce (2004: 19) asked speakers of Kera (East Chadic, spoken in Chad, with no tradition of writing) to choose between two possible spellings for acoustically CVCV̲CV words, where the middle vowel was analyzed as epenthetic. The speakers chose CVCCV spellings, suggesting that the middle vowel was not part of their conscious segmentation of the word. On the other hand, when I have asked Lebanese Arabic speakers to write colloquial pronunciations (which are not usually written, as orthography follows Classical Arabic), they do write in the epenthetic vowels. This suggests that speakers' consciousness of epenthetic vowels may differ from language to language.

Van Donselaar *et al.* (1999) argue that in Dutch, where vowel epenthesis is optional ([tʏlp] ~ [tʏlə̲p]), speakers treat the form without epenthesis as canonical. In an experiment, Dutch speakers were asked to perform different language-game-like reversals on monosyllables and disyllables: subjects were to reverse monosyllables segment by segment, changing [tap] to [pat], and reverse disyllables syllable by syllable, changing [hotel] to [tɛlho]. Over 90 percent of words with vowel epenthesis were treated like monosyllables, so that [tʏlə̲p] 'tulip' changed to [plʏt] rather than [lə̲ptʏ]. The authors suggest that speakers have a unitary representation for the forms with and without epenthesis. It might be objected, however, that the experiment is contaminated by orthographic differences between lexical schwa, which is written, and epenthetic schwa, which is not. Another objection, raised by a reviewer, is that [ləptʏ] is not a possible word in Dutch, due to its final lax vowel.

Speakers may be particularly likely to lack awareness of the kind of weak epenthetic vowels often called "excrescent" (discussed further in §6). For example, Harms (1976) reports that Finnish speakers are unaware of an epenthetic schwa that is easily perceived by some non-native speakers:

> [mɛləkein] (*melkein*) 'almost' has essentially the same vowel qualities ([ɛ, ə, ei]) and relative durations as the English verb *delegate* – [dɛləgeit]. From a descriptive phonetic point of view, the Finnish [epenthetic] schwa and the English reduced-vowel schwa represent very nearly identical classes of vowel sounds; i.e., they vary over a wide central area, with their range of variation conditioned by the preceding and following segments. But here the similarity ends. The schwa in the above Finnish forms is purely transitional in nature. Speakers perceive these forms as containing only two syllables, not three.

Few studies of vowel epenthesis have probed the intuitions of native speakers about the vowels, and it would be useful to have data from more languages on how speakers perceive epenthetic vowels, including how the vowels are written, treated in metrics, and treated in language games (see CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY).

6 What distinguishes an "excrescent" vowel?

A number of proposals distinguish a special class of epenthetic vowels often called "excrescent" (Levin 1987) or "intrusive" (Hall 2006). These terms are usually used for vowels that are noticeably phonetically weaker than other vowels. Typically, excrescent vowels are short in duration and centralized in quality. The excrescent vowel may have a quality not present in the language's lexical vowel system; for example, excrescent schwa may exist in a language that otherwise has no schwas. Excrescent vowels are systematically ignored by other phonological processes. The commonly expressed insight is that excrescent vowels are a kind of phonetic effect, likely a transition between consonant articulations.

A classic example of excrescent vowels is the short vowels that occur in consonant clusters in Piro (Arawakan), as shown in (14). Matteson and Pike (1958) note that these vowels differ from the short phonemic vowels of Piro (/i e o a ï/) in several ways. The excrescent vowels are subject to extensive free variation. Their quality can be highly variable, as in /hwï/ below, where the excrescent vowel has been recorded with five different qualities. Also, in some cases the presence of the excrescent vowel varies with "syllabification" of a consonant, as in /whene/ below. The vowels cannot bear any kind of stress, and they are of much shorter duration than lexical vowels. In terms of timing, the authors report that "in the rhythm of a phrase, a consonant plus the transition vocoid corresponds in timing to a single consonant rather than to a sequence of consonant plus vowel." The excrescent vowels fail to block a pattern of co-articulatory rounding that is blocked by other vowels. In Piro orthography, the excrescent vowels are not written.

(14) *Excrescent vowels in Piro* (Matteson and Pike 1958)

/kwalï/	k^{ə}walï ~ k^{ʊ}walï	'platform'
/tkatʃi/	t^{ɨ}katʃi	'sun'
/ʃjo/	ʃ^{i}jo	'bat'
/hwï/	h^{ɨ}wï ~ h^{u}wï ~ h^{ə}wï ~ h^{i}wï ~ h^{ɯ}wï	'O.K.'
/whene/	w̥hene ~ w^{ə}hene ~ w^{u}hene ~ w^{i}hene ~ w^{ɯ}hene	'child'

Based on the vowels' exceptional phonological and phonetic characteristics, the authors analyze them as "non-phonemic transitional vocoids." Vowels with similar characteristics occur in Finnish (Harms 1976), Sanskrit (Allen 1953: 173), South Hamburg German (Jannedy 1994), and other languages listed in Hall (2006).

Recently, a number of authors have formalized similar ideas about excrescent vowels in an Articulatory Phonology framework. Articulatory Phonology (Browman and Goldstein 1986, 1992) treats abstract articulatory gestures as primitives, and allows the grammar to regulate the timing of articulatory gestures with respect to one another. Vowel-like percepts can be created when two consonant gestures are phased to have a low degree of overlap with one another, leaving a period between the consonant constrictions where the vocal tract is relatively open (Browman and Goldstein 1992). See Gafos (2002) and Hall (2006) for arguments that excrescent vowels lack an independent gesture, and hence are not present as phonological units in the way that lexical vowels (and most epenthetic vowels) are.

7 How does vowel epenthesis interact with other processes?

One of the most interesting characteristics of epenthetic vowels is their tendency to interact opaquely with other phonological processes. It is common for phonological patterns to treat epenthetic vowels as if they were not present. This observation has many theoretical interpretations. Some argue that epenthetic vowels are representationally defective: Piggott (1995), for example, argues that some epenthetic vowels are weightless, lacking a mora. Other approaches handle opaque interactions through rule ordering, with the epenthetic vowels being inserted late in the derivation. Here, I will focus on the empirical issues to be explained, with examples of the kinds of interactions that have been reported.

7.1 Metrical patterns

Syllables whose nuclei are epenthetic vowels frequently fail to count as syllables in patterns such as stress assignment, minimal word requirements, and the conditioning of open syllable lengthening. This section gives an example of epenthesis interacting with each of these processes.

In Lebanese Arabic, a closed penult is stressed when it contains a lexical vowel, as in (15a), but not when it contains an epenthetic vowel, as in (15b) (see also CHAPTER 124: WORD STRESS IN ARABIC).

(15) *Stress–epenthesis interaction in Lebanese Arabic*

a.	/fihim-na/	fi.'him.na	'he understood us'
b.	/fihm-na/	'fi.hi̲m.na	'our understanding'

In words without a closed penult, stress normally falls on the final syllable if it is superheavy, i.e. CV:C or CVCC, as in (16a), and on the antepenult otherwise, as in (16b). Again, vowel epenthesis disrupts the pattern. If an epenthetic vowel is inserted into a final CC cluster, breaking up what would otherwise be a final superheavy syllable, stress is assigned to the penult, as in (16c). This is the only case in which a light penult can be stressed.

(16) *Lebanese Arabic* (Haddad 1984a)

a.	/nazzal-t/	naz.'zalt	'I brought down'
b.	/katab-it/	'ka.ta.bit	'she wrote'
c.	/katab-t/	ka.'ta.bi̲t	'I wrote'

For all of the patterns above, stress is simply assigned as if the epenthetic vowel were absent. The only exception to this generalization is an epenthetic vowel inserted in an underlying CCCC sequence. In this case alone, the epenthetic vowel is treated the same as a lexical vowel for stress. In (17), the epenthetic vowel falls in a closed penult, and is stressed, as is normal for a heavy penult (cf. (15a)).

(17) /katab-t-l-ha/ ka.tab.'ti̲l.ha 'I wrote to her'

Such patterns, where epenthetic vowels are visible to stress under some circumstances but invisible in others, also occur in Mohawk (Michelson 1989) and Selayarese (Broselow 1999).

In languages that require words to have a minimal size, epenthetic vowels may not count in determining this size. Mohawk, for example, requires each lexical word to contain two syllables, as in (18a). A verbal stem containing only one syllable is augmented with an epenthetic [i], as in (18b). Mohawk also inserts an epenthetic [e] after the first consonant of certain CC and CCC clusters. This [e] counts for metrical purposes if it is in a closed syllable, but not if it is in an open syllable. Hence, a two-syllable word containing an open epenthetic syllable, as in (18c), is augmented with epenthetic [i] as well. However, a two-syllable word containing epenthetic [e] in a closed syllable is not augmented, as seen in (18d).

(18) *Minimal word augmentation in Mohawk* (Michelson 1989)

a.	/k-hnˈuu-s/	ˈkhniː.nus	'I buy'
b.	/k-jʌ-s/	ˈi̲k.jʌs	'I put it'
c.	/s-riht/	ˈi̲.se̲.riht	'cook!'
d.	/s-rho-s/	ˈse̲r.hos	'you coat it with something'

This interaction highlights another interesting problem: the fact that there may be multiple vowel epenthesis processes in a single language, which differ in whether they are metrically "visible."

Epenthetic [e] in Mohawk also shows another type of metrical invisibility: it fails to trigger a rule by which stressed vowels lengthen in an open syllable. In (19a) we see this rule apply normally. In (19b), it appears that the stressed [i] is an open syllable, since the following epenthetic vowel has syllabified [r] as an onset; yet the stressed syllable fails to lengthen.

(19) *Stressed vowel lengthening in Mohawk*

a.	/wak-ashet-u/	wa.kas.ˈheː.tu	'I have counted it'
b.	/s-riht/	ˈi̲.se̲.riht	'cook!'

In sum, although epenthetic vowels are usually added in order to syllabify stray consonants, the syllables they form do not necessarily count as syllables for other aspects of the phonology.

7.2 *Segmental processes*

In some cases, epenthetic vowels fail to condition other segmental processes, such as deletion or allophonic variation, in the same way that lexical vowels condition them. In Dutch, for example, underlying /ən/ is optionally reduced to [ə], as in (20a). Yet when schwa epenthesis occurs before /n/, as in (20b), the epenthetic schwa does not condition deletion of the following [n]. Some speakers thus eliminate underlying /ən/, yet create surface [ən] through epenthesis.

(20) *Dutch* [n]-*deletion* (Booij 1995; Hall 2006)

a.	*regen*	/reɣən/	→	reɣən ~ reɣə		'rain'
	horen	/horən/	→	horən ~ horə		'to hear'
b.	*hoorn*	/horn/	→	horn ~ horə̲n	*horə̲	'horn'

Similarly, Herzallah (1990) describes a Palestinian Arabic dialect in which a pharyngealized [rˤ] loses its pharyngealization before lexical [i], but not before epenthetic [i] (CHAPTER 25: PHARYNGEALS).

Just as different epenthetic vowels within a single language may show different metrical behavior, they may also differ in whether they condition other segmental processes. For example, in Tiberian Hebrew, one kind of epenthetic vowel does condition spirantization in following stops, and another does not. Normally a stop becomes a fricative after vowels, as in (21a). One type of epenthetic vowel, which splits up final CC clusters in non-derived words, also causes spirantization. In (21b), we see /b/ spirantize to [β] following the epenthetic [e]. But another epenthetic vowel, which occurs in final clusters of a guttural and a following consonant, does not condition spirantization. In (21c), the /t/ following the epenthetic vowel is realized as [t] rather than [θ].

(21) *Tiberian Hebrew spirantization* (McCarthy 1979)

a.	/katab+t/	→	kaθaβt	'you (FEM SG) wrote'
b.	/kelb/	→	kele̱β	'dog'
c.	/ʃalaħ+t/	→	ʃalaħa̱t	'you (FEM SG) sent'

Thus, there is variation both within and between languages in how vowel epenthesis interacts with other processes.

8 How does epenthesis happen in loanwords?

Typological studies of vowel epenthesis frequently consider loanword data side by side with cases of epenthesis within languages, under the assumption that similar phonological mechanisms produce both (e.g. Broselow 1982; Kitto and de Lacy 1999; among many others). Since vowel epenthesis is particularly common in loanwords, loanword data have played a large role in theorizing on epenthesis, probably more than most other phenomena. However, I would like to argue that conflating loanword and native-language epenthesis is a serious methodological mistake. A growing body of evidence suggests that epenthesis in loanwords differs from epenthesis within languages in its formal characteristics, and may have different causes and functions. For this reason, facts about loanword epenthesis are reviewed here separately from within-language epenthesis, to highlight some likely empirical differences between the two kinds of epenthesis. I will also include some references to epenthesis in "interlanguage," which is the language produced by second language learners. While interlanguage and loanwords are not the same thing, they are related in the sense of both involving language contact, and many loanwords may arise historically from interlanguage forms (see also CHAPTER 95: LOANWORD PHONOLOGY).

8.1 Perceptual origin?

There is considerable debate over whether epenthesis in loanwords happens through perceptual errors by speakers of the borrowing language. Traditionally, it was assumed that a speaker of the borrowing language (likely a bilingual) would hear a foreign word, construct some reasonably accurate representation of the way the word was pronounced in the foreign language, and then alter that representation to fit the phonotactics of the borrower's native language. But Peperkamp and Dupoux (2002) argue that the borrower is likely to perceive the foreign word incorrectly, and that these perceptual errors are the main source of phonological alterations in loanwords (see also CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY and CHAPTER 95: LOANWORD PHONOLOGY for further discussion).

One piece of evidence for this view comes from Japanese, which inserts an epenthetic vowel to remove illegal codas in loanwords (only a nasal or the first half of a geminate can be a coda). The epenthetic vowel is [o] after [d] and [t], and [ɯ] elsewhere.

(22) *Japanese loanwords from English* (Itô and Mester 1995)

faito	'fight'
fesɯtibarɯ	'festival'
sɯfiŋkɯsɯ	'sphinx'

Dupoux *et al.* (1999) argue that Japanese speakers actually believe they hear this [ɯ] in the pronunciation of foreign CC clusters. In a perception experiment, Japanese and French listeners were asked to judge whether a middle vowel was present in nonsense words like [ebzo] and [ebɯzo]. For words like [ebzo], where no middle vowel was acoustically present, most Japanese listeners reported hearing a vowel, while most French listeners did not. Japanese listeners also had great difficulty in discriminating between tokens like [ebzo] and [ebɯzo] in an ABX discrimination test. Dupoux *et al.* point out that Japanese [ɯ] is frequently devoiced and shortened, and shows considerable allophonic variation. Knowing this may make listeners likely to fill in an illusory [ɯ] when they hear consonants with no vowel between them.

The idea that epenthesis in loanwords has a perceptual origin is controversial; see Rose and Demuth (2006), Smith (2006), and Uffmann (2007) for arguments that perceptual factors cannot account for all facets of loanword adaptation. Nevertheless, we will see below several additional arguments that perceptual factors play a special role in loanword vowel epenthesis.

8.2 *Function of vowel epenthesis*

For within-language phonology, epenthesis usually occurs to repair an input that does not meet the language's phonotactic or metrical requirements. In most cases, epenthesis in loanwords can be analyzed as having the same function, like the Japanese examples in (22). Yet surprisingly, there is at least one case where speakers add epenthetic vowels to loanwords that were phonotactically permissible in the borrowing language without the vowel. Korean (Kang 2003) frequently epenthesizes a final vowel after English loanwords ending in a stop, as in the examples below.

(23) *English loanwords in Korean* (Kang 2003: 223)

gag	→	kækɨ
pat	→	pʰætɨ
tube	→	tʰjupɨ

There is no phonotactic need to add vowels to these words. The consonants /k t p/ are among the acceptable codas of Korean, occurring in native words such as [kæk] 'guest', so epenthesis cannot be explained as a means of syllabifying stray consonants. Kang argues that the purpose of the vowel is to maximize perceptual similarity between the English word and the Korean word. English has more release of final stops than Korean does, and Kang claims that to Korean listeners, the release of a final stop of an English word sounds vocalic. She shows that final vowel insertion in loanwords from English is most common in precisely the environments where final stop release is most common in English, such as after voiced stops and when the preceding vowel is tense. Thus, epenthesis may be a means of preserving phonetic details of the source language, rather than a repair.

8.3 *Relation to native phonology*

The epenthetic vowel used in loanwords often differs from any vowel epenthesis process that exists in the native phonology, and epenthesis may be used in loanwords in contexts where other repairs would be used in the native phonology.

In Japanese, for example, consonant clusters that arise through morpheme concatenation in the native language are repaired through deletion of one of the consonants, as shown in (24). Yet consonant clusters in loanwords are repaired with vowel epenthesis, as in (22).

(24) *Deletion in Japanese native phonology* (McCawley 1968; Smith 2006)

non-past /-ɾɯ/		*causative* /-sase/		
/jom-ɾɯ/	jo.mɯ	/jom-sase/	jo.ma.se	'read'
/tob-ɾɯ/	to.bɯ	/tob-sase/	to.ba.se	'fly'

Karimi (1987) reports a similar case for Farsi: CCC clusters are subject to consonant deletion in the native phonology, but repaired through epenthesis in loanwords and interlanguage.

In general, vowel epenthesis seems to be a heavily favored repair type in loan adaptation, more than in native phonologies. Uffmann (2007) surveys case studies of loanword adaptation and concludes that consonant deletion is a marginal phenomenon, compared to epenthesis. Paradis and LaCharité (1997) invoke the "Preservation Principle," which states that segmental material is maximally preserved (see also CHAPTER 76: STRUCTURE PRESERVATION: THE RESILIENCE OF DISTINCTIVE INFORMATION). Hence, adding extra segments is less undesirable than deleting segments from the source word. It is possible that the prevalence of vowel epenthesis in loanwords is related to its prevalence in interlanguage. Jenkins (2000) observes, based on a corpus of conversations between non-native speakers of English, that more misunderstandings are caused by deletion of consonants than by addition of vowels. If bilinguals are aware of this fact and therefore favor vowel epenthesis in their interlanguage pronunciations, then any loanwords based on these interlanguage forms would also tend to favor vowel epenthesis.

8.4 *Quality*

As in native language phonology, epenthetic vowels in loanwords may have a default quality or copy their quality from nearby consonants or vowels. However, the patterns of vowel quality in loanwords are often strikingly complex in ways that are not common (and perhaps not attested at all) in native language epenthesis.

Consider the patterns of epenthetic vowel place in words borrowed from English or Afrikaans into the southern Bantu language Sotho, as described in Rose and Demuth (2006). This study examines only the front–back dimension of epenthetic vowel place. In word-initial CC clusters, the epenthetic vowel is back when it follows a labial (25a), and front when it follows a coronal (25b). When the initial

C is velar, the epenthetic vowel copies the place of the following vowel, as in (25c). In word-medial or word-final /CC/ clusters, usually the vowel copies its place from the preceding vowel, as in (25d). (A few further sub-patterns are ignored here. Only epenthetic vowels discussed in the text are underlined.)

(25) *Epenthesis in loanwords in Sotho* (Rose and Demuth 2006)

	source word	borrowed form	
a.	blɪk	bʊ̲leke	'tin can, dish'
b.	truwn	tɪ̲ronɪ	'throne'
c.	xʀɑːf	kʰɑ̲rɑfu	'spade'
d.	hibruw	hebe̲ru	'Hebrew'

Sotho also shows epenthesis for minimal word purposes within the native vocabulary, but in this case, the epenthetic vowel is always [ɪ], regardless of context. Sotho is not the only case where vowel epenthesis in loanwords follows such a complex pattern; Uffmann (2007) analyzes similarly complicated rules for epenthetic vowel quality in Shona, Sranan, Nyarwanda, and Samoan, each of which shows an interplay between copying the features of consonants, copying the features of vowels, and insertion of default features.

An informal survey of the literature gives the impression that such complex effects of phonological context on vowel quality are more or less confined to loanword epenthesis. Within languages, it is far more common to find epenthetic vowels of default quality, as in Arabic, or relatively simple kinds of copying, such as always copying in one direction, as in the Welsh pattern in (12). An extensive typological comparison of the formal qualities of vowel epenthesis in loanwords and non-loanwords would be a valuable contribution to understanding the difference between them.

Another important difference between loanword and native language epenthesis is that epenthesis in loanwords is often not fully predictable. As we saw in the Korean examples in (23), epenthesis in a given location may be optional, and in languages like Shona and Sotho, "rules" for epenthetic vowel quality in loanwords are not exceptionless. Uffmann (2007: 9–13) argues that loanword epenthesis needs to be studied by looking for statistical patterns in large corpora of loanwords, because incorrect generalizations are easily reached from impressionistic or limited data.

Both the complexity and unpredictability of some loanword epenthesis patterns may indicate that these patterns have not been internalized by speakers as true phonological "rules" – again, an argument for not considering them side by side with language-internal epenthesis.

8.5 Vowel placement

The problem of where to place an epenthetic vowel arises in loanword phonology in the same way as in native language phonology: initial CC clusters, or medial CCC clusters, can potentially be split in two ways.

In some cases, epenthesis location in loanwords or interlanguage appears to follow the same placement pattern as the borrowing language shows in its native epenthesis patterns. For example, we saw in (11) that Iraqi and Egyptian Arabic differ in how they break up word-medial CCC clusters in the native phonology: Iraqi puts the epenthetic vowel after the first consonant, and Egyptian after the second. These dialects differ in exactly the same way in how they epenthesize into CCC clusters in interlanguage phonology, as seen in (26). This pattern can be explained by the same mechanism, directionality of syllabification, that is commonly used to explain vowel placement in the native phonologies of these languages.

(26) *Iraqi* vs. *Egyptian epenthesis in* CCC *clusters* (Broselow 1987)

			Iraqi	Egyptian
native language	/kitab+t+l+V/	→	ki.ta.bit.la	ki.tab.ti.lu
interlanguage	*children*	→	*chilidren*	*childiren*

Yet in other cases the placement of the epenthetic vowel is not explainable as a transfer of native language epenthesis rules, and cannot be analyzed through directional syllabification alone. Fleischhacker (2001) presents a typological study of epenthesis in initial CC(C) clusters in loanwords and interlanguage, focusing on the question of whether the vowel precedes the cluster (VCC) or breaks up the cluster (CVC). She shows that in many languages, the placement of the vowel depends on what kind of consonants are in the cluster, as in the Egyptian Arabic examples in (27). In word-initial clusters consisting of a voiceless sibilant plus a stop, it is cross-linguistically more common to insert a vowel before the first consonant, as in (27a), while in word-initial clusters of an obstruent and sonorant, it is more common to place the vowel between the consonants, as in (27b).

(27) *Egyptian Arabic epenthesis in interlanguage* (Broselow 1987)

a.	*study*	→	istadi
	special	→	izbasjal
	ski	→	iski
b.	*sweater*	→	siwetar
	slide	→	silaid

Fleischhacker argues that the reason for this pattern is that epenthetic vowels are inserted where they will cause the least perceptual difference between the foreign word and the epenthesized adaptation (a theory which follows the P-map hypothesis of Steriade 2003). She presents an experiment in which English listeners were asked to judge auditory similarity between English words and modifications of those words with epenthetic vowels in different locations. Words beginning with sibilant–stop clusters, like *spar*, were judged more similar to versions with epenthesis before the cluster ([əspɑr]) than to versions with epenthesis within the cluster ([səpɑr]). Words beginning with obstruent–sonorant clusters, like *flit*, were judged more similar to versions with epenthesis within the cluster ([fəlɪt]) than to versions with epenthesis before the cluster ([əflɪt]). The results of the perception experiment thus match the cross-linguistic tendencies in epenthetic vowel placement, and add to the body of arguments that perceptual factors have a special role in loanword epenthesis.

9 Conclusion and suggested directions for future research

In the discussion above, I have tried to highlight some of the main empirical questions about vowel epenthesis, and to show that vowel epenthesis processes are greatly heterogeneous. A better understanding of vowel epenthesis will require work on two dimensions. One is detailed case studies of individual languages, in particular studies that combine the traditional, structural description of vowel epenthesis with attention to the acoustics, articulation, and perception of the epenthetic vowels, and also probe speaker intuitions about the vowels. Epenthetic vowels in Dutch are probably currently the best-studied in this regard, and it would be useful to have similar experiments done with epenthetic vowels in other languages. It would be interesting to examine whether the phonetic nature of an epenthetic vowel (for example, whether it is acoustically identical to a lexical vowel) correlates with any aspect of its phonological behavior (for example, whether it is visible to other phonological processes in the same way that lexical vowels are). The second area is typological work that looks for correlations between different characteristics of epenthetic vowels. Often, typological studies that focus on one variable, such as vowel quality, have lumped together vowel epenthesis processes that differ on other important parameters, such as whether the epenthesis occurs in native words or loanwords, whether the vowels are excrescent or not, whether they are morphologically conditioned, etc. However, it is possible that there may be relations between these variables. For example, it would be interesting to see more systematic comparisons of epenthesis in loanwords *vs.* native phonology, given the growing evidence that these processes may work differently.

REFERENCES

Abdul-Karim, Kamal. 1980. Aspects of the phonology of Lebanese Arabic. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Allen, W. Sidney. 1953. *Phonetics in Ancient India*. London: Oxford University Press.

Awbery, Gwenllian M. 1984. Phonotactic constraints in Welsh. In Martin Ball & Glyn E. Jones (eds.) *Welsh phonology*, 65–104. Cardiff: University of Wales Press.

Booij, Geert. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.

Borgstrøm, Carl H. 1937. The dialect of Barra in the Outer Hebrides. *Norsk Tidsskrift for Sprogvidenskap* 8. 71–242.

Borgstrøm, Carl H. 1940. *A linguistic survey of the Gaelic dialects of Scotland*, vol. 1: *The dialects of the Outer Hebrides*. Oslo: Aschehoug.

Bosch, Anna & Kenneth J. de Jong. 1997. The prosody of Barra Gaelic epenthetic vowels. *Studies in the Linguistic Sciences* 27(1). 2–15.

Bosch, Anna & Kenneth J. de Jong. 1998. Syllables and supersyllables: Evidence for low level phonological domains. *Texas Linguistic Forum* 41. 1–14.

Broselow, Ellen. 1982. On predicting the interaction of stress and epenthesis. *Glossa* 16. 115–131.

Broselow, Ellen. 1987. Non-obvious transfer: on predicting epenthesis errors. In Ioup & Weinberger (1987), 292–304.

Broselow, Ellen. 1992. Parametric variation in Arabic dialect phonology. In Ellen Broselow, Mushira Eid & John J. McCarthy (eds.) *Perspectives on Arabic linguistics IV*, 7–45. Amsterdam & Philadelphia: John Benjamins.

Broselow, Ellen. 1999. Stress, epenthesis, and segment transformation in Selayarese loans. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 25. 311–325.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Browman, Catherine P. & Louis Goldstein. 1992. "Targetless" schwa: An articulatory analysis. In Gerard J. Docherty & D. Robert Ladd (eds.) *Papers in laboratory phonology II: Gesture, segment, prosody*, 26–56. Cambridge: Cambridge University Press.

Clements, G. N. 1986. Syllabification and epenthesis in the Barra dialect of Gaelic. In Koen Bogers, Harry van der Hulst & Martin Mous (eds.) *The phonological representation of suprasegmentals*, 317–336. Dordrecht: Foris.

Côté, Marie-Hélène. 2000. Consonant cluster phonotactics: A perceptual approach. Ph.D. dissertation, MIT.

Davidson, Lisa & Maureen Stone. 2003. Epenthesis versus gestural mistiming in consonant cluster production: An ultrasound study. *Proceedings of the West Coast Conference on Formal Linguistics* 22. 165–178.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.

Donselaar, Wilma van, Cecile Kuijpers & Anne Cutler. 1999. Facilitatory effects of vowel epenthesis on word processing in Dutch. *Journal of Memory and Language* 41. 59–77.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25. 1568–1578.

Fagyal, Szuszanna. 2000. Le retour du e final en français parisien: Changement phonétique conditionné par la prosodie. In *Actes du XIIe Congrès International de Linguistique et Philologie Romane*, vol. 3: *Vivacité et diversité de la variation linguistique*, 151–160. Tübingen: Narr.

Fleischhacker, Heidi. 2001. Cluster-dependent epenthesis asymmetries. *UCLA Working Papers in Linguistics 7, Papers in Phonology* 5. 71–116.

Gafos, Adamantios I. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20. 269–337.

Gick, Bryan & Ian Wilson. 2006. Excrescent schwa and vowel laxing: Crosslinguistic responses to conflicting articulatory targets. In Louis M. Goldstein, Douglas Whalen & Catherine T. Best (eds.) *Laboratory phonology* 8, 635–659. Berlin & New York: Mouton de Gruyter.

Gouskova, Maria & Nancy Hall. 2009. Acoustics of epenthetic vowels in Lebanese Arabic. In Steve Parker (ed.) *Phonological argumentation: Essays on evidence and motivation*, 203–225. London: Equinox.

Haddad, Ghassan. 1984a. Problems and issues in the phonology of Lebanese Arabic. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Haddad, Ghassan. 1984b. Epenthesis and sonority in Lebanese Arabic. *Studies in the Linguistic Sciences* 14. 57–88.

Hall, Nancy. 2003. Gestures and segments: Vowel intrusion as overlap. Ph.D. dissertation, University of Massachusetts, Amherst.

Hall, Nancy. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23. 387–429.

Harms, Robert T. 1976. The segmentalization of Finnish "nonrules." *Texas Linguistic Forum* 5. 73–88.

Herzallah, Rukayyah. 1990. Aspects of Palestinian Arabic phonology: A non-linear approach. Ph.D. dissertation, Cornell University.

Ioup, Georgette & Steven H. Weinberger (eds.) 1987. *Interlanguage phonology: The acquisition of a second language sound system*. Cambridge: Newbury House.

Itô, Junko. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7. 217–259.

Itô, Junko & Armin Mester. 1995. Japanese phonology. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 817–838. Cambridge, MA & Oxford: Blackwell.

Jannedy, Stefanie. 1994. Rate effects on German unstressed syllables. *OSU Working Papers in Linguistics* 44. 105–124.

Jenkins, Jennifer. 2000. *The phonology of English as an international language*. Oxford: Oxford University Press.

Johnston, Raymond Leslie. 1980. *Nakanai of New Britain: The grammar of an Oceanic language*. Canberra: Australian National University.

Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20. 219–273.

Karimi, Simin. 1987. Farsi speakers and the initial consonant cluster in English. In Ioup & Weinberger (1987), 305–318.

Kawahara, Shigeto. 2007. Spreading and copying in phonological theory: Evidence from echo epenthesis. In Leah Bateman, Michael O'Keefe, Ehren Reilly & Adam Werle (eds.) *Papers in Optimality Theory III*, 111–143 Amherst, MA: GLSA.

Kiparsky, Paul. 2003. Syllables and moras in Arabic. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*, 147–182. Cambridge: Cambridge University Press.

Kitto, Catherine & Paul de Lacy. 1999. Correspondence and epenthetic quality. In Catherine Kitto & Carolyn Smallwood (eds.) *Proceedings of AFLA VI: The Sixth Meeting of the Austronesian Formal Linguistics Association*, 181–200. Toronto: Department of Linguistics, University of Toronto.

Kuijpers, Cecile & Wilma van Donselaar. 1997. The influence of rhythmic context on schwa epenthesis and schwa deletion in Dutch. *Language and Speech* 41. 87–108.

Ladefoged, Peter, Jenny Ladefoged, Alice Turk, Kevin Hind & St. John Skilton. 1998. Phonetic structures of Scottish Gaelic. *Journal of the International Phonetic Association* 28. 1–41.

Levin, Juliette. 1987. Between epenthetic and excrescent vowels. *Proceedings of the West Coast Conference on Formal Linguistics* 6. 187–201.

Martin, Pierre. 1998. À Québec, a-t-on *l'schwa*? In Yves Duhoux (ed.) *Langue et langues: Hommage à Albert Maniet*, 163–180. Leuven: Peeters.

Martinez-Gil, Fernando. 1997. Word-final epenthesis in Galician. In Fernando Martínez-Gil & Alfonso Morales-Front (eds.) *Issues in the phonology and morphology of the major Iberian languages*, 270–340. Washington, DC: Georgetown University Press.

Matteson, Esther & Kenneth L. Pike. 1958. Non-phonemic transition vocoids in Piro (Arawak). *Miscellanea Phonetica* 3. 22–30.

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague & Paris: Mouton.

Michelson, Karin. 1989. Invisibility: Vowels without a timing slot in Mohawk. In Donna Gerdts & Karin Michelson (eds.) *Theoretical perspectives on Native American languages*, 38–69. New York: SUNY Press.

Miner, Kenneth L. 1979. Dorsey's Law in Winnebago-Chiwere and Winnebago accent. *International Journal of American Linguistics* 45. 25–33.

Ní Chiosáin, Máire. 1995. Barra Gaelic vowel copy and (non-)constituent spreading. *Proceedings of the West Coast Conference on Formal Linguistics* 13. 3–15.

Oftedal, Magne. 1956. *A linguistic survey of the Gaelic dialects of Scotland*, vol. 3: *The Gaelic of Leurbost, Isle of Lewis*. Oslo: Aschehoug.

Olson, Kenneth S. 2003. Word minimality and non-surface-apparent opacity in Mono. Paper presented at the 4th World Conference of African Linguistics, Rutgers University.

Paradis, Carole & Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33. 379–430.

Pearce, Mary. 2004. Kera foot structure. Unpublished ms., University College London.

Peperkamp, Sharon & Emmanuel Dupoux. 2002. A typological study of stress "deafness." In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology 7*, 203–240. Berlin & New York: Mouton de Gruyter.

Piggott, Glyne L. 1995. Epenthesis and syllable weight. *Natural Language and Linguistic Theory* 13. 283–326.

Rice, Keren. 2008. Review of de Lacy (2006). *Phonology* 25. 361–371.

Rose, Yvan & Katherine Demuth. 2006. Vowel epenthesis in loanword adaptation: Representational and phonetic considerations. *Lingua* 116. 1112–1139.

Smith, Jennifer. 2006. Loan phonology is not all perception: Evidence from Japanese loan doublets. In Timothy J. Vance & Kimberly A. Jones (eds.) *Japanese/Korean Linguistics 14*, 63–74. Palo Alto: CLSI.

Smorodinsky, Iris. 2002. Schwas with and without active control. Ph.D. dissertation, Yale University.

Steriade, Donca. 1994. Licensing by cue. Unpublished ms., University of California, Los Angeles.

Steriade, Donca. 2003. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. Unpublished ms., University of California, Los Angeles.

Susman, Amelia L. 1943. The accentual system of Winnebago. Ph.D. dissertation, Columbia University.

Uffmann, Christian. 2007. *Vowel epenthesis in loanword adaptation*. Tübingen: Max Niemeyer Verlag.

Watson, Janet C. E. 2007. Syllabification patterns in Arabic dialects: Long segments and mora sharing. *Phonology* 24. 335–356.

68 Deletion

John Harris

1 Introduction

In language, we often come across situations where a morpheme shows up in two alternating phonological shapes, one of which contains a vowel or consonant segment that is missing from the other. For example, in Samoan the root meaning 'twist' has two alternants: [milos] and [milo]. The form ending in [s] is found before a vowel-initial suffix (as in [milos-ia] 'be twisted'), while the form lacking the [s] is found when the root falls at the end of a word.

In such situations, the two alternants usually stem from a single historical source, and we can attribute the alternation between segment and zero to the action of some sound change. The question then is whether or not the historical form contained the segment that alternates in the present-day forms. Either the segment was absent from the original form and has since been inserted in certain phonological contexts, or it was present and has since been deleted in certain contexts. Which of these scenarios is correct depends on whether the segment's occurrence is phonologically predictable or not.

In the case of Samoan, the [s] must have been part of the original form of the root meaning 'twist', because its presence is unpredictable. There is no regular sound change that could have inserted the consonant without also incorrectly inserting it in other morphemes. The unpredictability of the root-final consonant is confirmed by observing roots that show alternations with consonants other than [s], such as [oso] – [osof-ia] 'jump', [tau] – [taul-ia] 'cost'. The conclusion then is that Samoan has undergone a change that deleted consonants at the ends of words.

Whole-segment deletion is a pervasive phenomenon in the world's languages. Much of the terminology phonologists use to describe it – including the term *deletion* itself – dates from nineteenth-century philology. Although the terms were initially applied to historical deletion processes, they have subsequently been extended into synchronic phonology. This is largely due to a well-established tradition of assuming that when a sound change affects a grammar it can remain there as an active phonological process. According to this model of grammar, a synchronically live process allows regularly alternating forms to be derived from a single underlying form that strongly resembles the historical form (CHAPTER 93: SOUND CHANGE). In our Samoan example, this means that the form [milo] is derived from underlying /milos/ through the operation of a synchronic process that deletes a consonant when it is final in the word.

General terms used to describe whole-segment deletion include *elision, loss, drop,* and *truncation*. Although these terms continue to prove useful for descriptive purposes, they retain a strong flavor of the philological tradition within which they were conceived. There are at least two connotations that more recent research has shown should not be allowed to determine how we approach synchronic deletion. First, there is a procedural flavor to the terminology: deletion might suggest that a phonological form is derivationally altered by the irretrievable elimination of a sound. Second, there is an implication that what gets deleted is a phoneme-sized unit – an impression undoubtedly reinforced by the practice of using alphabetic transcription to present the relevant data. As we'll see, neither of these connotations accurately reflects how deletion is treated in modern phonological theory.

This chapter is laid out as follows. §2 catalogues the main types of vowel and consonant deletion, sticking as far as possible to theory-neutral descriptions. §3 discusses different approaches to how deletion effects are represented in phonological grammars. The next two sections examine the phonological conditions under which deletion occurs. §4 addresses the issue of what causes consonant deletion and reviews claims that it is driven by an imperative to simplify syllable structure. §5 focuses on vowel deletion and evaluates the assumption that it inevitably triggers resyllabification. The chapter concludes in §6 by considering an alternative to traditional derivational approaches to deletion.

2 Segmental deletion

2.1 *Consonant deletion*

The following forms expand the Samoan example we started with (data originally from Pratt 1911, cited in Bloomfield 1933):

(1) *Samoan*

	Simple	*Perfective*	
a.	olo	oloia	'rub'
	aŋa	aŋaia	'face'
	tau	tauia	'repay'
b.	api	apitia	'be lodged'
	sopo	sopoʔia	'go across'
	milo	milosia	'twist'
	oso	osofia	'jump'
	tau	taulia	'cost'
	asu	asuŋia	'smoke'
	ŋalo	ŋalomia	'forget'

The form of the perfective suffix, [-ia], is evident from the examples containing vowel-final roots in (1a). The examples in (1b) contain consonants that appear in the perfective but not in simple forms. Since the alternating consonants vary unpredictably from one word to another, we can conclude that each belongs to a root rather than the suffix. The phonological process responsible for the alternation can be summarized as follows: a consonant is deleted at the end of a word. The result is that root-final consonants are elided when word final but survive when prevocalic.

The effect of deletion in Samoan has been to bar any consonant from appearing word finally. In a more restricted version of the process, final deletion only targets consonants of a certain type. In Lardil, for example, the only type of consonant permitted finally is apical. Stem-final consonants, which show up before a suffix vowel, delete finally if not apical; compare (2a) with (2b) (Hale 1973).

(2) *Lardil*

	Bare noun	*Accusative*	
a.	pirŋen	pirŋen-in	'woman'
	kentapal	kentapal-in	'dugout'
b.	ŋalu	ŋaluk-in	'story'
	taŋku	taŋkuŋ-in	'oyster'
	murkuni	murkunim-an	'nullah'

Lardil illustrates another restricted form of final deletion, where the targeted consonant is preceded by another consonant. The effect of simplifying final consonant clusters in this way can be seen in (3), where the second of two stem-final consonants drops when word final. The process is fed by an independent process of final vowel deletion (on which more below) and itself feeds the deletion of final non-apical consonants exemplified in (2).

(3) *Lardil*

Input	*Output*	
/jukarpa/	jukar	'husband'
/wuluŋka/	wulun	'fruit'
/kantu-kantu/	kantukan	'red'

As the form [kantukan] in (3) attests, cluster simplification in Lardil affects any type of consonant, including apicals. In other languages, a more restricted version of this process shows sensitivity to the type of consonant involved. In Catalan, for example, final cluster simplification targets coronals but not other places; compare (4a) with (4b) (Mascaró 1983).

(4) *Catalan*

	Masculine	*Feminine*	
a.	əskerp	əskerpə	'shy'
	orp	orβə	'blind'
	ʎark	ʎarɣə	'long'
b.	al	altə	'tall'
	for	fortə	'strong'
	bɛr	bɛrðə	'green'
	san	santə	'saint'
	prufun	prufundə	'deep'
	blaŋ	blaŋkə	'white'

Cluster simplification can also target non-final consonants. Here, which member of a cluster drops varies in a way that is often attributed to differences in syllabification (on which more below). Pali illustrates the case where the second of two consonants deletes; as shown in (5), historical liquids (evident in the cognate Sanskrit forms) have been lost post-consonantally (Zec 1995).

(5)	*Sanskrit*	*Pali*	
	prati	paṭi	'against'
	traana	taana	'protection'
	kramati	kamati	'walks'

In syllabic terms, the deletion exemplified in (5) simplifies an onset cluster. The reverse situation, where it is the first of two non-final consonants that drops, is typically described as coda deletion. Consider the example of Diola Fogny, where the only permitted type of coda–onset cluster is a partial geminate consisting either of a nasal plus homorganic obstruent or of a liquid plus coronal obstruent (Sapir 1965; Itô 1986). As illustrated in (6a), this sequence can arise through morphological derivation, in which case it survives in output (with appropriate adjustments for homorganicity).

(6) *Diola Fogny*

	Input	*Output*	
a.	/ni-gam-gam/	nigaŋgam	'I judge'
	/ku-boɲ-boɲ/	kubomboɲ	'they sent'
	/na-tiiŋ-tiiŋ/	natiintiiŋ	'he cut through'
b.	/let-ku-jaw/	lekujaw	'they won't go'
	/ujuk-ja/	ujuja	'if you see'
	/kob-kob-en/	kokoben	'yearn'

If, however, the juxtaposition of two morphemes creates a consonant sequence other than a partial geminate, the first consonant is deleted. This is illustrated by the elision of the stops in the examples in (6b).

The deletion of one consonant before another is often accompanied by compensatory lengthening (see CHAPTER 64: COMPENSATORY LENGTHENING), where one segment lengthens to make up for the loss of a neighbor. The compensation can be undertaken by either the following consonant or the preceding vowel. The first scenario is illustrated by the development of earlier Romance to later Italian [nokte] > [notːe] 'night', the second by earlier to later English [nixt] > [niːt] ('night', subsequently diphthongized to [najt]).[1]

2.2 *Vowel deletion*

Vowel sequences lacking an intervening consonant are cross-linguistically dispreferred. Whenever morpheme concatenation threatens to create a hiatus configuration of this sort, languages can take various measures to resolve it

[1] Singleton consonants are sometimes observed to elide intervocalically, as in Turkish *inek* 'cow (NOM)' – *ineği* 'cow (POSS)'. In this environment, deletion is almost always the final stage of historical lenition (see below and CHAPTER 66: LENITION).

(CHAPTER 61: HIATUS RESOLUTION). One of the most favored of these is to delete one of the vowels, either the first, as in French (see (7a)), or the second, as in Karok (see (7b); Bright 1957).

(7)

		Input	*Output*	
	a.	*French*		
		/lə ami/	lami	'the friend (MASC)'
		/la ami/	lami	'the friend (FEM)'
	b.	*Karok*		
		/ni-axjar/	nixjar	'fill (1SG)'
		/ni-uksup/	nikʃup	'point (1SG)'
	c.	*Ganda*		
		/ba-ezi/	beːzi	'sweepers'
		/ba-ogezi/	boːgezi	'speakers'

As with consonant cluster simplification, vowel deletion under hiatus can, depending on the language, be accompanied by compensatory lengthening. We see this in the Luganda examples in (7c), where a stem-initial vowel lengthens to make up for the loss of a prefix vowel (Clements 1986).

Vowel deletion can also occur between consonants. Syncope, as this process is called, is typically sensitive to stress or to the vowel's position relative to a word's edge. The examples in (8) illustrate two forms of stress-sensitive syncope in English (syncope-prone vowels underlined).

(8) *English*

a. potato, parade, career
b. opera, factory, chocolate, reference, family, camera

Syncope in English, which is both lexically and phonetically variable, targets unstressed syllables in two environments (Bybee 2001): (a) a word-initial unfooted syllable (as in (8a)) and (b) between a stressed and an unstressed syllable where the consonant following the targeted vowel is a sonorant and more sonorous that the consonant preceding (as in (8b)). The effect of the second pattern is to contract a trisyllabic sequence into a bisyllabic trochaic foot.

One type of positionally conditioned syncope targets the middle syllable of a trisyllabic sequence located at either the left or the right edge of a word. The left-edge scenario is illustrated in (9a) by Tonkawa (Hoijer 1946), the right-edge one in (9b) by Tagalog (Kenstowicz and Kisseberth 1979).

(9)	a.	*Tonkawa*		
		Input	*Output*	
		/pit͡sena-n-oʔ/	pit͡snanoʔ	'he is cutting it'
		/we-pit͡sena-n-oʔ/	wep͡tsenanoʔ	'he is cutting them'
	b.	*Tagalog*		
		Bare root	*Patient*	
		bukas	buksin	'open'
		kapit	kaptin	'embrace'
		laman	lamnin	'fill'

In Tonkawa, it is the second vowel in a word that syncopates. As shown in (9a), this means that the first vowel of a root (here, the /i/ of input /picena/) shows up when it is also the first vowel in the word but is suppressed when the root is preceded by a prefix vowel. Meanwhile, the second root vowel (/e/ in /picena/) is suppressed when the root is unprefixed but shows up under prefixation. In Tagalog, the last vowel in a root shows up when it is also the last vowel in the word but is suppressed when a suffix vowel follows. A similar effect is seen in Turkish and Hindi.

Deletion can also target vowels at the absolute edges of words, usually when the affected syllable is unstressed or in some other way non-prominent. We have already seen a word-final example (apocope) in Lardil (see (3)). Less common is the word-initial equivalent (aphaeresis), as in colloquial English *'bout*, *'lectric*.

3 Modeling segmental deletion

3.1 *Linear-segmental analyses of deletion*

The first basic question a model of segmental deletion needs to provide an answer to is this: how is an alternant that retains a segment related to one that lacks it? The best-established approach to this question can be broadly defined as derivational: the alternants are derived from a single underlying form that contains the segment, which is then removed under certain phonological conditions by some mechanism that typically recapitulates historical deletion. For several generations, this basic assumption has remained largely unchallenged in mainstream phonological theory.

The derivational approach raises a second basic question: what impact does the deletion mechanism have on the representations it operates on? Over the years, there have been quite radical changes in how phonologists answer this question.

Interpreted derivationally, the term *deletion* might suggest a scenario where a segment disappears without trace. Certainly that is one of the readings implicit in the tradition of representing segments as phoneme-sized units strung together linear-fashion in phonological forms. The reading is reinforced by modeling the deletion mechanism as a rule that transforms an underlying or input form containing a given phoneme into a surface or output form that lacks it (for a textbook exposition of how deletion is treated in linear-derivational theory, see Kenstowicz and Kisseberth 1979). Combining these linear and derivational assumptions yields an analysis of Samoan along the lines of (10).

(10)	Input form	/milos/	'twist'	/milos-ia/	'twist (PASS)'
	Rule C → / __ #	↓		↓	
	Output form	[milo]		[milosia]	

This type of derivation can be described as destructive: the rule ("delete a word-final consonant") destroys information that is present in the input. A question that immediately arises is whether this information is recoverable. That is, can a language learner or listener reconstruct an underlying segment on the evidence of an output form from which it has been removed? In the example in (10), could a learner/listener retrieve the underlying /s/ in Samoan /milos/ on the basis of encountering surface [milo]? If the output form [milo] were the only available evidence, the answer would almost certainly have to be no.[2] On the face of it, no information is present in [milo] that would alert the learner/listener to the presence of /s/ in the adult talker's underlying form. In declarative theories of phonology, this lack of recoverability is enough to disqualify deletion as a derivational device (see e.g. Bird 1995; Scobbie *et al.* 1996).

For most phonologists, however, this definition of recoverability is too narrow. There is other evidence the learner/listener can potentially draw on to reconstruct an input segment that fails to appear in the output. In cases such as Samoan, the most obvious alternative source of evidence lies in the availability of alternants that preserve the deletion-prone segment in the output. According to a broader interpretation of recoverability, exposure to the [s] in a suffixed form such as [milosia] allows the learner to construct an underlying form /milos/, which can then be accessed whenever occurrences of the alternant without [s] are encountered. (For a summary of how this line of argumentation was deployed in earlier generative phonology, see again Kenstowicz and Kisseberth 1979.)

3.2 *Non-linear analyses of deletion*

Alternations are not the only source of evidence that can be exploited to recover deleted segments. A segment can sometimes leave its mark on output forms even in its absence. Historically, this situation can arise when a segment exerts some influence on a neighboring segment before being deleted. Crucially in such cases, deletion of the segment does not completely undo the effects of its influence.

By way of illustration, consider again the examples of hiatus-breaking vowel deletion exemplified by the three languages in (7). The situation in French (7a) and Karok (7b) is just as we would expect if deletion is viewed as targeting phonemes: one vowel vacates the representation, and the other segments simply shuffle up, leaving no evidence that the vowel was ever present in the input. In Ganda, on the other hand, deletion is accompanied by compensatory lengthening of the remaining vowel (7c). This suggests that, although the vowel-quality properties of the affected segment have been removed, its position within the word has been preserved. The deletion is thus only partial. A similar effect is witnessed in tonal stability, where a tone that is cut loose by the deletion of the vowel to which it is initially associated survives by attaching itself to a neighboring vowel (see CHAPTER 45: THE REPRESENTATION OF TONE).

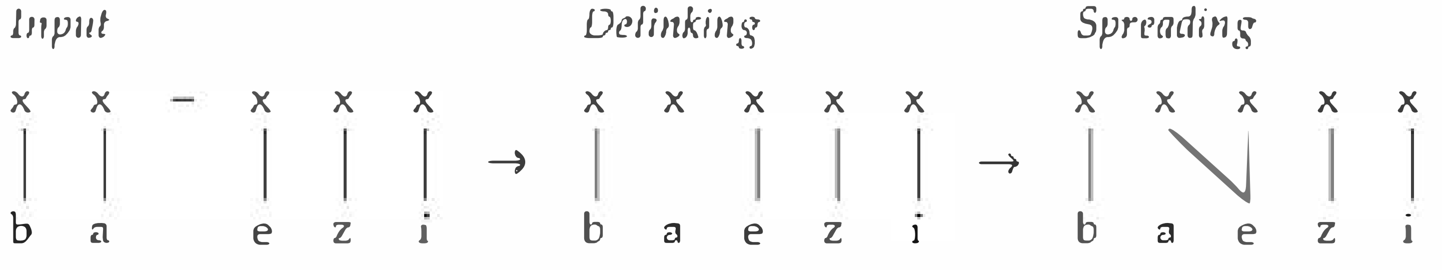
Stable weight and tone effects of this type are just part of a large body of evidence indicating that it is wrong to think of segments as indissoluble phonemic units. Rather, they are composites of non-linearly linked properties, each of which can be independently targeted by phonological processes, including deletion. Embracing this insight opens up different perspectives on deletion than that offered

[2] The caveat "almost certainly" is necessary here in light of recent studies showing that a speaker's production of a given form can be influenced in phonetically fine-grained ways by the existence of another, morphologically related form (see, for example, Ernestus and Baayen 2007). If this were true of the Samoan case, it would mean for example that the [o]s of the output forms [milo] and [ŋalo] would be phonetically distinct; each would bear some phonetic trace of the deleted root-final consonant that appears in the alternant found in the passive forms [milos-ia] and [ŋalom-ia], respectively. As far as I know, this effect has never been reported for Samoan deletion.

by linear-phonemic theory. In particular, the representational technology of non-linear theory has a significant impact on how we conceive of the notion of segment in deletion processes.

In a non-linear representation, a sharp distinction is drawn between the feature content that specifies a segment's phonetic quality and the position the segment occupies in syllable structure (CHAPTER 14: AUTOSEGMENTS; CHAPTER 54: THE SKELETON). In principle, deletion could target either of these entities or both simultaneously. Simultaneous targeting, equivalent to the phonemic conception of deletion, is what is required for the type of situation illustrated by French in (7a) and Karok in (7b).

Compensatory lengthening, such as we see in Ganda in (7c), requires an operation that removes the targeted segment's feature content but leaves its syllabic position intact. The usual non-linear way of implementing this is along the lines of the derivation /ba-ezi/ → [beːzi] in (11) (from Clements 1986; see also Kavitskaya 2002; CHAPTER 64: COMPENSATORY LENGTHENING; and the papers in Wetzels and Sezer 1986). The x-slots represent positions in syllable structure (the syllabification details are not relevant at this point); the alphabetic symbols are shorthand for complexes of feature specifications; and the lines indicate associations between feature complexes on the one hand and syllabic positions on the other.

(11)


Deletion is implemented by an operation that delinks the feature complex specifying *a* from its syllabic position. Compensatory lengthening consists in the insertion of a new link between *e* and the position vacated by *a*.

Representing deletion as in (11) relies on the principle that a segment must be linked to a syllabic position in order to be phonetically expressed (Goldsmith's 1990 Linkage Condition). Based on this principle, regular alternations involving deletion can be modeled as the attachment *vs.* non-attachment of a segment to a position. There are two main ways in which this notion has been implemented. One is illustrated by the delinking step in (11): the alternating segment has an underlying association to a position but loses it under certain phonological conditions. The other approach is to assume that the segment lacks an underlying association but acquires one under the complementary set of conditions (for applications of this analysis to *liaison* consonants in French, see for example Encrevé 1988 and Charette 1991). Applied to our Samoan example, the latter analysis would posit an unassociated *s* at the end of /milos/. Before a vowel, this "floating" segment attaches to a position and thus surfaces (as in [milos-ia]); elsewhere it is left stranded and thus phonetically unrealized (as in [milo]).

3.3 *Stray erasure*

There is a consensus in the literature that unsyllabified or "stray" segments get erased and thus fail to be phonetically realized (McCarthy 1979; Steriade 1982;

Itô 1986). However, there is less than general agreement about where this erasure occurs – whether within the phonological grammar or in phonetics. These two views can be summarized as follows:

(12) a. *Phonological erasure*
Unsyllabified segments are absent from phonological output.
b. *Phonetic erasure (Containment)*
All input segments are present in phonological output; unsyllabified segments are phonetically erased.

The notion in (12b) that all input segments are "contained" in output, including those that remain unsyllabified, was adopted by early OT (McCarthy and Prince 1993; Prince and Smolensky 1993). Although the notion has since been abandoned in favor of phonology-internal erasure (12a) in most OT work since Correspondence Theory (McCarthy and Prince 1995), it continues to figure in some current output-oriented approaches (see for example van Oostendorp 2004, 2007).

The two principles in (12) are empirically distinguishable, at least if we limit ourselves to a consideration of the effect each can have on phonological output. On its own, phonological erasure makes the prediction that a deleted segment will leave no trace of itself in an output form. Containment, on the other hand, predicts that, since an unlinked segment is still present in output, it should be capable of influencing or even triggering processes that affect linked segments.

There is at least one type of evidence that, it is generally agreed, can be seen as conforming to the prediction made by Containment. It involves floating tones, which are tones that lack an association to a vowel, often as a result of their original host vowel being deleted (see CHAPTER 45: THE REPRESENTATION OF TONE). Where such a tone remains unassociated in output, it can usually be seen to influence the pitch of a following tone, either raising it (*upstep*, indicating a floating high) or lowering it (*downstep*, a floating low). This suggests an analysis under which tones can be delinked without being erased.

More controversial is the claim that similar evidence can be found to support the notion that non-tonal features or feature complexes can also be delinked without being phonologically erased. Most of the relevant evidence involves derivational opacity.

A derivation is said to be opaque if it produces forms showing the effects of processes applying in specific ways not predicted by regular phonological conditions: either a process fails to occur where it would be expected to (underapplication), or it occurs where it would not be expected to (overapplication). It is worth considering this issue in this chapter, since deletion processes figure very prominently in the literature on opacity. This is because of their inherent potential to eliminate segments that trigger other processes.

As an illustration of opacity, consider the case of fortition in Cypriot Greek (Newton 1972; Coutsougera 2002). The input to the process is a glide resulting from a general Greek process that desyllabifies *i* before a vowel, as in /psumi-u/ → [psumj-u] 'bread (GEN)' (cf. [psumi] 'bread (NOM)'). The basic fortition pattern in Cypriot Greek is illustrated in (13a) and (13b): the glide hardens to a velar stop after *r* and to a palatal stop after other oral consonants.

(13) *Cypriot Greek*

	Input	*Output*	
a.	/ʃeri-a/	ʃerka	'hands'
	/teri-azo/	terkazo	'I match'
b.	/mmati-a/	mmaθca	'eyes'
	/pulluði-a/	pulluθca	'little birds'
c.	/ðonði-a/	ðoɲca	'teeth'
	/vasti-ete/	vascete	'he is held'
d.	/xarti-a/	xarca	'papers'
	/karði-a/	karca	'heart'

As shown in (13b), the clusters produced by gliding and hardening are also subject to independent processes of spirantization and voice assimilation.

The opacity arises in forms that are affected by another independent process, illustrated in (13c), which simplifies input three-consonant clusters by deleting the medial segment, as in /vasti-ete/ → [vascete] (*[vastcete]). The opaque examples appear in (13d). They contain clusters where *r* is followed by a palatal stop rather than the velar variant that the regular conditions on hardening would lead us to expect (cf. (13a)). Looking at the input forms, we can see where the source of the opacity lies: *r* is separated from the gliding/hardening site by a consonant. The intervening consonant fails to appear in output as a result of being deleted by the cluster simplification process. Expressed in traditional serial-derivation terms, hardening must precede deletion (essentially following in the steps of historical sound changes). An ordered-rule derivation of [xarca], for example, runs something like this: /xarti-a/ → xartja → xartca → [xarca]. At the point in the derivation where hardening applies, the consonant adjacent to the target segment is not *r* but *t*, which determines that the hardened consonant is palatal in output.

There are two main output-oriented approaches to opacity involving deletion, distinguishable on the basis of whether they subscribe to the phonological or the phonetic version of stray erasure in (12).

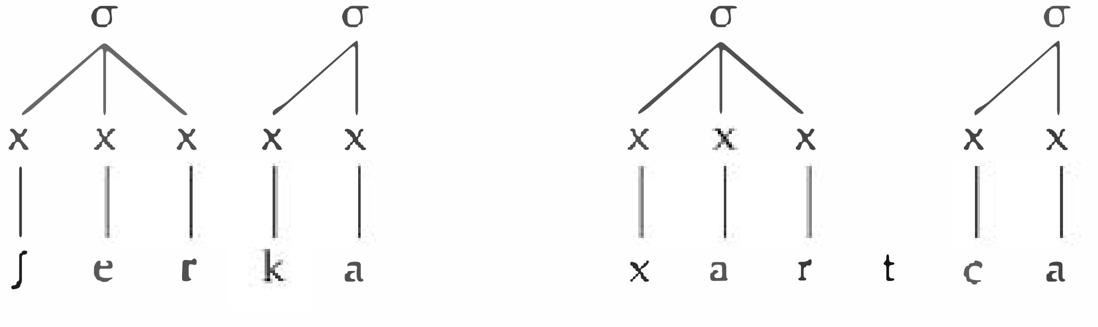
In the version of OT assuming phonological erasure, the basic approach to opacity is to allow the grammar to select a winning output candidate by cross-referring to a losing candidate that would figure as an intermediate form in an ordered-rule analysis. For example, in Cypriot Greek the attested output form [xarca] would be judged a more optimal output than *[xarka] on the grounds that it is the more similar of the two to non-surfacing *[xartca]. (For discussion of different approaches to how this cross-referencing can be achieved in OT, see McCarthy 2007.)

Containment approaches to deletion are equipped to treat opacity in a less obviously serial manner. Continuing with our Cypriot Greek example, consider the following analysis of [ʃerka] *vs.* [xarca].

(14) a. [ʃerka] 'hands' b. [xarca] 'papers'

input /ʃeri-a/ /xarti-a/

output



This analysis is of a type that was standard in early OT. It incorporates the widely held assumption that syllable structure is not present in lexical representation and is supplied by GEN, the mechanism that generates the set of candidate output forms.

In (14a), all of the input segments are parsed into syllable structure in output. The *k* of [ʃerka] is the hardened output of *i*, the velarity reflecting its adjacency to *r*.

In (14b), we see the effect of cluster simplification, driven by a high-ranked constraint banning complex codas (on which more later). The input *t* fails to be syllabified in output. In line with the Linkage Condition, a segment that is "underparsed" in this way is not phonetically realized. Nevertheless, in line with Containment, it remains in the output representation, where it can influence segments that fail to syllabify. In this example, *t* is in a position not only to trigger hardening but also to block *r* from causing the hardened output to be velar.

4 Conditions on consonant deletion

4.1 *Consonant deletion and syllabification*

What are the phonological conditions that favor whole-segment deletion? In answering this question, we need also to take into account processes of obstruent devoicing and consonant weakening, since these occur under similar conditions (see CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION and CHAPTER 66: LENITION). In fact, deletion figures prominently in what is perhaps the most reliable historical definition of phonological strength: segment A is weaker than segment B if A passes through a B stage on its way to deletion (Hyman 1975; Harris 1994).

There is a well-established tradition of approaching the question from the viewpoint of syllable structure. This is based on two main assumptions: certain syllabic positions are particularly favorable to deletion, and deletion changes the syllabification of the phonological forms it targets. These assumptions are themselves founded on a widely accepted model of syllabification that can be summarized as follows:

(15) *Standard syllabification model*

a. *Sonority*
Syllable nuclei always correspond to sonority peaks (typically vowels).

b. *Word edges*
i. A word-initial consonant forms a syllable onset.
ii. A word-final consonant forms a syllable coda.

These assumptions represent what can be considered the "standard" view of syllabification (see Vennemann 1972 and the literature summarized in Blevins 1995 and Zec 2007; see also CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE, and CHAPTER 55: ONSETS). However, it has increasingly been called into question, and this inevitably impacts on the validity of syllable-based analyses of deletion.

4.2 *Consonant deletion and syllabic markedness*

There are certain phonological contexts where consonants are especially vulnerable to deletion, particularly in clusters or at the end of a word. From the perspective of the standard model in (15), deleting a consonant in these contexts pushes syllable structure towards a less marked state. Removing a consonant from a cluster can open a previously closed syllable, simplify an onset, or reduce the size of a complex coda. Deleting a singleton consonant at the end of a word can be understood as having the same effect.

On the other hand, there are phonological contexts that are resistant to deletion. This is especially true of consonants at the beginning of a domain such as the word, stem, or foot, and to a lesser extent of word-internal prevocalic consonants. Here too we can detect a syllabic preference, in this case for syllables to contain at least one onset consonant.

There is a long tradition of interpreting these patterns to mean that consonant deletion is actively driven by a preference for less marked syllable structure. The tradition has been updated in OT by formulating the preferences in terms of markedness constraints (see CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS; also Prince and Smolensky 1993, Zec 2007, and the papers in Féry and van de Vijver 2003). In a given language, deletion can occur if any of these constraints outrank countervailing constraints that call for input consonants to be faithfully preserved in output. In what follows, this is the format I will employ to present the standard syllable-based approach to deletion, focusing on the following constraints:

(16) a. *Syllabic markedness constraints*
Codas
NoCoda
A syllable must be open.
NoComplexCoda
A coda must contain no more than one consonant.
Onsets
Onset
A syllable must have an onset consonant.
NoComplexOnset
An onset must contain no more than one consonant.

b. *Segmental faithfulness constraint*
MaxC
An input consonant must be preserved in output.

4.3 *Onset simplification*

(17) compares the grammars of two language types, one with complex onsets, the other without. As illustrated in (5), Sanskrit represents type (17a), while Pali represents (17b). The table shows two different output analyses of a schematic input form /CCV/, where CC potentially syllabifies as a complex onset. In grammar (17a), NoComplexOnset is outranked by MaxC, which allows the two input consonants to show up as an onset cluster. The reverse ranking in grammar (17b) forces onsets to be simplex: one input consonant takes up the only slot available in the onset, and the other drops (symbolized by C̲.)

(17) *Input* /CCV/

Language	*Output*	*Structure*	*Constraint ranking*
a	.CCV	Complex onset	MaxC >> NoComplexOns
b	.CC̲V	Simplex onset	NoComplexOns >> MaxC

A similar simplification effect is widely reported in first language acquisition. A child acquiring a language with complex onsets typically starts out by deleting the more sonorous of two onset consonants (as in *bring*, *blue*). The deletion evidently reflects a developmental stage where the only type of onset in the child's syllabic inventory is simplex. As is also typical of early phonology, the normally developing child nevertheless perceives the adult distinction between simple and complex onsets (as in *boo vs. blue*). This suggests that the liquid in consonant–liquid clusters is present in the child's lexical representations, but is phonologically erased in output as result of failing to find a berth in the child's simplified onset (see Bernhardt and Stemberger 1998 and the literature reviewed there).

4.4 Coda deletion

The syllable coda is widely regarded as the consonant deletion site *par excellence*. Under the standard syllabification view in (15), the coda subsumes consonants in three phonological contexts: (a) a word-final singleton consonant, (b) any member of a word-final cluster, and (c) the first member of a word-internal heterosyllabic cluster.

In an OT grammar, deletion of final singleton consonants (illustrated above by Samoan (1) and Lardil (2b)) results from MaxC being outranked by NoCoda. The reverse ranking defines languages that permit final consonants. This is shown in (18).

(18) *Input* /VC]/

Language	*Output*	*Structure*	*Constraint ranking*
a	VC.]	Final closed σ	MaxC >> NoCoda
b	V.<u>C</u>]	Final open σ	NoCoda >> MaxC

This basic treatment extends to consonant deletion in non-final codas, such as occurs in Diola Fogny (6).

As to the deletion of absolute word-final consonants in clusters (illustrated by Lardil in (3) and Catalan in (4)), the standard syllable-driven account is that it is motivated by a markedness imperative to reduce the complexity of codas. In OT terms, final cluster simplification occurs in languages of the type shown in (19b), where NoComplexCoda outranks MaxC (Prince and Smolensky 1993). The reverse ranking yields languages of type (19a), those with final clusters. The fact that words can be consonant final at all indicates that NoCoda is ranked relatively low in both of these grammars.

(19) *Input* /VCC]/

Language	*Output*	*Structure*	*Constraint ranking*
a	VCC.]	complex coda	MaxC >> NoComplexCoda, NoCoda
b	VC.<u>C</u>]	simplex coda	NoComplexCoda >> NoCoda, MaxC

Word-internal cluster simplification in Diola Fogny is selective in the type of coda it targets, deleting only those that do not form part of a partial geminate (see (6)). On their own, the syllabic markedness constraints in (16) are not enough to derive this selective behavior. What is required is an additional constraint that bans a coda from bearing or "licensing" certain feature specifications unless they are assimilated from a following onset. (The formulation of constraints of this type is based to a large extent on proposals by Itô 1986 and Goldsmith 1989.) With this constraint ranked high, a faithful output candidate such as *[u.juk.ja] is non-optimal, since the coda [k] bears its own place and manner specifications. Under these circumstances, the lower-ranked constraint NoCoda asserts itself, and the optimal candidate is the coda-less form [u.ju.ja] (an example of what in OT is called The Emergence of the Unmarked; see chapter 58: the emergence of the unmarked).

4.5 *Rethinking coda deletion*

According to the standard view in (15b.ii), a word-final consonant has the same syllabic status as a word-internal coda. This makes the strong prediction that whenever consonant deletion targets codas it should strike both of these positions simultaneously (chapter 36: final consonants).

There are certainly languages that bear this prediction out. In Lardil, for example, the set of non-apical singleton consonants that drop finally (as in (2b)) is largely the same as the set of consonants that are excluded from internal codas. In Samoan, the word-final deletion of root-final consonants shown in (1) coincides with an absence of internal closed syllables. Another example is provided by non-rhoticity in various languages, where historical *r* drops in exactly the combination of environments defined by the final-coda analysis. For example, in non-rhotic English, constricted *r* is suppressed both in an internal coda (as in *carnal*) and word finally (as in *car*).

Examples such as these represent some of the core evidence in favor of the assumption that consonants in internal codas and word-final positions are syllabified the same (again see Blevins 1995 for a summary of the relevant literature). However, alongside these examples we find deletion evidence that is difficult to square with the assumption. On the one hand are languages that preserve internal codas but either delete or lack word-final consonants; examples include Italian, colloquial Tamil, and Pali. On the other are languages that lack internal codas but have final consonants; examples include Kejaman (Strickland 1995), Yapese (Piggott 1999), and Central Sentani (Hartzler 1976).

Where historical final consonants can be reconstructed for the first of these syllabically "hybrid" types of language, they have either been deleted or survive as onsets followed by an epenthesized vowel. In Pali, for example, deletion has wiped out historically word-final consonants, which can be reconstructed on the basis of a comparison with cognate forms in Sanskrit (see (20a)) (Zec 1995). However, as shown in (20b), the process has not deleted internal codas, which survive as the first position of a partial or full geminate.

(20)		*Sanskrit*	*Pali*	
	a.	tatas	tato	'therefrom'
		punar	puno/puna	'again'
		pra:patat	papata	'hurled down'
	b.	danta	danta	'tamed'
		sapta	satta	'seven'
		karka	kakka	'precious stone'
		valka	vakka	'tree bark'
		kara	kaɳɳa	'ear'

It is true that internal codas in Pali have lost much of their historically contrastive feature content. However, the coda position itself has remained in place, picking up most or all of its feature interpretation from the following onset.

The existence of syllabically hybrid languages of the Pali type is inconsistent with the prediction that deletion will target internal codas and final consonants simultaneously. It could be further interpreted as undermining the assumption that final consonant deletion is motivated by syllabic structure independently of word structure. Does a final consonant delete because of its position in the syllable or because of its position within the word? Of course we could just say that a final consonant is a sub-type of coda, one that is particularly susceptible to deletion. But this is at best unparsimonious: it would be simpler just to say that deletion targets a consonant that is final in the word rather than having to say that it is final in both the word and the syllable.

In any event, the coda analysis faces additional problems, some of which become evident when we look more closely at consonant deletion in word-final clusters.

4.6 *Final cluster simplification*

According to the analysis outlined in §4.4, final cluster simplification results from the operation of the constraint NoComplexCoda in (16a). Two main objections have been leveled at this description. One has to do with the actual number of consonants permitted finally compared to internally, and the other with the phonotactics of final compared to internal clusters.

Cross-linguistically, there is a clear numerical asymmetry between final clusters and internal codas. Many languages, including English, which allow two or more consonants in final position only allow up to one consonant in an internal coda (Harris 1994; Duanmu 2008). It is hard to come up with clear examples of languages where the two positions exhibit equal complexity (still less languages where internal codas are actually more complex than final clusters).

Moreover, in languages showing the numerical mismatch between the two positions, it is noteworthy that the phonotactics of final clusters typically mimic those of initial or internal clusters containing onsets. Two basic patterns are attested in languages of this type. In one pattern, found in English, French, and Irish, for example, final clusters with a falling or level sonority profile (*mp, lt, pt*, etc.) share their phonotactics with internal coda–onset clusters (CHAPTER 53: SYLLABLE CONTACT). In the other – found in French, Polish, and Icelandic, for example – rising-sonority clusters (*pl, gl, dr*, etc.), share their phonotactics with internal or initial complex onsets (Harris 1994; Dell 1995; Harris and Gussmann 2002). In some languages, the two patterns overlap in three-consonant clusters; in French, for example, we find [rkl] both medially (as in [sɛrkle] 'circle (VB)') and finally (as in [sɛrkl] 'circle (N)').

If final clusters are treated as complex codas, the problem posed by cluster evidence from languages of this type is that the same set of phonotactic restrictions has to be assigned to two or even three different syllabic conditions. There have been two different responses to this problem, both of which challenge the claim that consonant deletion in final clusters is driven by some need to simplify complex codas. One persists with the notion that phonotactics are syllabically conditioned but proposes an alternative syllabification of final clusters to the standard one.

The other denies that cluster phonotactics are conditioned by syllable structure in the first place.

According to the first proposal, the simplest analysis of final clusters is to syllabify them in the same way as the internal clusters with which they show phonotactic parallels (Harris 1994; Dell 1995; Harris and Gussmann 2002). That is, final clusters of the non-rising sonority type (*mp*, *lt*, etc.) take the form coda plus onset, while those of the rising-sonority type (*pl*, *gr*, etc.) are complex onsets. While this captures the phonotactic parallels between final and internal clusters, it is questionable whether it provides a syllabic motivation for final cluster simplification. This is because there is a cross-linguistically strong preference for the second consonant to be targeted, regardless of whether the cluster is of the rising-sonority type or the non-rising (see the language surveys in Blevins 1995 and Côté 2004).

French provides us with an example of final deletion in rising-sonority clusters. In some varieties of the language the liquid is deleted in final clusters of this type, as in [pov<u>r</u>] 'poor', [minist<u>r</u>] 'minister' (Laks 1977; Côté 2004).

Deletion in final clusters of non-rising sonority is illustrated by the Catalan examples in (4). Feminine forms such as [askerpə] and [forta] show internal coda-onset clusters of falling sonority. The same clusters occurred historically in final position in masculine forms (Badia Margarit 1951). Of these final clusters, only those that are heterorganic survive into present-day Catalan, as in [əskerp] (see (4a)). The rest – those consisting of partial geminates – have been subject to simplification through deletion of the final consonant, as in /fort/ → [for] (see (4b)).

One conclusion that has been drawn from facts such as these is that the context for final cluster simplification is best stated in terms of word-finality without reference to syllable structure at all (as argued by Côté 2004, for example). A further reason for reaching this conclusion is the fact that the deletion that targets final clusters does not also automatically target the same clusters when they occur elsewhere in the word. For example, the dropping of the liquid in French final rising-sonority clusters does not also target initial and internal onsets. Final liquid deletion in French is thus not symptomatic of some more general simplification of complex onsets, such as has occurred in Pali for example (see (5)). By the same token, dropping the obstruent in Catalan final homorganic clusters, as in (4b), does not also target the same cluster when it occurs word internally.

Acknowledging that final cluster simplification might be best described without reference to syllable structure naturally raises the question of whether the same might be said of deletion in other phonological contexts. This line of argumentation has led some phonologists to claim that all patterns of consonant distribution should be expressed in a strictly linear fashion by referring exclusively to immediately adjacent segments and word boundaries (see e.g. Steriade 1999; Blevins 2003). At issue here is the fundamental question of why certain phonological positions promote deletion in the first place.

4.7 *What causes consonant deletion?*

What is it about consonants in word-final and pre-consonantal positions that makes them especially vulnerable to deletion? The answer is likely to be tied to the fact that these positions are also favorable sites for neutralization. Deletion is just one of a collection of process types that target consonants in these positions, and all of these processes tend to have the effect of neutralizing segmental contrasts. For example, these are the preferred contexts where we find that obstruent devoicing neutralizes laryngeal contrasts, debuccalization neutralizes place contrasts (e.g. /p t k / → [ʔ]), and vocalization neutralizes manner contrasts (e.g. /l r/ → [j]) (see CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). Deletion, too, can be seen as neutralizing, if it is understood as suspending the contrast between the presence and absence of a consonant, i.e. "merger with zero," as historical linguists put it (see Campbell 2004).

According to one proposal, the neutralizing tendency of certain positions can be attributed to the fact that they provide weaker auditory-perceptual cues to consonants than other positions (see e.g. Steriade 1999; Wright 2004). This point can be illustrated by comparing the cues projected by oral stops in different positions. The offset or release phase of a plosive provides more robust cues to its identity than does its approach phase (Bladon 1986; Ohala 1990). Moreover, the offset is most robustly cued when it is released onto a following vowel. Before another consonant or at the end of words, offset cues can be attenuated or suppressed altogether. In the first instance, they may be masked by the closure phase of the following consonant. In the second, since the end of a word often coincides with the end of an utterance, a final stop often offsets into silence.

Weakened cueing potential reduces the reliability with which listeners are able to detect consonant contrasts in particular positions, thereby increasing the likelihood that these contrasts will be eroded over time (see Ohala 1981, 1990). This would explain why preconsonantal and word-final positions are the most favorable sites for the neutralization of voice, place, and manner contrasts. Deletion, it can be argued, is an extreme manifestation of this overall effect (Côté 2004).

The question of whether consonant distribution and deletion are motivated by cueing potential or syllabic position continues to be debated. (For critiques of the "licensing-by-cue" approach, see for example Gerfen 2001 and Kochetov 2006.) It is not clear whether the two notions are in fact incompatible: the positions across which cueing potential is differentially distributed can in principle be stated in syllabic rather than linear terms.

On the face of it, neutralizing processes in general and deletion in particular seem to be communicatively dysfunctional, in that they suppress information that might otherwise be used to help keep words distinct from one another. However, the particular location of favored neutralizing positions within the word means that the impact of such processes on lexical distinctiveness is not as deleterious as it might otherwise have been. In lexical access, listeners rely much more heavily on phonological information at the beginning of words than towards the end (cf. Nooteboom 1981; Hawkins and Cutler 1988). It is no surprise, then, that of all phonological positions word-initial is the most resistant to deletion.

5 Conditions on vowel deletion

5.1 *Syllabification and vowel deletion*

If it is clear why there is a strong tradition of viewing consonant deletion as being driven by a pressure to simplify syllable structure, it is also easy to understand why no parallel tradition exists for vowel deletion. Deleting a vowel almost always increases syllabic markedness, at least according to the standard syllabification account. Apocope, illustrated by Lardil in (3), creates a final closed syllable (V.CV.] > VC.]). Syncope, illustrated by English in (8) and by Tonkawa and Tagalog in (9), creates a closed syllable and thus also a consonant cluster (V.CV.CV > VC.CV).

There is probably only one pattern of vowel deletion that can be straightforwardly viewed as reducing syllabic markedness: the type of hiatus-resolving elision seen in (7). The second vowel in a hiatus configuration occupies a syllable without an onset, a marked situation (acknowledged in OT by the ONSET constraint in (16a)).

Although apocope and syncope may not be syllabically motivated, they can generally be seen to be subject to other kinds of prosodic conditioning, specifically involving metrical or word structure or some combination of both. Moreover, the positions targeted by the two types of deletion can be broadly identified as prosodically weak or non-prominent. The processes are not always sensitive to stress, but when they are they typically target unstressed vowels. In stress-conditioned apocope, for example, the targeted vowel occurs either in the weak position of a foot or in an unfooted syllable. The emergence of word-final consonants in Catalan (see (4)) and certain other Romance languages is due to historical apocope of this type (Badía Margarit 1951; Lief 2006).

5.2 *Resyllabification?*

According to the sonority-driven model of syllabification in (15a), syncope and apocope necessarily have a much more profound impact on syllabification than consonant deletion: a vowel forms a local sonority peak and thus projects its own syllable nucleus, so removing it inevitably unleashes resyllabification. For example, by removing a final sonority peak, apocope forces a preceding consonant, originally an onset, into the coda of the preceding syllable.

This account of apocope predicts that the resyllabified consonant should take on the kind of coda-like behavior it did not exhibit when it was an onset. For example, it should now be able to trigger closed-syllable shortening in the preceding vowel. The prediction is not generally borne out. This is not surprising, since for stress purposes a final consonant typically behaves extrametrically: unlike an internal coda, it does not contribute to the weight of the syllable occupied by the preceding vowel (see Hayes 1995; CHAPTER 43: EXTRAMETRICALITY AND NON-FINALITY). Let us briefly consider two examples where developments accompanying apocope have not followed the path predicted by the final-coda analysis.

Modern English bears the marks of a limited form of historical closed syllable shortening, as a result of which certain consonants (very broadly speaking, non-coronals) can now appear in a word-internal coda only after a short vowel (see, for example, Myers 1987; Harris 1994). As the alternations in (21a) show, there is no parallel restriction when the corresponding consonants are word final.

(21) *English*

a. *perceptive* *perceive*
   reduction *reduce*

b. hɔpə > hɔːp > hoːp > howp *hope*
   bækə > bæːk > beːk > bejk *bake*

In Middle English, apocope of unstressed schwa was accompanied by lengthening of the stressed vowel in a preceding open syllable (Minkova 1991). If consonants rendered word final by apocope had resyllabified as codas, the non-coronals amongst them would have been expected either to prevent or to undo lengthening of a preceding vowel. As the examples in (21b) illustrate, that did not happen. Apocope in English thus failed to disturb the general Germanic pattern whereby a singleton word-final consonant, unlike an internal coda, has no influence on the length of a preceding vowel.

In Sesotho, final *i* has undergone apocope under certain quite specific phonological conditions. The effect of the change is evident in the locative suffix [-eŋ], which derives historically from [-eni] (the form still attested in some of Sesotho's Bantu sister languages) (Doke and Mofokeng 1957). There is clear evidence that apocope has not resulted in the nasal being resyllabified as a coda. One indication involves the widespread Bantu process that lengthens penultimate vowels in phrases of a certain type (with accompanying tonal effects not relevant here; see CHAPTER 114: BANTU TONE). The examples in (22a) illustrate this process in Sesotho (where vowel length is not lexically contrastive).

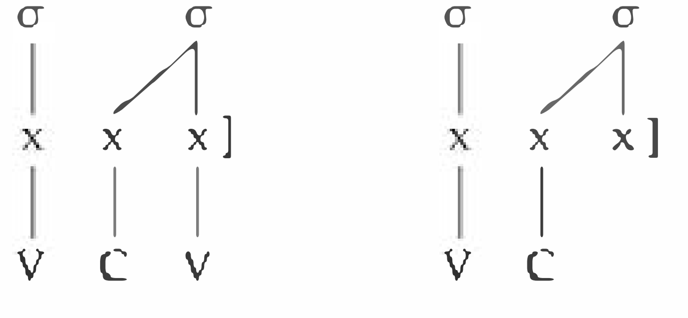
(22) *Sesotho*

a.	hase muːthu	'It's not a person'	
	kraː tla	'I am coming'	
b.	re teŋ lapeːŋ	'We are at home'	*laːpeŋ
	ba ile sedibeːŋ	'They have gone to the well'	*sediːbeŋ

If the final [ŋ] of the locative were a coda, penultimate lengthening would be predicted to target the vowel immediately preceding the suffix. However, as the forms in (22b) show, this is incorrect: the extra length falls instead on the suffix vowel itself. This suggests that the apocope that produced [-eŋ] has left the original bisyllabic structure of [-eni] undisturbed.

The more general conclusion we might draw from the evidence represented in (21) and (22) is that apocope has little or no impact on syllable structure. In fact it is just the sort of evidence that has led some phonologists to conclude that a word-final consonant is not a coda but the onset of a syllable with a phonetically unexpressed nucleus (see e.g. Giegerich 1985; Kaye 1990; Burzio 1994; Harris 1994; Piggott 1999; Harris and Gussmann 2002; Scheer 2004). According to this account, apocope only targets the feature content of a vowel while leaving its nuclear position untouched. This is schematized in (23) (where we abstract away from the issue of how the suppression of feature content is best represented).

(23) a. *Pre-apocope* b. *Post-apocope*



(A similar analysis has been applied word internally to syncope; see for example Charette's 1991 treatment of schwa in French.)

In (23), a consonant exposed to the right edge of a word by apocope remains syllabified as an onset. Since at no point does the consonant become a coda under this analysis, it is predicted not to trigger closed-syllable shortening. This is consistent with the scenario exemplified by English in (21). Similar reasoning would explain why penultimate lengthening in Sesotho targets the vowel immediately preceding word-final [ŋ] rather than the vowel before that, as shown in (22). The vowel before [ŋ] counts as the penultimate nucleus in the phrase, because there is another to its right, namely the final empty nucleus heading the syllable containing the nasal (as in [lapeːŋØ]).

Bearing on the question of whether apocope triggers resyllabification is the fact that the process is often reported to be phonetically continuous along the dimensions of duration and periodicity (see e.g. Silva 1994; Gordon 1998; Myers 2005). Similarly continuous effects are found with syncope; examples include Japanese (Beckman 1996), and the two types of syncope in English shown in (9) (Bybee 2001).

Phonetically continuous vowel deletion raises an awkward question for the standard model of syllabification: at what point do we decide that a fleeting vowel stops projecting a local sonority peak and causes a preceding consonant to resyllabify into a coda? This is particularly problematic where the gradience occurs within the speech of individual speakers. The sonority model suggests an implausible scenario in which a speaker's output flickers between one syllabification and the other. Under a stable-nucleus analysis, syllabification remains unaffected by vowel deletion, regardless of whether it is phonetically continuous or not. In the case of gradience, what varies is the manner in which the affected nucleus is phonetically expressed.

A further difference between the stable-nucleus and sonority-driven approaches to syllabification is that they offer empirically distinct perspectives on the relation between consonants flanking a syncope site. Under a sonority-driven analysis, the consonants start out as separate onsets but become syllabically adjacent after syncope. Newly formed clusters that happen to conform to existing phonotactic restrictions are wrongly predicted to be phonetically indistinguishable from clusters already existing outside the syncope context. In English, for example, the liquids in pairs such as *pa̲rade – prayed* and *po̲lite – plight* show differences in duration and voice onset time that listeners are able to utilize in word discrimination (Price 1980). One conclusion that might be drawn from this is that *p* and *r* are not phonotactically adjacent in *pa̲rade* in the way that they are in *prayed*. This is consistent with the view that, rather than triggering resyllabification, syncope of a vowel leaves the flanking consonants in phonotactically independent onsets separated by a stable but variably expressed nucleus.

6 Conclusion

Over the years, there has been a significant shift in the way phonologists model phonological processes in general and whole-segment deletion in particular. Previously, deletion processes were conceived of as rules that remove segments from linear strings of input phonemes. That view gave way to one according to which deletion rules selectively target elements within non-linear representations. Later, input-oriented rules were largely abandoned in favor of the notion that deletion results from the operation of output-oriented constraints.

For all the differences amongst the various approaches reviewed here, there remains an important shared assumption: that morphologically related forms showing a regular alternation between a segment and zero should be derived from a single lexical source. However, this idea too has been increasingly called into question.

Under the standard derivational analysis of deletion in Samoan, recall, the historical consonant-final form of an alternating root such as [milo] – [milos(-ia)] is preserved in a single underlying representation and is deleted in final position by a synchronic analogue of sound change (/milos/ → [milo]). The main argument in favor of this analysis is economy: speakers need memorize only one lexical form of each alternating root instead of two.

This reasoning was already being questioned as early as 1973 by Hale, citing evidence from Maori, a Polynesian relative of Samoan that shares the same historical deletion of word-final consonants. The evidence strongly suggests that speakers of present-day Maori have re-analyzed the original morphology of alternating forms by treating the formerly root-final consonant as now belonging to the suffix. This yields historical reparsings such as [awhit-ia] > [awhi-tia] 'embrace (PASS)' and [hopuk-ia] > [hopu-kia] 'catch (PASS)'. If this is correct – and it is corroborated by more recent evidence adduced by Eliasson (1990) – it indicates that the segment–zero alternation in this case is now a matter of allomorphy rather than regular phonology (CHAPTER 99: PHONOLOGICALLY CONDITIONED ALLOMORPH SELECTION): for each root, speakers simply memorize the appropriate consonant-initial suffix. (In fact, one of the re-analyzed suffix forms, *-tia*, has a much wider lexical distribution than the others and is the one now used by default in loanwords, neologisms, and code-switching; Eliasson 1990.) Similar historical restructurings involving consonant deletion have been reported for other languages (see e.g. the survey of the Sulawesi group of Western Malayo-Polynesian languages in Sneddon 1993).

Evidence from historical restructurings of this type presents an obvious challenge to any derivational treatment of segment deletion, be it formulated in terms of input-oriented rules or output-oriented constraints. In any case, the economy argument in favor of unique underlying forms is no longer as persuasive as it might have once seemed, since there are now known to be practically no limits on the storage capacity of lexical memory (Braine 1974; Landauer 1986).

Since Hale's (1973) paper, models of lexical storage and access have emerged that allow us to capture phonological connections between morphologically related forms without necessarily deriving them from a single lexical source. Part of the process of auditory word recognition involves a lexical search that leads to a specific neighborhood containing a number of forms with similar phonological characteristics (see the literature review in McQueen and Cutler 1997). Even if two alternants of the same morpheme are stored as separate lexical entries, they are thus likely to show up in the same search, depending on how phonologically similar they are to one another. In the Samoan deletion case, related forms such as *milo* and *milosia* can have separate addresses in the lexicon (which would account for their allomorphic behavior) and yet still be close neighbors and thus be accessed in unison. The fact that one form lacks a consonant that is present in the other is of course a source of dissimilarity. However, this is mitigated by the location of the dissimilarity – away from the initial portion of the forms that is known to provide the most valuable information in word recognition.

ACKNOWLEDGMENTS

Thanks to the following for their valuable comments on earlier drafts of this chapter: Beth Hume, Marc van Oostendorp, Shanti Ulfsbjörninn, and two anonymous reviewers.

REFERENCES

Badia Margarit, Antonio. 1951. *Gramática histórica catalana.* Barcelona: Noguer.

Beckman, Mary E. 1996. When is a syllable not a syllable? In Takashi Otake & Anne Cutler (eds.) *Phonological structure and language processing: Cross-linguistic studies*, 95–123. Berlin & New York: Mouton de Gruyter.

Bernhardt, Barbara H. & Joseph P. Stemberger. 1998. *Handbook of phonological development from the perspective of constraint-based nonlinear phonology.* San Diego: Academic Press.

Bird, Steven. 1995. *Computational phonology: A constraint-based approach.* Cambridge: Cambridge University Press.

Bladon, Anthony. 1986. Phonetics for hearers. In Graham McGregor (ed.) *Language for hearers*, 1–14. Oxford: Pergamon Press.

Blevins, Juliette. 1995. The syllable in phonological theory. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 206–244. Cambridge, MA & Oxford: Blackwell.

Blevins, Juliette. 2003. The independent nature of phonotactic constraints: An alternative to syllable-based approaches. In Féry & van de Vijver (2003), 375–403.

Bloomfield, Leonard. 1933. *Language.* New York: Holt.

Braine, Martin D. S. 1974. On what might constitute learnable phonology. *Language* 50. 270–299.

Bright, William. 1957. *The Karok language.* Berkeley & Los Angeles: University of California Press.

Burzio, Luigi. 1994. *Principles of English stress.* Cambridge: Cambridge University Press.

Bybee, Joan. 2001. *Phonology and language use.* Cambridge: Cambridge University Press.

Campbell, Lyle. 2004. *Historical linguistics: An introduction.* 2nd edn. Cambridge, MA: MIT Press.

Charette, Monik. 1991. *Conditions on phonological government.* Cambridge: Cambridge University Press.

Clements, G. N. 1986. Compensatory lengthening and consonant gemination in LuGanda. In Wetzels & Sezer (1986), 37–77.

Côté, Marie-Hélène. 2004. Consonant cluster simplification in Quebec French. *Probus* 16. 151–201.

Coutsougera, Photini. 2002. The semivowel and its reflexes in Cypriot Greek. Ph.D. dissertation, University of Reading.

Dell, François. 1995. Consonant clusters and phonological syllables in French. *Lingua* 95. 5–26.

Doke, Clement M. & S. Machabe Mofokeng. 1957. *Textbook of Southern Sotho grammar.* Cape Town: Longman.

Duanmu, San. 2008. *Syllable structure: The limits of variation.* Oxford: Oxford University Press.

Eliasson, Stig. 1990. English–Maori language contact: Code-switching and the free morpheme constraint. In Rudolf Filipović & Maja Bratanić (eds.) *Languages in contact: Proceedings of the 12th International Conference of Anthropological and Ethnological Sciences, Zagreb 1988*, 33–49. Zagreb: Institute of Linguistics.

Encrevé, Pierre. 1988. *La liaison avec et sans enchaînement: Phonologie tridimensionnelle et usages du français.* Paris: Éditions du Seuil.

Ernestus, Mirjam & R. Harald Baayen. 2007. Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch. *Language and Cognitive Processes* 22. 1–24.

Féry, Caroline & Ruben van de Vijver (eds.) 2003. *The syllable in Optimality Theory.* Cambridge: Cambridge University Press.

Gerfen, Chip. 2001. A critical view of licensing by cue: Codas and obstruents in Eastern Andalusian Spanish. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory*, 183–205. Cambridge: Cambridge University Press.

Giegerich, Heinz J. 1985. *Metrical phonology and phonological structure: German and English.* Cambridge: Cambridge University Press.

Goldsmith, John A. 1989. Licensing, inalterability and harmonic rule application. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 25. 145–156.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology.* Oxford & Cambridge, MA: Blackwell.

Gordon, Matthew. 1998. The phonetics and phonology of non-modal vowels: A cross-linguistic perspective. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 24. 93–105.

Hale, Kenneth. 1973. Deep–surface canonical disparities in relation to analysis and change: An Australian example. In Thomas A. Sebeok (ed.) *Current trends in linguistics*, vol. 11: *Diachronic, areal, and typological linguistics*, 401–458. The Hague: Mouton.

Harris, John. 1994. *English sound structure.* Oxford: Blackwell.

Harris, John & Edmund Gussmann. 2002. Word-final onsets. *UCL Working Papers in Linguistics* 14. 1–42.

Hartzler, Margaret. 1976. Central Sentani phonology. *Irian: Bulletin of Irian Jaya Development* 5. 66–81.

Hawkins, John A. & Anne Cutler (1988). Psycholinguistic factors in morphological asymmetry. In John A. Hawkins (ed.) *Explaining language universals*, 280–317. Oxford: Blackwell.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Hoijer, Harry. 1946. Tonkawa. In Harry Hoijer, L. Bloomfield, M. R. Haas, A. M. Halpern, F. K. Li, S. S. Newman, M. Swadesh, G. L. Trager, C. F. Voegelin & B. L. Whorf (eds.) *Linguistic structures of native America*, 289–311. New York: Viking Fund.

Hyman, Larry M. 1975. *Phonology: Theory and analysis.* New York: Holt, Rinehart & Winston.

Itô, Junko. 1986. Syllable theory in prosodic phonology. Ph.D. dissertation, University of Massachusetts. Published 1988, New York: Garland.

Kavitskaya, Darya. 2002. *Compensatory lengthening: Phonetics, phonology, diachrony.* London & New York: Routledge.

Kaye, Jonathan. 1990. "Coda" licensing. *Phonology* 7. 301–330.

Kenstowicz, Michael & Charles W. Kisseberth. 1979. *Generative phonology: Description and theory.* New York: Academic Press.

Kochetov, Alexei. 2006. Testing licensing by cue: A case of Russian palatalized coronals. *Phonetica* 63. 113–148.

Laks, Bernard. 1977. Contribution empirique à l'analyse socio-différentielle de la chute de /r/ dans les groupes consonantiques finals. *Langue française* 34. 109–125.

Landauer, Thomas K. 1986. How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science* 10. 477–493.

Lief, Eric A. 2006. Syncope in Spanish and Portuguese: The diachrony of Hispano-Romance phonotactics. Ph.D. dissertation, Cornell University.

Mascaró, Joan. 1983. *La fonologia catalana i el cicle fonològic.* Bellaterra: Universitat Autònoma de Barcelona.

McCarthy, John J. 1979. Formal problems in Semitic morphology and phonology. Ph.D. dissertation, MIT. Published 1985, New York: Garland.

McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory.* London: Equinox.

McCarthy, John J. & Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.

McQueen, James M. & Anne Cutler. 1997. Cognitive processes in spoken-word recognition. In W. J. Hardcastle & John Laver (eds.) *The handbook of phonetic sciences*, 566–585. Oxford: Blackwell.

Minkova, Donka. 1991. *The history of final vowels in English: The sound of muting*. Berlin & New York: Mouton de Gruyter.

Myers, Scott. 1987. Vowel shortening in English. *Natural Language and Linguistic Theory* 5. 485–518.

Myers, Scott. 2005. Vowel duration and neutralization of vowel length contrasts in Kinyarwanda. *Journal of Phonetics* 33. 427–446

Newton, Brian. 1972. *Cypriot Greek: Its phonology and inflections*. The Hague: Mouton.

Nooteboom, Sieb G. 1981. Lexical retrieval from fragments of spoken words: Beginnings vs. endings. *Journal of Phonetics* 9. 407–424.

Ohala, John J. 1981. The listener as a source of sound change. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 17(2). 178–203.

Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 258–275. Cambridge: Cambridge University Press.

Oostendorp, Marc van. 2004. Phonological recoverability in dialects of Dutch. Unpublished ms., Meertens Institute, Amsterdam (ROA-657).

Oostendorp, Marc van. 2007. Derived environment effects and Consistency of Exponence. In Sylvia Blaho, Patrik Bye & Martin Krämer (eds.) *Freedom of analysis?*, 123–148. Berlin & New York: Mouton de Gruyter.

Piggott, Glyne L. 1999. At the right edge of words. *The Linguistic Review* 16. 143–185.

Pratt, George. 1911. *Pratt's grammar dictionary and Samoan language*. Reprinted 1977, Apia: Malua Printing Press.

Price, P. J. 1980. Sonority and syllabicity: Acoustic correlates of perception. *Phonetica* 37. 327–343.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Sapir, J. David. 1965. *A grammar of Diola-Fogny*. Cambridge: Cambridge University Press.

Scheer, Tobias. 2004. *A lateral theory of phonology*, vol. 1: *What is CVCV, and why should it be?* Berlin & New York: Mouton de Gruyter.

Scobbie, James M., John Coleman & Steven Bird. 1996. Key aspects of declarative phonology. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, vol. 2, 685–709. Salford: ESRI.

Silva, David J. 1994. The variable elision of unstressed vowels in European Portuguese: A case study. *University of Texas at Arlington Working Papers in Linguistics* 1. 79–94.

Sneddon, J. N. 1993. The drift towards final open syllables in Sulawesi languages. *Oceanic Linguistics* 32. 1–44.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT.

Steriade, Donca. 1999. Alternatives to syllable-based accounts of consonantal phonotactics. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.) *Item order in language and speech*, 205–245. Prague: Karolinum Press.

Strickland, Simon. 1995. *Materials for the study of Kejaman-Sekapan oral tradition*. Kuching: Sarawak Museum.

Vennemann, Theo. 1972. On the theory of syllabic phonology. *Linguistische Berichte* 18. 1–18.

Wetzels, W. Leo & Engin Sezer (eds.) 1986. *Studies in compensatory lengthening*. Dordrecht: Foris.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 34–57. Cambridge: Cambridge University Press.

Zec, Draga. 1995. The role of moraic structure in the distribution of segments within syllables. In Jacques Durand & Francis Katamba (eds.) *Frontiers of phonology: Atoms, structures, derivations*, 149–179. London & New York: Longman.

Zec, Draga. 2007. The syllable. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 161–194. Cambridge: Cambridge University Press.

69 Final Devoicing and Final Laryngeal Neutralization

Gregory K. Iverson
Joseph C. Salmons

1 Introduction

In this chapter, we survey a set of phenomena that have traditionally been given the simple rubric "final devoicing." This name, however, clearly conflates a number of different phonological phenomena – deletion of other laryngeal features, even feature addition – and the relevant general pattern is better characterized as "final laryngeal neutralization." Widely attested across the languages of the world, final laryngeal neutralization represents a prototypical positional merger of phonological contrasts. Nonetheless, the attested patterns vary along several dimensions; they provide challenges to current phonological frameworks, on the one hand, and allow good testing grounds for them, on the other. In particular, the topic is highly relevant to the ongoing debates over the relationship between universal grammar and language change in shaping sound systems, such as Blevins (2004, 2006) *vs.* Kiparsky (2006, 2008) (see also CHAPTER 93: SOUND CHANGE).

In §2, after briefly reviewing some basic data, we provide a cross-linguistic survey of attested patterns in which (phonological) laryngeal features neutralize at right edges of prosodic constituents, while §3 introduces some aspects of the phonetics of final laryngeal neutralization with an eye toward what those aspects mean for the phonology of such processes. In §4, we then give an overview of the role of prosodic domains along which such patterns arise and generalize in sound change. In §5, we explore the major question in current phonology connected with this issue: the relationship between historical development and Universal Grammar. §6 summarizes and concludes.

We focus on laryngeal features, leaving aside broader neutralizations, such as loss of length distinctions in final position (e.g. Kümmel 2007: 133–136), although it is important to note that these are often connected (see Trubetzkoy 1977: 74). We also restrict discussion to processes affecting obstruents, although final vowel devoicing and other such phenomena are attested, as surveyed by Barnes (2002: ch. 3). Finally, we focus on "dynamic," alternation-supported neutralizations, leaving aside the "static" absence of contrasts, such as those in Thai and varieties of Quechua, in which laryngeal contrasts found in onsets (or word-initial position, etc.) are merely missing in codas (or word-final position, etc.), without morphophonemic alternation.

2 The phonological typology of final laryngeal neutralization

2.1 *Some basic data*

In many languages of the world, underlyingly voiced speech sounds do not show glottal pulsing at the ends of words, with the result that they are largely or often entirely indistinguishable from voiceless sounds. This occurs frequently with all obstruents, as in Catalan (Hualde 1992) or as illustrated here by nominal alternations from Polish (Rubach 1997: 553):

(1)

nom sg		*nom pl*		
klub	[p]	*klub-y*	[b]	'club'
majonez	[s]	*majonez-y*	[z]	'mayonnaise'
staw	[f]	*staw-y*	[v]	'pond'
kandelabr	[pr]	*kandelabr-y*	[br]	'lamp'

As the last example demonstrates, Polish shows a variety of wrinkles in devoicing. Phonologically, extrasyllabic consonants, /r/ in this case, do not block devoicing (see also CHAPTER 36: FINAL CONSONANTS). As Tieszen (1997) shows, Polish manifests complex regional and acoustic patterns, which include evidence for incomplete neutralization for at least some speakers.

Targets of this neutralization vary considerably across languages. Sometimes, not all obstruents alternate. In Turkish (Kopkallı 1993; Becker *et al.* 2008) most stops surface as voiceless finally (and with a degree of aspiration as well, per Vaux and Samuels 2005: 418), but may alternate with a voiced counterpart in suffixed word forms (data from Kopkallı 1993: 29; cf. also Nicolae and Nevins 2009; Feizollahi 2010):

(2) *Turkish final stop neutralization*

	nom sg	*acc sg*	
/kab/	[kap]	[kabɯ]	'container'
/kanad/	[kanat]	[kanadɯ]	'wing'

This pattern, however, does not extend to fricatives:

(3) *Turkish final fricative voicing alternation*

[af]	'pardon'	≠	[av]	'hunting'
[kas]	'muscle'	≠	[kaz]	'goose'

While our focus is on obstruent neutralizations, we note that, in other languages, final sonorants (CHAPTER 8: SONORANTS) may be subject to devoicing, as in Kaqchikel (Campbell 1998: 41):

(4) *Kaqchikel sonorant devoicing*

/kar/	[kar̥]	'fish'
/kow/	[koʋ̥v]	'hard'

We will not review the various technical approaches to laryngeal alternations among obstruents that have been offered in the past, such as Trubetzkoy's (1977: 71) analysis of German final neutralization in terms of archiphonemes, which treats the neutralized obstruents as identical neither with the underlying *Media* (lenis/voiced) nor *Tenuis* (fortis/voiceless). In modern work, these processes have been characterized as rules deleting the feature [voice] (or sometimes imposing [−voice]) or as constraints prohibiting that feature in final position, and that is our point of departure here as well.

But final laryngeal neutralization can also consist in the addition of a contrastive property, not just the loss of one exemplified in the cases involving removal of [voice] adduced above. In particular, the work of Vaux and Samuels (2005: 418–422) has shown that the addition of aspiration to stops in final position is cross-linguistically rather more common than has been appreciated. As they review for Kashmiri, contrasts between plain and aspirated final voiceless stops are merged in favor of the aspirated series, presumably via a rule that accrues the privative feature [spread glottis] (or [+spread glottis], in a binary system) to voiceless stops at the end of the word.[1]

(5) *Final aspiration in Kashmiri* (Vaux and Samuels 2005: 420, citing Syeed 1978)

	nom sg	*dat pl*	*agent pl*	
/wat/	[wat^h]	[watan]	[watau]	'way'
/kat^h/	[kat^h]	[kat^{h}an]	[kat^{h}au]	'story'

Similarly, in Klamath a three-way contrast among voiceless aspirated, ejective, and plain stops neutralizes to the aspirated series in word-final position. Such patterns clearly suggest feature addition.

(6) *Final aspiration in Klamath* (Vaux and Samuels 2005: 421, citing Blevins 1993)

/n'ep^h/	[n'ep^h]	'hand'	cf.	[n'ep^{h}eːʔa]	'puts on a glove'
/ntʃ'ek'/	[ntʃ'ek^h]	'in little bits'		[ntʃ'ek'aːni]	'small, little'
/nkak/	[nkak^h]	'turtle (sp.)'		[nkakam]	'turtle (ross)'

Vaux and Samuels take the existence of patterns such as these to justify the interpretation of aspiration as "unmarked" in languages that add this feature finally, attempting to salvage the widely subscribed view that neutralization in final position regularly entails merger to the unmarked member of a contrast (CHAPTER 2:

[1] Sadaf Munshi (personal communication) indicates that there are systematic exceptions to final aspiration in Kashmiri, apparently connected with patterns of historical apocope, e.g. [op] 'person who can't keep a secret'; [mot] 'madman'. She further points out (Munshi 2006: 58ff., personal communication) that, in Burushaski, the three-way voiced/aspirated/voiceless unaspirated contrast is regularly neutralized to voiceless unaspirated in final position, affecting stops, affricates, and fricatives, even in loanwords, whereas in Kashmiri final (aspirating) neutralization affects only stops.

CONTRAST; CHAPTER 4: MARKEDNESS), albeit now defined on a language-specific rather than universal basis. More generally, however, aspiration patterns such as Vaux and Samuels reveal have been simply disregarded, with final laryngeal neutralizations often all being treated without further differentiation as "final devoicing." The idea that such non-assimilatory neutralization involves feature loss rather than addition has been particularly strong: Lombardi (2001: 13, *passim*) starts from the position that "The laryngeal distinctions of voicing, aspiration, and glottalization are often neutralized to plain voiceless in coda position," exemplified by what she analyzes as the removal of [voice] in German codas. More recently, Kiparsky (2008: 46) uses the term "devoicing" to characterize the Korean process of final neutralization across three series – usually treated as lenis (laryngeally unmarked), aspirated, and tense. But the phonetic feature [voice] plays no phonological role in Korean on most views (cf. Avery and Idsardi 2001; Ahn and Iverson 2004), appearing only allophonically in the otherwise voiceless lenis series in intersonorant contexts because this is a position favorable to passive voicing. This aside, however, the characterization of final laryngeal neutralization generally as final devoicing has far-reaching implications for the nature of the phonological component in human grammar. Thus, Kiparsky (2006: 222) argues forcefully that "marked feature values are suppressed in 'weak' prosodic positions." On the question of how to formalize this, he writes:

> The right way to do it in my opinion is that constraints can single out marked feature values (but not unmarked feature values). From these, with certain additional assumptions, we can build a system of constraints that asymmetrically prohibit marked feature values in weak positions. In processual terms, it predicts the existence of coda devoicing (coda depalatalization, debuccalization, deaspiration, etc.) and excludes coda voicing (coda palatalization, buccalization, aspiration, etc.).

On this view, *ex nihilo* feature insertion or addition is impossible, prohibited by the "design of language." In the next section, however, we develop a typology of final neutralizations that includes a broad set of counterexamples to the claim that final neutralization invariably entails feature loss, or merger to the unmarked, but rather may involve motivated neutralization to marked feature values as conventionally (i.e. universally) construed.

2.2 *Overview*

We begin this section with an overview of laryngeal features, so as to provide a framework for discussing which ones participate in final neutralization and how. But it is clear at the outset that right edges of prosodic constituents are frequent loci of neutralization and reduction of many kinds. In this spirit, final laryngeal neutralization has often been regarded as a "subtype of final weakening" (Hock 1999: 19; Harris 2009; Honeybone, forthcoming), and thus related to lenition (CHAPTER 66: LENITION) and final consonant loss (CHAPTER 68: DELETION). Indeed, the view expressed just above (Kiparsky 2006) is that "weak positions" are actually governed by this directionality, so that final weakening in various forms is to be expected, but not final strengthening. We shall see, however, that a full typology of final laryngeal neutralization must also recognize the occurrence of final strengthening, or, as we shall refer to it, final fortition.

2.3 *Laryngeal features*

Now common in discussions of neutralization of "voicing" and other distinctions involving glottal states is the perspective known as "laryngeal realism" (cf. Iverson and Salmons 1995, 2003, 2006, 2007, 2009, the name taken from Honeybone 2005), which we will adopt here. On this view of laryngeal phonology, three privative features are considered to be sufficient to represent the known relevant contrasts in languages: [voice], [spread glottis] (henceforth [spread]), and [constricted glottis] ([constricted]) (see also CHAPTER 17: DISTINCTIVE FEATURES). In two-way systems, a "voice" language such as Spanish distinguishes marked voiced stops ([voice]) from unmarked voiceless unaspirated stops ([ ], using a blank space to indicate the absence of a phonological specification), whereas an "aspiration" language like English distinguishes marked aspirated or fortis stops ([spread]) from unmarked lenis (albeit often passively voiced) stops ([ ]). And a "glottalic" language like K'ekchi distinguishes marked ejectives ([constricted]) from unmarked, typically voiceless unaspirated stops ([ ]).

Combinations of these possibilities also exist to make up three-way contrasts, as in the aspirated–voiced–plain system of Thai ([spread], [voice], [ ]), the aspirated-ejective–plain system of Klamath ([spread], [constricted], [ ]) or the aspirated–implosive–plain system of Vietnamese ([spread], [constricted & voice], [ ]). Four-, five-, and even six-way systems are also attested, as laid out first by Ladefoged (1973) and charted under the present feature system by Iverson and Salmons (1995). For example, the four-way system of Hindi adds murmured (breathy or voiced aspirated) stops to the three types of distinctions found in Thai, via paradigmatic as well as syntagmatic combination of [voice] with [spread] ([voice], [spread], [voice & spread], [ ]). The thrust of this minimalist representation is thus to reconcile not only the phonetics but also the phonological, historical, and acquisitional behavior of speech sounds with their featural characterization.

Perhaps the most familiar neutralization of laryngeal contrasts among obstruents at the right edge of a prosodic constituent is the final neutralization process in German, documented comprehensively by Brockhaus (1995) and interpreted in the light of laryngeal realism by Iverson and Salmons (2007). Taking the realist perspective, German is an "aspiration language" in the sense described above, meaning that the laryngeal merger that takes place between fortis (aspirated) and lenis (passively voiced) obstruents syllable-finally is in fact final fortition (as implied by the descriptive German grammatical term, *Auslautverhärtung*), not final devoicing. (The domain varies regionally and stylistically between syllable-final and word-final.) In contrast to its sister Dutch, then, where final neutralization is devoicing (given that Dutch is a "voice language" in the sense described above), German neutralizes final laryngeal distinctions through feature addition rather than loss. The two types are illustrated below:

(7) a. *Final devoicing:* /d/ → [t] (Dutch)

d]$_\sigma$	Phonemic contrast:	/d/	/t/
ǂ		|	|
[voice]		[voice]	[ ]

b. *Final fortition:* /d̥/ → [tʰ] (German)

```
d̥ ]σ        Phonemic contrast:    /d̥/      /tʰ/
↑                                   |        |
[spread]                           [ ]    [spread]
```

In some languages, lack of required final release or other factors may make it difficult to distinguish between feature addition and feature removal, but the former pattern is securely and robustly attested.[2] For instance, in his treatment of Kashaya (Pomoan), Buckley (1994: 87–88) illustrates the rule of "Coda Aspiration" with respect to the palatal stop /c/:

(8) /s'uwac-i/ → s'uwaci 'dry it! (SG)'
/s'uwac-me-ʔ/ → s'uwacʰmeʔ 'dry it! (FORMAL)'

In this language, an underlyingly "plain" or laryngeally unmarked stop at the end of a word-internal syllable becomes "aspirated phonologically, and not simply subject to some rule of obligatory final release at the phonetic level" (Buckley 1994: 88). The aspiration rule also applies to word-final stops in loanwords, such as /'cajnikʰ/ 'teapot' (cf. Russian /ʧajnik/) and /ʃaki'taqʰ/ 'puffin' (cf. Alutiiq /ʃaki'ta-q/) (1994: 100).

Kashaya, in fact, shows a three-way contrast among plain, aspirated, and glottalic stops, but other processes besides coda aspiration contribute to the non-occurrence of plain finals. Thus, a suffix in a morphological category like the assertive (/=ʔ/) combines with a stem-final plain stop to yield a word-final glottalic (see Fallon 2002 on the notion of "fusion"):

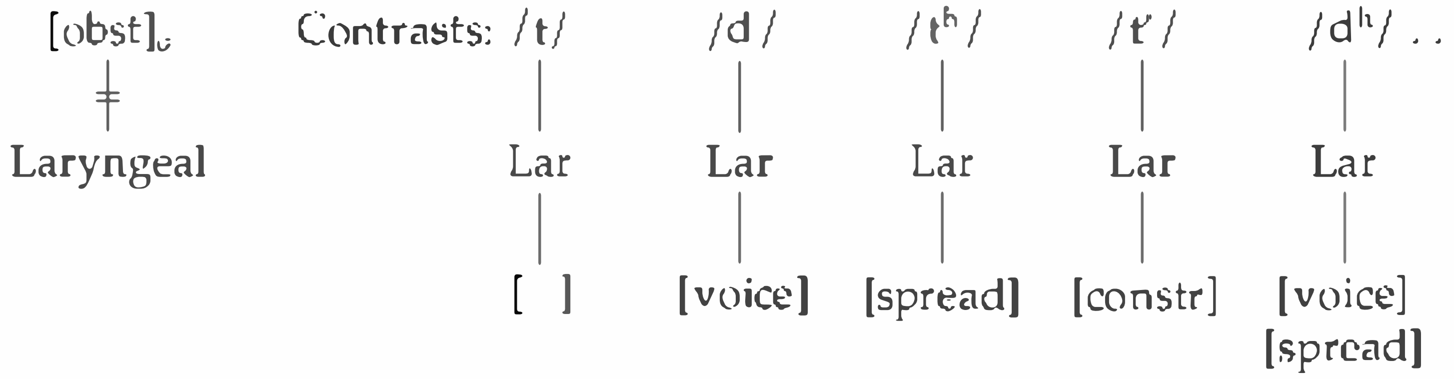
(9) /qahmat=ʔ/ [qahmát'] 'he's angry' cf. /qahmat/ [qahmáʔ] 'angry'

At the same time, non-verbs ending in a glottalic retain that ejective articulation (as in [hosiq'] 'screech owl'), whereas word-final plain stops in the native vocabulary, analyzed as laryngeally empty, debuccalize to glottal stop ([qahmáʔ] ← /qahmat/). In other words, the language aspirates plain stops in native word-internal codas (and word finally in loanwords), retains word-final phonemically aspirated and glottalic stops, but debuccalizes remaining word-final stops, with the result that final plain stops are phonetically absent.

As already noted above, a collection of similar cases, including Kashmiri and Klamath, has been adduced by Vaux and Samuels (2005). Ejectives and plain stops in Klamath neutralize to aspirates word finally (Blevins 1993), and Yu (2008, personal communication) reports the same pattern for Washo. In Kashaya, however, plain voiced stops do not undergo the neutralization to aspirates that affects the language's ejectives, just as in Kashmiri the phonemically voiced stops escape the final aspiration to which plain stops are subjected (Vaux and Samuels 2005: 419). The range of these phenomena suggests that final fortition, in addition to the expression in (7b), may be accompanied by the loss of all marked laryngeal content, not just the removal of [voice], as expressed in (7a):

[2] Throughout, we will see complex interaction between release features and laryngeal neutralization. Like Rice (2009: 316), we understand even release features to be phonologically relevant.

(10) *Final laryngeal delinking* (/d tʰ tʼ dʰ . . . / → [t])



The operation in (10) thus accounts for final devoicing, as in Dutch or Polish, but also for final deaspiration, as in Korean; and the effect of this generalized delinking for languages, which combines voicing with aspiration (e.g. Sanskrit) or other properties, is to neutralize all contrasting laryngeal manner types to the plain, voiceless unaspirated type. On this interpretation, then, final neutralization of the most common kind is effected by delinking of the Laryngeal node, with the consequence that the loss of any one of the contrasting laryngeal features in the system implies loss of all the others, too, if any (CHAPTER 27: THE ORGANIZATION OF FEATURES). This appears to be accurate, as final devoicing in more complex systems goes hand in hand with the removal of other contrastive laryngeal gestures as well.

Final fortition, on the other hand, consists in imposition of the feature [spread], as per (7b), but appears to affect only laryngeally unmarked obstruents. Thus, aspiration in Kashaya accrues to final plain (unmarked) stops, but not to final ejectives (marked as [constricted]), and aspiration in Kashmiri similarly accrues to final plain stops, but not to phonemically voiced ones. Final fortition, in sum, affects the class of laryngeally empty obstruent configurations as per the refinement of (7b) given in (11), attracting the feature [spread] to a final Laryngeal node that is otherwise empty, or unmarked. The situation in Klamath or Washo then falls into place as a combination of final laryngeal delinking (10) feeding into final fortition (11).

(11) *Final fortition* (/t d̥/ → [tʰ] (Kashmiri, Klamath, German)

As noted for Kashmiri above, neutralization to [spread] often leaves other, already laryngeally specified series untouched. For instance, Koyukon, Hupa, and Tlingit appear to aspirate unmarked stops finally, neutralizing with the [spread] series, but in each language the ejective series does not participate (Vaux and Samuels 2005: 418–441).

Final neutralization via the addition of other features appears to be far less common, according to the literature of which we are aware. Neutralization to a phonologically voiced member of an opposition may occur, but appears to be very rare, and controversial, as reviewed in recent discussion about the analysis of Lezgian (Yu 2004; Blevins 2006, drawing data from Haspelmath 1993). This Nakh-Daghestanian language possesses a four-way laryngeal distinction in onsets (voiceless, aspirated, voiced, ejective). In some monosyllabic noun classes, ejective and voiceless obstruents neutralize to voiced word-finally or after consonants, with suffixed forms showing the underlying obstruent and unsuffixed stems the neutralized realization (data from Yu 2004):

(12)	pab	pap-a	'wife'
	rug	ruk^{w}-adi	'dust'
	q'eb	q'ep'-ini	'cradle'
	t'ib	t^{h}jp'-er	'owl'

As Blevins (2004, 2006) emphasizes, there is no natural phonetic or aerodynamic reason to make stops voiced in final position, where modal voicing in obstruents is difficult to maintain as well as to perceive; accordingly, the regular occurrence of final voicing, although in principle learnable, is expectedly uncommon. In this instance, the phenomenon arose coincidentally as the product of separate but converging historical events. Specifically, the synchronic pattern in Lezgian is the result of "natural" sound changes involving old voiced obstruents in both medial and final position: the medials underwent gemination (and became voiceless) and subsequently degeminated, leaving the historically voiced word-finals to stand in alternation with now voiceless word-medial stem-finals.

Blevins (2004, 2006) suggests other candidates for synchronic final voicing, perhaps most clearly in Somali, where historically a neutralizing medial voicing process was followed by final vowel loss, thus creating a pattern of apparent final voicing. Kiparsky (2006: 225) reanalyzes Somali (and some other cases), however, as an aspiration language in our sense, rather than a voice language. That is, Somali final "voiced" stops are phonetically lenis and unaspirated, contrasting with aspirated stops marked by [spread]. This view appears to be consistent with available descriptions of Somali obstruent phonetics, most explicitly that of Orwin (1993), although the possibility remains open that some dialects of the language may employ [voice] rather than [spread].

Utterance-final position provides a potential universal starting point or trigger for anticipatory devoicing (to the following silence), whereas no such starting point exists to trigger utterance-final voicing. Phonologically, as a non-assimilatory addition of the feature [voice], final voicing is poorly motivated to begin with inasmuch as feature additions are almost always sourced in assimilation. On the other hand, in the structurally parallel case of final fortition as described above, non-assimilatory addition of [spread glottis] appears to serve a prosodic edge-marking function in association with the greater acoustic salience of fortis/aspirated over lenis/voiced consonants. This function is not compatible with final voicing, and would appear to be irrelevant, too, to the other laryngeal feature, which in principle could be added finally, [constricted glottis].

Thus, neutralization *per se* to glottalic consonants is not securely attested in the literature, although superimposition of a glottal stop on final voiceless oral stops is seemingly common. In the terminology of Michaud (2004: 120), three different kinds of glottal gestures (glottal stop, glottal constriction, and creaky voice/laryngealization):

> can be characterized phonetically as follows: (1) *Glottal stop* is a gesture of closure that has limited coarticulatory effects on the voice quality of the surrounding segments. (2) *Glottal constriction* (also referred to here as *glottal interrupt*) is a tense gesture of adduction of the vocal folds that extends over the whole of a syllable rhyme. (3) *Laryngealization* (i.e. lapse into *creaky voice*), resulting in irregular vocal fold vibration, is not tense in itself. *Glottalization* is used as a cover term for laryngealization and glottal constriction . . . .

Final neutralization directly via any of these three gestures is not known to us, nor are we familiar with final glottalization in the sense of Michaud's (2) and (3). At the same time, a familiar allophonic and optional instance of (1) is found in many varieties of English, e.g. in *bat* [b̥æt̚ˀ], perhaps as a prosodic right-edge marker appearing on fortis stops. This pattern of "glottal reinforcement" then sets the scene for the loss of supralaryngeal stops in some dialects, especially with /t/, e.g. *bat* [b̥æʔ], *bottle* [b̥aʔɫ̩].

Under the Dimensional Theory of feature representation advanced by Avery and Idsardi (2001, forthcoming), in fact, the gesture [constricted glottis] (which characterizes glottal closure, among other phenomena) is implemented via the dimension of Glottal Width, as a complement to the contrary gesture [spread glottis]. On this approach to feature organization, a relationship between these two contradictory gestures is thus predicted, so that an aspiration language such as English or German (with contrastive [spread glottis]) naturally gravitates toward implementation of its Glottal Width dimension in final position as either aspirated, when released (as in *bat* [b̥ætʰ]), or as glottalically closed, when unreleased (as in *bat* [b̥ætˀ]). As noted, many American speakers realize such codas simply as a glottal stop, with no supralaryngeal occlusion.

A striking illustration of the complementary relationship between aspiration and glottalization is found in McFarland's (2007) description of Filomeno Mata Totonac: "Glottal consonants [h] or [ʔ], or spread/constricted glottis features are required at certain domain edges in Filomeno Mata Totonac, and are disallowed domain-internally." The process is thus not neutralizing, as the language makes no laryngeal contrasts, but the differing manifestation of Glottal Width in domain-edge consonants is entirely predictable: glottalization in sonorants, aspiration in obstruents.

Thai – with its three-way laryngeal contrast of voiceless, voiced, and aspirated – shows a pattern of glottalization similar to English, albeit with a difference in frequency. As Esling *et al.* (2005: 388) describe it: "In English, unreleased final glottal reinforced oral stops [ˀp], [ˀt], and [ˀk] are optional allophonic variants, but in Thai they are the norm." Michaud (2004) reports that dialects of Chinese that retain Sino-Tibetan historical final stops (Fujian and Cantonese) typically accompany these with a glottal stop, too, and this co-articulation is taken as a step on the historical path toward loss of final oral stops altogether, first via debuccalization leaving only the glottal stop, as in modern Burmese and Min and Hakka Chinese, then loss of all trace (save tone) of the original stops, as in modern Mandarin.

Central to accounting for the attested patterns of featural addition in neutralization is doubtless the often-noted (e.g. Blevins 2004: 98–99, 2006: 138) tendency of languages not to require or allow final release of stops, thus partially or entirely obscuring distinctions carried by release features. One superficially exceptional-looking pattern underscores the role of release in such neutralizations: in Chong

(Mon-Khmer; Silverman 2006: 79–80), final stops are unreleased, but the language nonetheless maintains a distinction between root-final (= word-final, since the language has no suffixes) glottalized and non-glottalized stops. This is suggested by Silverman to correlate with the timing of glottalization: Chong has pre- rather than post-glottalization. That is, because of the timing of the gesture, the distinction between phonologically glottalized and non-glottalized stops is not dependent on final release.[3]

The literature contains a number of other similar instances from the languages of the world, where some set of final obstruents adds glottalization, without leading to neutralization; see Gurevich (2004: 137ff., 151ff.) on optional glottalization of final voiced stops in Lahaul (Pattani) and glottalization of voiceless stops before consonants or juncture in Maidu. In all these cases, glottal stop is present in the language, but the languages lack contrastive glottalization in the obstruent system. Similarly, the phoneme inventory of Dumi (Tibeto-Burman; van Driem 1993: 52–59) includes both /h/ and /ʔ/, the former restricted to onsets and the latter to codas. The language has three stop series: voiceless unaspirated, voiced, and breathy or murmured. The voiceless series appears finally unreleased with "simultaneous glottal stop," e.g. /leŋghok/ [lɛŋgɔʔk̚] 'throat'. The other two series do not appear in codas in native words, so that the effect of this glottalization is not neutralizing either. The related language Limbu (van Driem 1985: 7–16) shows similar patterns of glottalization of voiceless stops in codas without neutralization.

Debated as a possible case of final voicing (Blevins 2006; Kiparsky 2006), Tundra Nenets reinforces all consonants prepausally with a glottal stop (Salminen 1997: 31–32; see also Janhunen 1986: 81–83), which appears variably in word-internal codas as well.

Other less secure cases are closely parallel, like the glottalization reported for all three voiced stops (/b d g/) in syllable codas in Kamassian (an extinct South Samoyedic language), from Kümmel (2007: 187–188):

(13) b → ʔb ~ ʔ / __ #

As in all other languages we know to have final glottalization, this reinforcement is not contrastive in Kamassian. Like Chong, this involves pre- rather than post-glottalization and, like English, it appears to be connected with facultative loss of final stops for some speakers, as noted above for forms like [b̥æʔ] *bat*. Kümmel in fact suggests that loss of the original coda consonant may have triggered the glottalization. On the other hand, Barnes (2002: 210ff.) follows Hyman (1988) in arguing that "at least in many cases the epenthetic final glottal stop so common in the languages of the world is ultimately the phonologization of allophonic phrase-final creak." In any case, the occurrence of word- or phrase-final glottalic gestures does not itself lead directly to obstruent neutralization, as far as we can tell, but rather does so only indirectly through concomitant loss of oral gestures.

The full possible typology of final laryngeal neutralization, then, includes the patterns shown in (14).

[3] If correct, note that this suggests how phonetics can shape phonology: the gestural and timing patterns used to realize a feature (see Henton *et al.* 1992) appear to correlate with what does or does not happen phonologically, but see also Howe and Pulleyblank (2001).

(14) *Typology of possible final laryngeal neutralizations*

a. deletion of [voice] (Polish, Dutch, Catalan)
b. deletion of [spread] (Korean)
c. deletion of [constricted] (Hup, as discussed below)
d. deletion of [voice] and [spread] (Sanskrit, Burushaski)
e. deletion of [voice] and [constricted] (no clear cases)
f. insertion of [voice] (probably Lezgian, possibly Somali dialects)
g. insertion of [spread] (Kashmiri, Eastern Armenian, Kashaya, German)
h. insertion of [constricted] (final glottalization; no clear neutralizing cases)
i. insertion of [voice] and [spread] (final murmuring; no clear cases)
j. insertion of [voice] and [constricted] (final laryngealization; no clear neutralizing cases)

Of these types, deletion of [voice] (a) and insertion of [spread] (g) are widely and securely attested.[4] While Kiparsky (2008: 46) asserts that coda neutralizations in general "typically" go to unmarked values, he concludes that "the direction of voicing neutralization [is] universal" (2008: 53). As we have seen, however, at least in the particular instance of laryngeal neutralizations, the strong interpretation of this claim is simply false. Indeed, it is unclear to us whether (a) is even significantly more common among voice languages than (g) is among aspiration languages.

Neutralization via the deletion of other contrasting laryngeal features is exemplified in the deaspiration of (b) and the combination of deaspiration with devoicing in (d). Deglottalization via the deletion of [constricted], on the other hand, appears to be rare, but it is found in Hup (see below; on the view that laryngeal features are privative, thus ruling out the insertion of [−voice] or [−spread], see also CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION). And the devoicing in (a) need not necessarily lead to the loss of [constricted]: in Dhaasanac (Tosco 2001: 19–20; Blevins 2006: 143), for instance, word-final obstruents devoice but do not lose contrastive [constricted], so that the implosive series surfaces in final position as voiceless glottalized stops, [ˀp ˀt ˀk]. Fallon (2002) further shows that [constricted] often functions independently of other laryngeal features in other settings.

While the insertion specifically of [voice] (f) is perhaps only marginally instantiated, the deletion of [spread] (b) appears to be tied to the more general removal of all marked laryngeal features. Rather commonplace, however, are the opposites of these: the deletion of [voice] and the insertion of [spread]. Like deletion of [constricted] (c), insertion of contrastive [constricted] (h), (i) is not securely attested, although the literature on this question is more limited to date. Notable is that

[4] Indeed, the association between [voice] in our sense and devoicing is strong enough that van Rooy and Wissing (2001: 326) raise the possibility that phonemic [voice] might imply final devoicing. For [voice] languages that do not appear to have devoicing, like French and Ukrainian, they write that "the phonetics and phonology of these languages need careful investigation, as it appears from this paper that much of what appears to be variable from a phonological perspective can be explained from a phonetic perspective." While final devoicing in [voice] languages may be underreported, the widespread tendency for devoicing could also be the result of a strong historical bias toward deletion of [voice], which would find support in a variety of phonetic factors, many discussed below.

non-contrastive glottalization via insertion of [constricted], or laryngealization via insertion of simultaneous [constricted & voice], is a frequent optional process across a set of languages, whereas final murmuring via insertion of [voice & spread] is not known to us either as a neutralization or as an allophonic enhancement.

In complex laryngeal systems, with three-way or more contrasts, it appears often that not all series neutralize in a merger context. In the Amazonian language Hup, the glottalized series of stops (variably but non-contrastively voiced elsewhere) merges in final position with the plain voiceless series, although the neutralization is perhaps not always complete in nasal contexts (Epps 2008, personal communication). In the same environment, however, the language's phonemically voiced plain stops do not devoice (and so do not merge with voiceless and deglottalized finals), but rather post-nasalize, making them phonetically even more distinct from the other series. (Obstruent post-nasalization is an enhancement suggestive of "hypervoicing" in the sense of Henton *et al.* 1992.)

In summary, then, (a), (b), (d), and (g) are relatively well-attested neutralizations; (c), (e), and (f) seem reasonable, but are not widely attested; (h) is marginal (but learnable); and (i) and (j) appear to be unattested and are perhaps impossible, as we know of no solid cases of neutralization to compound features. Below, in §4, we consider the diachronic paths that may have given rise to some of these asymmetries.

3 How the phonetics of final neutralization informs the phonology

Phonetically, final devoicing in particular has invited appeals to the aerodynamics of speech, according to which utterance or breath group-final edges are produced with reduced pulmonary pressure; others have attributed final devoicing to a phonological assimilation to the following silence (cf. Hock 1999 on both points). And for stops, in particular, it has been argued that modal voicing is difficult to maintain in general; see Gamkrelidze (1975), although Westbury and Keating (1986) advise that solid data on this point are limited.

While neutralization is typically categorical phonologically, the phonetic cues to laryngeal distinctions, including those in final position, have proven to be remarkably complex. With respect to the ostensibly simple issue of English "voicing," Lisker (1986) alone catalogues 16 distinct cues for stops in medial position, including glottal pulsing, consonant and vowel duration, and changes in fundamental frequency and in the first formant. Lisker notes, however, that the list is hardly exhaustive, and more recent work shows that still other factors, including amplitude, also play a role. The function of these cues varies of course by prosodic context, and while stressed word-initial position may be captured relatively straightforwardly by measurement of Voice Onset Time delay, final position proves particularly elusive. In fact, Rodgers *et al.* (2010) show that for a set of speakers from the American Upper Midwest, a different range of acoustic characteristics provides the best correlation for distinguishing word-final /t/ from /d/ in both frame sentences and running speech. These include RMS amplitude, rate of change of RMS amplitude, and, even more specifically, amplitude of individual harmonics and formants. In short, the full acoustic picture of final laryngeal distinctions is far from clear, even for a well-studied language like English.

We lay out in the next paragraphs some implications that this phonetic complexity carries for the phonology of final laryngeal neutralization. First, we argue that a successful account must pay attention to the role of multiple cues and trading relations in producing distinctions (§3.1). Second, we review briefly the possibility that final laryngeal distinctions undergo at most "incomplete neutralization," concluding that complete phonological neutralization is attested (§3.2). This leads to a third point reaching beyond phonetics, namely the range of other effects that come into play in neutralization (§3.3).

3.1 *Multiple cues*

While an extensive body of research shows that final neutralization can be complete (see below for references), the presence of multiple possible and actual phonetic cues to laryngeal distinctions raises questions about the mapping between the phonetics of such cues and phonological contrasts. (For one view on the topic, see Kingston *et al.* 2008.) Evidence from variation in American English points to the importance of "trading relations" among cues. Purnell *et al.* (2005a, 2005b) show that, over several generations of real- and apparent-time speech from eastern Wisconsin, speakers have systematically changed how they realize these contrasts finally, from exploiting actual glottal pulsing early on to later relying on duration of the preceding vowel.[5] A perception test showed that listeners who were speakers of other varieties of American English did not, in general, have difficulty interpreting either set of cues to laryngeal distinctions. That is, the phonetic realization of the distinction has changed over time in this region, but the phonological distinction, even for outsiders, has remained stable.

3.2 *Incomplete neutralization in production*

Although many cases of final laryngeal neutralization have been described as effecting complete merger, a long thread of work has argued that neutralization is sometimes not truly complete, often in the service of challenging the notion of phonological contrasts (e.g. Dinnsen and Charles-Luce 1984; Port and O'Dell 1985; Charles-Luce and Dinnsen 1987). Fourakis and Iverson (1984) attribute the incompleteness effect reported for German to a laboratory artifact introduced by awareness on the part of participants (who presumably also spoke English, a language with a final distinction) as to the purpose of the experimentation, thus resulting in partially hypercorrect pronunciations of finals with traces of their morphophonemically lenis properties rather than as fully neutralized fortis obstruents. But under conditions in which the real purpose of the experiment is concealed in the guise of a morphological exercise (strong verb conjugation) rather than presented as an (apparently intimidating) evaluation of pronunciation, participants neutralized German final obstruents completely. Jassem and Richter (1989) found that final neutralization under conditions such as these is also complete in Polish – *contra* Slowiaczek and Dinnsen (1985) – and Kim and Jongman (1996) report complete neutralization of final manner contrasts in Korean.

[5] In contrast to the productions of earlier generations, the youngest group of speakers appears to be developing final laryngeal neutralization. Evidence of devoicing in northwestern Indiana can be found in José (2009), with reference to a range of other studies. The pattern appears to be stable over time in that community.

3.3 *Incomplete neutralization in perception*

Studies seeking to establish perceptually incomplete neutralization ultimately face challenges beyond what listeners glean from the acoustic signal, which is not, it turns out, the only clue that listeners have to whether even nonce forms belong to one or another class (CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). Ernestus and Baayen (2003: 6), for instance, hypothesize that "Speakers recognize that there is neutralization and base their choice for the underlying representation on the distribution of the underlying representations among existing morphemes, serving as exemplars." They show that the Dutch lexicon has clear asymmetries with regard to the phonotactics of final underlying voice (cf. also Ernestus and Baayen 2007 and other papers in van de Weijer and van der Torre 2007). The distribution of word-final labial stops in Dutch, for example, is heavily biased toward underlyingly voiceless, but labial fricatives are even more heavily biased toward voiced.[6] A production experiment showed that speakers treated novel forms in line with those patterns, leading Ernestus and Baayen (2003: 31) to conclude:

> First, our data show that the underlying [voice] specification of final obstruents in Dutch is predictable to a far greater extent than has generally been assumed. It is predictable not only for linguists having computerized statistical techniques at their disposal, but also for naive speakers, since they use this predictability in language production. Second, we see that the predictability is based on the similarity structure in the lexicon.

Moreover, it has proven challenging for even well-designed experiments to control for issues like orthography. Warner *et al.* (2004, 2006) report a set of very low-level acoustic differences between underlying voiced and voiceless obstruents in Dutch, but then later follow up with new experiments showing that "incomplete neutralization may be entirely caused by orthographic differences" (Warner *et al.* 2006: 292).[7] Indeed, Fourakis and Iverson (1984) noted that previous laboratory investigations of German final neutralization had found acoustic traces of lenis articulation in the orthographic *d* of *und* 'and' and the *g* of *weg* 'away'. Both of these forms represent non-alternating and therefore uncontroversially phonemically fortis stops in modern German that happen only for historical reasons to be spelled with the lenis graphemes *d* and *g* rather than fortis *t* and *k*, respectively.

Aside from extra-phonological influences such as these, studies of ordinary speech, including present-day Wisconsin English and many cases cited by Blevins and others, show that early stages in the historical development of final laryngeal neutralization are prone to being both variable and partial, parallel in some ways to patterns familiar from vocalic "near-mergers" (Di Paolo 1988; Labov 1994). Moreover, even complete neutralizations are often recoverable perceptually from the pragmatic context (morphological differences aside, German *Rat* 'advice' and

[6] This is perhaps surprising, given a heavy cross-linguistic bias for languages to have voiceless fricatives and no voiced ones (Maddieson 1984: 52ff.). That is, frequency in cross-linguistic inventories patterns quite differently from the lexical patterns found within Dutch.

[7] For a fuller range of evidence and discussion of the complex devoicing and other laryngeal phonology of Dutch, see van de Weijer and van der Torre (2007).

Rad 'wheel' are unlikely to be confusable even if they occur in the same discussion) or based on phonological gaps in lexical distribution. Thus, German morphophonemically fortis stops freely occur after lax vowels (*Ecke* 'corner', *Beck* 'brook'), but lenis stops rarely do (*Ebbe* 'ebb tide', *gib* 'give! (SG IMP)'),[8] and there are numerous underlying sequences of stem-final labial or velar lax stop preceded by tense vowel (*Lieb-* 'love'), but hardly any with underlying fortis stop in this context. These kinds of patterns allow Piroth and Janker (2004: 99–100) to observe that, "Due to lexical and morphological structure there are only very few minimal pairs of alternating paradigms with underlying voiceless vs. underlying voiced final obstruents." That is, the phonological contrast differentiates few homophones, so that there is little lexical competition in the sense of Blevins and Wedel (2009: esp. 169).

3.4 Summary

This section has surveyed some issues in the phonetics of final laryngeal distinctions, drawing especially on English, which maintains a distinction in almost all varieties and most contexts, and German, in which most varieties do not. First, we have argued here that phonetic cues to final laryngeal distinctions show remarkable complexity, and can change even as phonological contrasts remain stable in perception and production. Second, this informs the question of whether final neutralization is always complete or can be incomplete. At least when realized by feature deletion, it appears that complete phonological neutralization in ordinary speech is observable both acoustically and perceptually, but recovery of the neutralization is enhanced by considerations of pragmatics, skewed phonotactic distributions, spelling conventions, and lexical limitations on homophony.

4 The domains of final laryngeal neutralization and paths of change

Whether accomplished via feature removal or addition, final laryngeal neutralization is widely attested at all levels of the prosodic hierarchy, from utterance-final to phrase-final to word-final to syllable-coda position. In fact, the major route by which languages develop final laryngeal neutralization has been seen as running along that hierarchy, beginning with large units and moving downwards (e.g. Hock 1999; Blevins 2004, 2006).

For feature-removing neutralization, universal physical and acoustic motivations make voicing particularly challenging at the end of breath groups. By definition, at the end of breath groups pulmonary pressure reaches its minimum for that stretch of speech, but that can also be controlled and enhanced as part of speech. Lieberman (1967: 104) found that American English speakers lower subglottal air pressure during the last 150–200 msecs of sentences, with acoustic effects including falling fundamental frequency. Decreased pressure exacerbates the inherent difficulties of voicing stops in particular, with further biases by place of articulation, as noted above.

[8] In fact, speakers vary in having a tense and lax vowel in *gib*, and the standard reference work on standard pronunciation gives [giːp] for this form (Mangold 2005).

While neutralization at the end of longer stretches of speech is directly rooted in the physiology and physics of speech (pulmonary pressure and the aerodynamics of vocal fold vibration), movement down the prosodic hierarchy involves steadily broadening generalizations made by learners and speakers over generations, as argued by Iverson and Salmons (2009), Salmons (2010), and many others. Blevins (2006: 140–143) presents broad evidence from a wide range of families on historical paths of development of final laryngeal neutralization. "Early stages" of devoicing co-occur with prepausal or phrase-final position, and they may be variable, gradient, and sensitive to aerodynamic properties (like a preference for devoicing /g/), in addition to occurring only at the right edges of phrases (or presumably other longer stretches of speech). She posits an implicational hierarchy of such patterns, according to which languages may neutralize at the right edges of larger prosodic units and not at smaller ones, but never vice versa. For instance, numerous languages (Dhaasanac, Maltese, and some varieties of German) neutralize at the ends of words (and larger units), but not of syllables.[9] Of course the reverse pattern would be inherently odd structurally: word-final is coda position and phrase-final is word-final, for instance, so that coda neutralization should apply to higher levels. As observed above, the historical development of glottalization is reported to parallel this path over prosodic domains closely, as discussed above, but remains laryngeally non-neutralizing and often leads to the loss of final oral stops.

Overall, these patterns reflect the historical paths of development laid out in the theory of Evolutionary Phonology (Blevins 2004, 2006): laryngeal neutralization by means of feature addition, especially if accompanied by mandatory release, appears to function as an edge marker, arising via release of final stops in salient prosodic positions, whereas neutralization via feature loss appears driven by mandatory or facultative absence of release. Even if the motivation is different, edge marking shows distributions similar to those of neutralization by feature removal. For example, in the Tundra Nenets case discussed above, glottal reinforcement occurs with all consonants prepausally.

Some varieties of American English may be starting down this path today. As already noted above, young speakers of Upper Midwestern English show nascent word-final devoicing. Purnell *et al.* (2009) present evidence that may reflect an earlier stage of this process: 2008 Republican vice-presidential candidate Sarah Palin showed variable final neutralization, with a preference for phrase-final position; i.e. what looks like a slightly earlier stage of development than that found now in Wisconsin. Palin was raised in an Alaskan community settled overwhelmingly by 1930s emigrants from northern Wisconsin, Michigan, and Minnesota, and this colonial variety of Upper Midwestern English may preserve the earlier patterns.

9 Blevins's treatment does not involve detailed language histories, but see Mihm (2004) and Iverson and Salmons (2007) for the beginnings of a case study of German.

5 History and Universal Grammar in final laryngeal neutralization

In the preceding sections, we laid out two contrasting positions with regard to how to explain final laryngeal neutralization. Following the discussion in Kiparsky (2008), one position favors the view that grammatical structure constrains language change and the other favors the view that language change is the primary shaper of grammar. The former is the classic generative position, illustrated on this issue by the work of Kiparsky, and the latter is associated with various approaches, mostly recently Evolutionary Phonology, illustrated by the work of Blevins, both discussed above. Kiparsky (2008: 52) concludes that:

> The two programs can coexist without contradiction or circularity as long as we can make a principled separation between true universals, which constrain both synchronic grammars and language change, and typological generalizations, which are simply the results of typical paths of change.

We share that ecumenical spirit, and note that it can be difficult to find the seam between such true universals and typological generalizations. We have argued above that Kiparsky has proposed a "true universal" which does not ultimately hold up empirically. This suggests that Universal Grammar is leaner than has often been claimed, in line with many views emerging in the field today. The preceding sections aim to develop generalizations, some of which may prove to be "true universals," while others will clearly be "typological generalizations." In contrast to Kiparsky's approach, Blevins (2008: 107) concludes that "Within the phonological realm, there appear to be few, if any, substantive universals," with specific rejection of distinctive features as "substantive phonological universals," treating them instead as "emergent properties" (see also CHAPTER 17: DISTINCTIVE FEATURES). Full justification would go well beyond our assigned task here, but we are not prepared to abandon the core substance of phonology. Above, we have relied on abstract featural characterizations, but argued that they must be considered in the context of phonetic variability and an array of psycholinguistic factors. History, internal and external, shapes contrasts and features through that context.

6 Summary and conclusions

In the foregoing, we started from a synthetic survey of what is known at present about final devoicing and laryngeal neutralization generally. Along the way, we have identified a number of phonological patterns based on the currently available evidence. Assuming privative laryngeal contrasts, neutralization can occur toward either a marked or an unmarked feature configuration, that is, by feature removal or insertion. Another form of right-edge marking, glottalization, appears to occur in languages for which [constricted] is not contrastive, thus without triggering laryngeal neutralization. These points all bear on current phonological discussions, i.e. about featural representations, the role of prosody and history in synchronic phonology, and the nature of neutralization in "weak" positions. All of these discussions represent areas of potential progress.

While we have kept a focus on phonological theory, we have done so in the context of phonetics (especially perception), sound change, and prosody, all of which, we argue, are critical to a full understanding of final neutralization. Several conclusions follow, including these:

(i) *Featural characterization matters*: The deletion of [voice], or final devoicing *sensu stricto*, is pervasive although not ubiquitous cross-linguistically. Addition of [voice] is rare to the point that some doubt its existence. The addition of [spread], or final fortition, is well attested, if potentially less common than devoicing. Deletion of [spread] (often with other features) is relatively common. Laryngeal realism, we have suggested, provides a robust typological generalization that would be impossible on the traditional interpretation of phonemically lenis stops with [+voice], and of phonemically aspirated stops with [−voice].

(ii) *Phonetics, phonotactics, and other patterns matter*: Laryngeal distinctions are carried by a wide range of phonetic cues, even within a single variety or even speaker. Still, long-standing efforts to argue for "incomplete neutralization" in German or Dutch may reflect nothing about the acoustic signal, but much about the generalizations speakers/listeners are able to make based on their knowledge of phonotactics and lexical frequency. This is especially important because the phonetics of final laryngeal distinctions is particularly complex, a fact that has left open the possibility that some cue to the distinction may survive neutralization. Phonologically, however, evidence indicates that final neutralization in languages like Dutch and German is typically complete.

(iii) *Both structure and history matter*: Both the "design" of language and historical forces play significant roles in synchronic sound patterns. In the particular example at hand, some claims about the role of language design have proven to be overstated, but such refinements are the work of healthy science.

As noted in §2, numerous scholars have treated "final devoicing" and its relatives as particular forms of final weakening, synchronic or diachronic. Research to date on final laryngeal neutralization reveals widespread patterns of feature addition and fortition. As with laryngeal realism itself, we would suggest that final neutralizations can come about by either weakening – through feature loss – or strengthening – through feature addition, including glottal reinforcement where it is not contrastive in any of the languages surveyed.

This presentation raises a number of new questions, especially on the typological front. The data we are aware of suggest some striking and as yet unexplored patterns, like the tendency of [spread] to delete finally together with other laryngeal features (Sanskrit), while [voice] appears able to delete independently (Dhaasanac). The answers to questions about such patterns will doubtless sharpen our understanding of the subtle interactions between the human linguistic endowment and the historical patterns we are presented with as learners.

ACKNOWLEDGMENTS

We thank the editors for the invitation to contribute to this project and an anonymous reviewer for helpful comments. Pattie Epps, Beth Hume, José Hualde, Sadaf Munshi, Marc van Oostendorp, Tom Purnell, Eric Raimy, Blake Rodgers, and Alan Yu all provided helpful discussions on this topic and in some cases comments on earlier versions of this chapter. The usual disclaimers apply.

REFERENCES

Ahn, Sang-Cheol & Gregory K. Iverson. 2004. Dimensions in Korean laryngeal phonology. *Journal of East Asian Linguistics* 13. 345–379.

Avery, Peter & William J. Idsardi. 2001. Laryngeal dimensions, completion and enhancement. In Hall (2001), 41–70.

Avery, Peter & William J. Idsardi. Forthcoming. *Laryngeal phonology*. Cambridge: Cambridge University Press.

Barnes, Jonathan. 2002. Positional neutralization: A phonologization approach to typological patterns. Ph.D. dissertation, University of California, Berkeley.

Becker, Michael, Nihan Ketrez & Andrew Nevins. 2008. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish devoicing neutralization. Unpublished ms., Reed College, Yale University & Harvard University (ROA-1001).

Blevins, Juliette. 1993. Klamath laryngeal phonology. *International Journal of American Linguistics* 59. 237–279.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32. 117–166.

Blevins, Juliette. 2008. Consonant epenthesis: Natural and unnatural histories. In Good (2008), 79–107.

Blevins, Juliette & Andrew Wedel. 2009. Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26. 143–183.

Brockhaus, Wiebke. 1995. *Final devoicing in the phonology of German*. Tübingen: Niemeyer.

Buckley, Eugene. 1994. *Theoretical aspects of Kashaya phonology and morphology*. Stanford: CSLI.

Campbell, Lyle. 1998. *Historical linguistics: An introduction*. Cambridge, MA: MIT Press.

Charles-Luce, Jan & Daniel A. Dinnsen. 1987. A reanalysis of Catalan devoicing. *Journal of Phonetics* 15. 187–190.

Di Paolo, Marianne. 1988. Pronunciation and categorization in sound change. In Kathleen Ferrara, Becky Brown, Keith Walters & John Baugh (eds.) *Linguistic change and contact: Proceedings of the 16th Annual Conference on New Ways of Analyzing Variation in Language*, 84–92. Austin: University of Texas.

Dinnsen, Daniel A. & Jan Charles-Luce. 1984. Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12. 49–60.

Driem, George van. 1985. *A grammar of Limbu*. Berlin: Mouton de Gruyter.

Driem, George van. 1993. *A grammar of Dumi*. Berlin: Mouton de Gruyter.

Epps, Patience. 2008. *A grammar of Hup*. Berlin & New York: Mouton de Gruyter.

Ernestus, Mirjam & R. Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79. 5–38.

Ernestus, Mirjam & R. Harald Baayen. 2007. Intraparadigmatic effects on the perception of voice. In van de Weijer & van der Torre (2007), 153–172.

Esling, John H., Katherine E. Fraser & Jimmy G. Harris. 2005. Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka). *Journal of Phonetics* 33. 383–410.

Fallon, Paul D. 2002. *The synchronic and diachronic phonology of ejectives*. New York: Routledge.

Feizollahi, Zhaleh. 2010. Does Turkish implement a two-way voicing contrast as prevoiced vs. voiceless aspirated? Paper presented at the 84th Annual Meeting of the Linguistic Society of America, Baltimore.

Fourakis, Marios & Gregory K. Iverson. 1984. On the "incomplete neutralization" of German final obstruents. *Phonetica* 41. 140–149.
Gamkrelidze, Thomas V. 1975. On the correlation of stops and fricatives in a phonological system. *Lingua* 35. 231–361.
Good, Jeff (ed.) 2008. *Linguistic universals and language change*. Oxford: Oxford University Press.
Gurevich, Naomi. 2004. *Lenition and contrast: The functional consequences of certain phonetically conditioned sound changes*. New York & London: Routledge.
Hall, T. A. (ed.) 2001. *Distinctive feature theory*. Berlin & New York: Mouton de Gruyter.
Harris, John. 2009. Why final obstruent devoicing is weakening. In Kuniya Nasukawa & Phillip Backley (ed.) *Strength relations in phonology*, 9–46. Berlin & New York: Mouton de Gruyter.
Haspelmath, Martin. 1993. *A grammar of Lezgian*. Berlin: Mouton de Gruyter.
Henton, Caroline, Peter Ladefoged & Ian Maddieson. 1992. Stops in the world's languages. *Phonetica* 49. 65–101.
Hock, Hans Henrich. 1999. Finality, prosody, and change. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.) *Proceedings of LP '98: Item order in language and speech*, 15–30. Prague: Karolinum Press.
Honeybone, Patrick. 2005. Sharing makes us stronger: Process inhibition and segmental structure. In Philip Carr, Jacques Durand & Colin J. Ewen (ed.) *Headhood, elements, specification and contrastivity*, 167–192. Amsterdam & Philadelphia: John Benjamins.
Honeybone, Patrick. Forthcoming. *Theoretical historical phonology: Lenition, laryngeal realism, and Germanic obstruent shifts*. Oxford: Oxford University Press.
Howe, Darin & Edward Pulleyblank. 2001. Patterns and timing of glottalization. *Phonology* 18. 45–80.
Hualde, José Ignacio. 1992. *Catalan*. London & New York: Routledge.
Hyman, Larry M. 1988. The phonology of final glottal stops. *Proceedings of the Western Conference on Linguistics* 1. 113–130.
Iverson, Gregory K. & Joseph C. Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12. 369–396.
Iverson, Gregory K. & Joseph C. Salmons. 2003. Legacy specification in the laryngeal phonology of Dutch. *Journal of Germanic Linguistics* 15. 1–26.
Iverson, Gregory K. & Joseph C. Salmons. 2006. On the typology of final laryngeal neutralization: Evolutionary Phonology and laryngeal realism. *Theoretical Linguistics* 32. 205–216.
Iverson, Gregory K. & Joseph C. Salmons. 2007. Domains and directionality in the evolution of German final fortition. *Phonology* 24. 121–145.
Iverson, Gregory K. & Joseph C. Salmons. 2009. Naturalness and the lifecycle of sound change. In Patrick Steinkrüger & Manfred Krifka (ed.) *On inflection: In memory of Wolfgang U. Wurzel*, 89–105. Berlin: Mouton de Gruyter.
Janhunen, Juha. 1986. *Glottal stop in Nenets*. Helsinki: Suomalais-Ugrilainen Suera.
Jassem, Wiktor & Lutoslawa Richter. 1989. Neutralization of voicing in Polish obstruents. *Journal of Phonetics* 17. 317–325.
José, Brian D. 2009. Testing the apparent time construct in a young community: Steel City speech in and around Gary, Indiana on its 100th birthday. Ph.D. dissertation, Indiana University.
Kim, Hyunsoon & Allard Jongman. 1996. Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics* 24. 295–312.
Kingston, John, Randy L. Diehl, Cecilia J. Kirk & Wendy A. Castleman. 2008. On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics* 36. 28–54.
Kiparsky, Paul. 2006. The amphichronic program vs. evolutionary phonology. *Theoretical Linguistics* 32. 217–236.

Kiparsky, Paul. 2008. Universals constrain change; change results in typological generalizations. In Good (2008), 23–53.

Kopkallı, Handan. 1993. A phonetic and phonological analysis of final devoicing in Turkish. Ph.D. dissertation, University of Michigan.

Kümmel, Martin Joachim. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert.

Labov, William. 1994. *Principles of linguistic change*, vol. 1: *Internal factors*. Oxford: Blackwell.

Ladefoged, Peter. 1973. The features of the larynx. *Journal of Phonetics* 1. 73–84.

Lieberman, Philip. 1967. *Intonation, perception and language*. Cambridge, MA: MIT Press.

Lisker, Leigh. 1986. "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29. 3–11.

Lombardi, Linda. 2001. Why Place and Voice are different: Constraint-specific alternations in Optimality Theory. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 13–45. Cambridge: Cambridge University Press.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Mangold, Max. 2005. *Duden Aussprachewörterbuch*. 6th edn. Mannheim: Duden.

McFarland, Teresa. 2007. Glottal epenthesis at domain edges in Filomeno Mata Totonac. Paper presented at the International Conference on Totonac-Tepehua Languages, Banff. Available (June 2010) at www.arts.ualberta.ca/~totonaco/ICTTL.html.

Michaud, Alexis. 2004. Final consonants and glottalization: New perspectives from Hanoi Vietnamese. *Phonetica* 61. 119–146.

Mihm, Arend. 2004. Zur Geschichte der Auslautverhärtung und ihrer Erforschung. *Sprachwissenschaft* 29. 133–206.

Munshi, Sadaf. 2006. Jammu and Kashmir Burushaski: Language, language contact, and change. Ph.D. dissertation, University of Texas, Austin.

Nicolae, Andreea & Andrew Nevins. 2009. The phonetics and phonology of fricative non-neutralization in Turkish. Paper presented at the 40th Annual Meeting of the North East Linguistic Society, MIT.

Orwin, Martin. 1993. Phonation in Somali phonology. In Mohamed Abdi (ed.) *Anthropologie Somalienne*, 251–257. Besançon: University of Besançon.

Piroth, Hans Georg & Peter M. Janker. 2004. Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics* 32. 81–109.

Port, Robert F. & Michael O'Dell. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13. 455–471.

Purnell, Thomas C., Eric Raimy & Joseph C. Salmons. 2009. Defining dialect, perceiving dialect and new dialect formation: Sarah Palin's speech. *Journal of English Linguistics* 37. 331–355.

Purnell, Thomas C., Joseph C. Salmons & Dilara Tepeli. 2005a. German substrate effects in Wisconsin English: Evidence for final fortition. *American Speech* 80. 135–164.

Purnell, Thomas C., Joseph C. Salmons, Dilara Tepeli & Jennifer Mercer. 2005b. Structured heterogeneity and change in laryngeal phonetics: Upper Midwestern final obstruents. *Journal of English Linguistics* 33. 307–338.

Rice, Keren. 2009. Nuancing markedness: A place for contrast. In Eric Raimy & Charles Cairns (eds.) *Contemporary views on architecture and representations in phonology*, 311–321. Cambridge, MA: MIT Press.

Rodgers, Blake, Thomas C. Purnell & Joseph C. Salmons. 2010. Harmonic energy at vowel offset as a cue to post-vocalic VOICING contrasts in American English. Unpublished ms., University of Wisconsin, Madison.

Rooy, Bertus van & Daan Wissing. 2001. Distinctive [voice] implies regressive voicing assimilation. In Hall (2001), 295–334.

Rubach, Jerzy. 1997. Extrasyllabic consonants in Polish: Derivational Optimality Theory. In Iggy Roca (ed.) *Derivations and constraints in phonology*, 551–581. Oxford: Clarendon Press.

Salminen, Tapani. 1997. *Tundra Nenets inflection*. Helsinki: Suomalais-Ugrilainen Suera.

Salmons, Joseph C. 2010. Segmental phonological change. In Vit Bubenik & Silvia Luraghi (ed.) *A companion to historical linguistics*, 89–105. London & New York: Continuum.

Silverman, Daniel. 2006. *A critical introduction to phonology: Of sound, mind, and body*. London & New York: Continuum.

Slowiaczek, Louisa M. & Daniel A. Dinnsen. 1985. On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13. 325–341.

Syeed, Syed Mohammad. 1978. The Himalayan way of breathing the last: On the neutralizing status of the word-final aspiration in Kashmiri. *The Eastern Anthropologist* 31. 531–541.

Tieszen, Bozena. 1997. Final stop devoicing in Polish: An acoustic and historical account for incomplete neutralization. Ph.D. dissertation, University of Wisconsin, Madison.

Tosco, Mauro. 2001. *The Dhaasanac language: Grammar, texts, vocabulary of a Cushitic language of Ethiopia*. Cologne: Rüdiger Köppe Verlag.

Trubetzkoy, Nikolai S. 1977. *Grundzüge der Phonologie*. 7th edn. Göttingen: Vandenhoeck & Ruprecht.

Vaux, Bert & Bridget Samuels. 2005. Laryngeal markedness and aspiration. *Phonology* 22. 395–436.

Warner, Natasha, Erin Good, Allard Jongman & Joan A. Sereno. 2006. Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics* 34. 285–293.

Warner, Natasha, Allard Jongman, Joan A. Sereno & Rachèl Kemps. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32. 251–276.

Weijer, Jeroen van de & Erik Jan van der Torre (eds.) 2007. *Voicing in Dutch: (De)voicing – phonology, phonetics, and psycholinguistics*. Amsterdam & Philadelphia: John Benjamins.

Westbury, John R. & Patricia Keating. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22. 145–166.

Yu, Alan C. L. 2004. Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80. 73–97.

Yu, Alan C. L. 2008. The phonetics of quantity alternation in Washo. *Journal of Phonetics* 36. 508–520.

70 Conspiracies

CHARLES W. KISSEBERTH

1 Conspiracies: The essential argument

The paper "On the functional unity of phonological rules" (Kisseberth 1970; henceforth FUPR) made what in essence is a very simple argument. It claimed that in the phonologies of the world's languages, it is often the case that there are phonological structures that are either barred or required, and that (from a standard generative phonology point of view) multiple rules may be involved in guaranteeing that these structures are avoided or achieved. This observation seems to be undeniably accurate. FUPR, however, went further and suggested that it was not sufficient to simply recognize this truth about the world's phonologies, but that somehow (a) these barred/required structures should be an explicit part of the phonological system of a given language, and (b) grammars that utilize multiple means to achieve/avoid a certain structure are not to be viewed as more complex than grammars that use fewer rules for the same end. As we shall discuss below, these claims are not consistent with the prevailing notion in generative phonology that all significant linguistic generalizations are expressible in terms of simplifications in the formal system of rules and representations. FUPR suggested that there was instead a *functional* aspect to phonological rules that eluded the formal approach of early generative phonology. It should be emphasized that the use of the term "functional" in FUPR is distinct from later usage where functional refers to the idea that (a) phonological phenomena are motivated by, i.e. grounded in, phonetic considerations (summarized in Boersma 1997 as the "minimization of articulatory effort and maximization of perceptual contrast"), as well as the idea that (b) notions such as contrast, paradigmatic considerations and frequency may shape the phonological grammar. FUPR emphasized the existence of avoided/ preferred structures, but not what factors make the structure in question good or bad nor what factors favor one repair over another.

The idea that phonological rules "conspire" to avoid/achieve a given phonological structure is one that had been suggested to me originally by Haj Ross and George Lakoff on the basis of their syntactic work (as well as that of David Perlmutter). A close reading of major pre-generative linguists (particularly Boas and Sapir) seemed to support this notion, as these linguists not infrequently linked a particular phonological phenomenon (e.g. so-called "inorganic" vowels) to some claimed limitation on phonological structure. It was, however, Morris Halle who wisely suggested to me that I take a close look at the Yawelmani dialect of Yokuts if I wished to push this line of thought.

Yawelmani has, over the decades, been a point of reference for almost every approach to the essential problems of phonology. Newman (1944) provided the initial detailed description of Yawelmani (as well as three other dialects of the Yokuts language). His description of the language was in the tradition of Sapir. Harris (1944) and Hockett (1967, 1973) looked at the Yawelmani data from the point of view of American structuralism, and Kuroda (1967) reworked Newman's analysis in terms of standard generative phonology. Later, Archangeli (1984 and subsequent work) applied the principles of underspecification and non-linear phonology to Yawelmani. In the optimality-theoretic literature, Yawelmani has been critical to the analysis of opacity (cf. Cole and Kisseberth 1995; McCarthy 2007).

FUPR examined a number of the essential aspects of Yawelmani phonology, but while not differing radically from Kuroda's analysis, drew a conclusion that was out of the mainstream. What follows is a detailed summary of FUPR's account of Yawelmani.

Yawelmani words, in their surface form, consist of a sequence of syllables of the shape CV, CVV, CVC (where **VV** = long vowel). Thus all words begin with a single consonant and they may end either in a vowel or a single consonant. Internal to the word, vowels do not occur in succession: there is at least one and at most two consonants located between vowels. Vowels may be long or short, but when a vowel stands in the environment __ CC or __ C#, it can only be short. These observations lead to the conclusion that Yawelmani bans syllables with complex margins (i.e. consonant clusters in either onset or coda position) as well as trimoraic syllables.

In the timeframe of FUPR, syllables played no role in generative phonology (see CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE). As a consequence, all of the statements about phonological structures and all the rules formulated referred only to sequences of consonants and vowels. The discussion below follows the presentation in FUPR, but also translates it into a syllable-based analysis.

Underlying representations in Yawelmani are shaped in part by the above limitations on surface structure (CHAPTER 1: UNDERLYING REPRESENTATIONS). There are no prefixes in Yawelmani, so stems always occupy word-initial position. All stems in Yawelmani have an initial consonant. Thus there is no need to have a rule to insert a consonant at the beginning of vowel-initial words. On the other hand, while no stem contains a sequence of three consonants, some stems do end in two consonants. When such stems are followed by a consonant-initial suffix, we have a situation where the ban on triconsonantal clusters (or complex margins) is endangered. An epenthetic vowel is introduced between the first two consonants in the three-consonant sequence (CHAPTER 67: VOWEL EPENTHESIS). This vowel epenthesis phenomenon is illustrated by the data in (1). (We have supplemented the data cited in FUPR for purposes of exposition.)

(1)	*Underlying form of stem*	*Stem + aorist* /hin/	
	ˀilk	ˀilik-hin	'sing'
	ˀutj	ˀutuj-hun	'fall'
	logw	logiw-hin	'pulverize'
	ˀajj	ˀajij-hin	'pole a boat'

The stems in (1) appear in their underlying form when followed by a vowel-initial suffix (e.g. [ʔilk-al] 'might sing', [logw-ol] 'might pulverize', [ʔajj-al] 'might pole a boat', etc.), but with an epenthetic [i] when followed by a consonant-initial suffix like the aorist /-hin/. If these stems appear in word-final position (e.g. in the imperative), then a vowel is also epenthesized between the two consonants that stand in word-final position (due to the fact that a word can only end in a single consonant, as a consequence of the ban on complex codas). Thus the stem /ʔilk/ will appear as [ʔilik] if not followed by a suffix. The epenthetic vowel in Yawelmani is the high front unround vowel [i], though this vowel will appear as [u] when preceded by a high round vowel (cf. [ʔutuj-hun]).

In FUPR, the rule in (2) is postulated for Yawelmani:

(2) *Vowel Epenthesis*

Ø → i / C __ C {#, C}

These epenthetic vowels are clearly a means to avoid complex syllable margins. But Vowel Epenthesis is not the sole method for avoiding complex margins in Yawelmani. Specifically, there are two morphologically restricted cases where one of the consonants is deleted. One case involves two suffixes that have an initial consonant cluster (in both cases, the first consonant in the cluster is a laryngeal), where the first of these consonants is deleted in position after a consonant-final stem. For example, the suffix /-hnil-/ elides its initial consonant after a consonant-final stem like /gitiin-/ 'to hold under the arm' (this morpheme sequence occurs in the noun 'armpit' and there is other morphology and irrelevant phonology involved in the final form of this noun, [giten-nel-a-w]). The available data are not sufficient to make it clear whether the two suffixes in question are at all productive. The rule given in (3), if ordered before Vowel Epenthesis, will correctly delete the initial consonant of /-hnil-/.

(3) C → Ø / C + __ C
(in fact, the only consonants that occur in the environment C + __ C are [ʔ h])

The second case of consonant deletion occurs as an aspect of the phonology of certain suffixes which trigger moraic reduction in a preceding stem (in Newman's terminology, these suffixes require the "zero" form of a stem). For instance, stems with three consonants (e.g. /halaal-/ 'lift up') are converted to the shape CVCC- (e.g. [hall-]) in front of a suffix such as /-hatin/. This suffix, in turn, will elide its initial consonant, due to the prohibition against three-consonant sequences (i.e. complex margins). Again, we are not aware of whether a suffix such as /-hatin/ is productive. The rule given in (4), if ordered before Vowel Epenthesis, will correctly delete the initial consonant of the specified suffixes.

(4) C → Ø / CC + __
(affects only suffixes such as /-hatin/ that trigger the so-called "zero stem")

It is clear that both of these restricted deletion rules are functionally related to the vowel epenthesis phenomenon: they guarantee that an input that potentially violates the ban on triliteral consonant clusters and word-final consonant clusters (i.e. complex syllable margins) will in fact obey this constraint on the surface (CHAPTER 68: DELETION). Of course, all the potential violations of the ban on complex margins could be avoided by the single phenomenon of Vowel Epenthesis. The consonant deletion processes in (3) and (4) are not necessary in order to secure an outcome where there are no complex margins. However, what the existence of conspiracies tells us is that (in a rule-based approach to phonology) languages do not always opt to use the fewest number of rules possible to avoid an offending phonological structure.

Rules (2)–(4) constitute only one part of the Yawelmani conspiracy against complex margins. There is evidence to motivate the postulation of a rule in Yawelmani that elides a short vowel (either /i/ or /a/) in the environment VC __ CV. This rule can explain both cases where vowels in verb suffixes are elided, and also aspects of the nominal case system. Specifically, it deletes what Newman referred to as the "protective" vowel /a/ in a structure like /k'iliij + a + ni/ 'cloud (INDIRECT OBJECTIVE)' but not in a structure like /puulm + a + ni/ 'husband (INDIRECT OBJECTIVE)'. In an example like /k'iliij + a + ni/, deletion of /a/ yields an output where there is only one consonant in onset position and one consonant in coda position. On the other hand, there is no deletion of /a/ in /puulm + a + ni/, due to the fact that such a deletion would yield an output with three consonants in a row: [lmn], a sequence that would require either a complex onset or a complex coda.

The rule in (5) achieves the correct result.

(5) $V \rightarrow \emptyset$ / VC __ CV

Rule (5) does not bear any *formal* similarity to either the vowel epenthesis rule in (2) or the consonant deletion rules in (3) and (4). However, despite this lack of formal similarity, there is an obvious functional similarity: (5) deletes a vowel *unless* to do so would create violations of the ban on complex margins. Deleting a vowel is of course the opposite of inserting one from a formal point of view, but both actions, along with consonant deletion, reveal the overarching principle that complex margins are not permitted in Yawelmani.

There is yet one more aspect to the Yawelmani complex margin conspiracy. Word-final verbal suffixes of the shape -CV elide their vowel when preceded by a vowel-final stem but not a consonant-final stem. In other words, the vowel of these suffixes will elide unless its elision would produce a violation of the ban on complex margins. Thus the imperative suffix /-k'a/ loses its vowel after the stem /taxaa-/ 'take', yielding [taxa-k'], but no elision takes place after the stem /xat-/ 'eat', yielding [xat-k'a] rather than *[xat-k'], with a complex coda. This second vowel deletion rule is formulated in (6):

(6) $V \rightarrow \emptyset$ / V + C __ #

There is of course some formal similarity between (5) and (6), since both are vowel deletion rules and only apply in the event a VC structure precedes. However, since (6) is restricted to word-final vowels that are part of a CV suffix, there is not identity even in terms of the preceding structure: i.e. VC __ in (5) but V + C in (6).

We have now discussed the *rules* in Yawelmani that implement the ban on complex margins: (2)–(6). These rules, however, do not tell the entire story. In the standard generative phonology that prevailed at the time FUPR was written, it was proposed that *morpheme structure conditions* (CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS) served to restrict the shapes of morphemes in underlying representation. In Yawelmani, there are no morphemes that contain a sequence of three successive consonants. In other words, structures are avoided inside a morpheme if they could only be syllabified by creating a complex margin.

This aspect of conspiracies was developed at greater length in Kenstowicz and Kisseberth (1977), under the rubric of the "duplication problem." Specifically, it was noted that in theories where there are both "morpheme structure conditions" and also phonological rules, it is often necessary to repeat the same generalization both as a morpheme structure condition and also as a feature-changing rule. For example, underlying representations of morphemes may disallow a NC sequence where the nasal and the consonant are not of the same point of articulation, but a rule may still be required that converts a morpheme-final nasal to be homorganic with a consonant in initial position in the next morpheme (CHAPTER 81: LOCAL ASSIMILATION). As long as there is a morpheme structure component of the phonology that is distinct from the rules that account for alternations in the shapes of morphemes, then there will be conspiracies whereby structures that are banned in the underlying representations of morphemes will trigger morphophonemic alternations as well.

To summarize: in Yawelmani, a morpheme structure ban on trisyllabic consonant sequences prevents morphemes from having offensive segmental material to begin with. A rule of vowel epenthesis, as well as two minor rules of consonant deletion, prevents violations from occurring at the juncture of morphemes. Two vowel deletion phenomena are constrained in a fashion to prevent the creation of complex margins.

But having pointed out this conspiracy, what are we to make of it? FUPR suggested that although there is no way in which a *formal* unity can be found for these various rules, nevertheless the ban on triliteral and word-final consonant clusters (i.e. the ban on complex margins) should be part of the phonological grammar of Yawelmani. However, in standard generative phonology, a phonology is a set of representations and a set of ordered rules that derive a surface form from an input form. A ban on complex margins cannot be part of the phonology unless it participates in the derivation of surface forms from input forms; otherwise it is simply a useless appendage that has no basis for existence in a formal system. In an attempt to find some way to make the ban on complex margins a part of the phonology, FUPR suggested that bans of this sort might function as *derivational constraints*.

The idea of a derivational constraint is this. Suppose that we formulate rule (5) as (7):

(7) A word-medial (short) vowel elides.

Say that the application of this rule fails if its immediate output would violate the ban on triliteral or word-final consonant clusters (i.e. complex margins). This derivational constraint would allow a word-medial vowel to delete only if deletion does not produce an illicit structure. Of course, this notion of derivational constraint would radically alter the way in which rules apply, but it would mean that a constraint like the ban on complex margins in Yawelmani would have an actual role to play in derivations.

The notion of derivational constraints, however, is only a very partial account of the Yawelmani complex margin conspiracy. It is not evident how the ban on complex margins would play any role in the derivation of words where offending structures arise across morpheme boundaries. Because the notion of a derivational constraint did not solve the conspiracy problem, it did not play a significant role in phonology until considerably later.

2 Conspiracies: An historical overview

In order to fully understand the argument made in FUPR, one must begin with perhaps the central concern of early generative phonology: specifically, the question of how a language learner deduces the correct grammar from the data to which the language learner is exposed. The rough answer to this question that was advanced was that the learner adopts the "simplest" grammar. Simplicity, at least in terms of phonology, was taken to be determined by reference to the counting (particularly) of feature specifications both in rules and in lexical representations. A critical part of this enterprise was to design grammars so that the phonological patterns most commonly found in languages could be expressed in a simple fashion, while patterns that were never found in languages could be expressed only under great duress. An essential element of this enterprise was the building of a system of notation that would allow what the linguist understood to be the "same" or "related" phenomena to be subsumed under a single rule (a rule that while covering the observed data would often go beyond those data to make predictions about data the learner may not have encountered). (See Chomsky and Halle 1968 (*SPE*) for an extended discussion of this point of view. These ideas can be found throughout the entire early generative phonology literature.)

What FUPR showed was that in synchronic grammars one could find cases where obviously related phenomena could not be given a unitary treatment from the point of view of any available formal notation because they were related not in terms of their actions (insertions, deletions, feature changes, etc.), but rather the structural configurations that they either avoided or strove to achieve. As long as phonology was viewed as a theory of rules which married phonological actions to specific phonological contexts, a solution to the problem of conspiracies was impossible. The one partial solution suggested – derivational constraints – was a tentative step in the direction of separating the action (vowel deletion, in the Yawelmani case) from the context in which it occurs. It was, of course, not until the development of Optimality Theory (cf. Prince and Smolensky 1993) that a total solution emerged.

Although the FUPR paper itself focused on conspiracies in the synchronic grammars of specific languages, the paper developed out of my (ultimately unsuccessful) attempt to extend the ideas of "Chapter Nine" of Chomsky and Halle (1968) by developing a notion of "universal rules." What Chapter Nine attempted to do was to make it formally simpler for a grammar to conform to what is "natural" than for it to go against what is most natural. In terms of phonological representations, it did this by creating a system where an unmarked feature value was cost-free, while a marked feature value rendered the representation a more costly one. The consequence was that whenever possible, an underlying representation would contain an unmarked value rather than a marked value (since generative phonology claimed that the least costly grammar was always chosen over the more costly grammar if both grammars yielded the correct outputs). Chapter Nine went further, however, and attempted to extend the idea of making unmarked specifications cost-free when formulating phonological rules. However, it found only a very restricted way of achieving this goal. In particular, it proposed that if a phonological rule specifies a particular structural change, then markedness principles come into play to add other changes that follow naturally. For instance, Yawelmani has a vowel harmony rule whereby a vowel becomes round when preceded by a round vowel of the same height. This vowel harmony rule affects the vowel /i/ when it stands after the vowel /u/. However, when /i/ rounds, it also becomes back and surfaces as [u]. Chapter Nine suggested that in a case such as this, the vowel harmony rule simply specifies that a vowel acquires the feature [+round] and then markedness principles will automatically add the feature [+back].

In Chapter Nine, the only way that markedness considerations could play a role in "simplifying" the grammar was through this device of "linking" a natural structural change (e.g. the backing of a round vowel) to a language-specific rule (e.g. vowel harmony). As such, Chapter Nine had a proposal only for the case where a secondary structural change is natural given some primary structural change. It did not have an explanation for why the same primary changes occur under similar conditions in language after language (e.g. nasal assimilation, epenthesis of onsets, lenition).

In an earlier version of Kisseberth (1969), I attempted to develop (but later abandoned) the idea that grammars contain a set of universal rules (cost-free, so to speak). This proposal was motivated by the recognition that no matter how much tinkering one did with the system of notation, it would always be possible to state very simple rules that have linguistically implausible consequences. As a consequence, generative phonology's attempt to make more natural rules "easy" to formulate and less natural rules more difficult seemed fundamentally flawed.

The principal difficulties that the search for universal rules faced at the end of the 1960s included:

(i) In a given language, a very natural phonological process accrues a significant number of language-specific restrictions that reduce the generality of the rule needed to account for the phenomenon; it was unclear how to separate the essence of a rule from all the baggage required to properly delimit its scope of application.
(ii) *The conspiracy problem.* Specifically, there are multiple distinct actions (insertions, deletions, feature changes, sequencing changes) that are triggered by essentially the same context. Since a given language may utilize several rules to avoid/achieve a certain structural configuration, the scope of application of each rule must be delimited, obscuring the universal nature of the rule.
(iii) If "universal" means "in every language," why are there languages where these (proposed) universal rules are not in fact implemented in all cases or indeed at all?

Optimality Theory (Prince and Smolensky 1993 and a myriad of subsequent references), of course, ultimately provided a solution to these problems by (a) separating the actions themselves from the "rules" (now expressed as constraints),

(b) allowing constraints to interact with one so that different actions are favored in certain situations over other actions, and (c) postulating a type of constraint (faithfulness) that could suppress the effects of other constraints by outranking them.

As the preceding discussion indicates, FUPR developed out of the problem of developing a notion "universal rule." It could not succeed in solving the problem of conspiracies, since it assumed the existence of (learned) rules. Between FUPR and the optimality-theoretic solution to conspiracies and universal rules, there were many significant phonological developments. Most of these developments have some bearing on the conspiracy problem. Perhaps it will be useful to begin with an observation in McCarthy (1993: 169):

> The idea that constraints on well-formedness play a role in determining phonological alternations, which dates back at least to Kisseberth's (1970) pioneering work, has by now achieved almost universal acceptance. A tacit assumption of this program, largely unquestioned even in recent research, is the notion that valid constraints must state true generalizations about surface structure or some other level of phonological representation. Anything different would seem antithetical to the very idea of a well-formedness constraint.

McCarthy goes on to reject the point of view that the constraints that grammars conspire to enforce are necessarily true surface generalization. But it is important to understand the way in which the idea of conspiracies evolved and McCarthy here identifies a principal theme.

It is true that in the Yawelmani case the ban on complex onsets/codas is (largely) satisfied by the surface representations of the language. FUPR did not claim, however, that this was necessarily the case, but rather allowed for the possibility that a relevant constraint might be true of only a certain stage of the derivation. This conclusion was a necessary one, because FUPR did not propose to abandon the notion that phonological systems are systems where principles interact in a possibly complex way such that some principles may not be true of the surface but only of some other level of the representation. From that point of view, it could very easily be the case that a certain configuration is favored or disfavored through much of a derivation, only to have late, low-level rules derive surface forms where the principle in question is violated. It was well known, for instance, that commonly assumed laws of syllabification in English may be violated in fast speech. In other languages, violations may result from processes operative in careful speech as well.

Although FUPR was careful not to suggest that the phonological targets of conspiracies were surface targets, the discussion that evolved over the subsequent years generally emphasized the surface nature of the constraints that rules conspired to serve (e.g. Haiman 1972; Shibatani 1973; Sommerstein 1974). This emphasis is extremely significant, since it had the consequence that the essential point of FUPR was lost as phonological thinking veered in new directions. The question raised by Kiparsky (1968), "How abstract is phonology?" led several influential phonologists to move in various interrelated directions. On the one hand, "Natural Generative Phonology" (cf. Vennemann 1971, 1974; Hooper 1973, 1979) attempted to limit phonology to surface-true generalizations. On the other hand, the "Natural Phonology" of Stampe (1973) attempted to re-focus phonology away from the alternations observed in the shapes of morphemes (alternations which were often of restricted scope, were non-productive, and had exceptions, and which Stampe considered to be arbitrary and learned) towards processes that were "automatic," "exceptionless," and "innate," which could be observed in a variety of domains such as language acquisition, fast speech, unguarded speech, drunken speech, language games, etc. Both Natural Generative Phonology and Stampe's Natural Phonology did not survive, for a quite simple reason: both approaches essentially removed the many examples of very regular, productive morphophonemic processes from the scope of phonology, since they were usually not surface-true generalizations due to exceptions or interactions with independent phenomena, etc. However, Stampe's Natural Phonology did have a lasting impact, in that it eventually led to the so-called "Lexical Phonology" approach, which made a significant attempt to distinguish among principles that obtained in the lexicon and principles with a wider scope of application that were not dependent on the particulars of morphological structure (cf. Kiparsky 1982, 1985; Mohanan 1986, 1995). To this day this remains a critical issue in working out a comprehensive theory of phonology (CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME).

The debate with regard to the "abstractness" of phonology (i.e. the extent to which surface forms may differ from their underlying sources and what sorts of evidence are required in order to postulate a divergence between surface and underlying structure) had considerable implications for the concerns of FUPR, as well as the ultimately related notion of universal rules. If phonology has little abstractness, and if most of the rules that had been proposed during the early years of generative phonology were not really rules of phonology at all, then perhaps arguments such as the one presented in FUPR are irrelevant. If the various rules that we claimed conspire to avoid complex margins in Yawelmani are not in fact real rules of the language, then the argument in FUPR is no argument at all. And if the only phonological rules are ones that are directly represented by overt surface forms, then explaining how a phonological system is learned no longer seems so challenging, and appeal to universal considerations is less necessary.

The abstractness controversies of the early 1970s, however, were never really resolved, but instead phonologists turned to a new approach to phonological analysis and universals, namely an approach that emphasized the development of phonological representations from which outputs could be predicted with a minimal appeal to rules and rule interaction. The motto of phonology became "if the representations are right, then the rules will follow" (McCarthy 1988: 84). For our purposes, we will refer to this as *representational phonology* (cf. Goldsmith 1976, 1990; Clements and Goldsmith 1984; Clements 1985; and many other references). Although representational phonology had very considerable successes (e.g. the autosegmental approach provided substantial insights into the complicated tonal phenomena of Bantu languages and the vowel harmony patterns of a variety of languages; CHAPTER 114: BANTU TONE; CHAPTER 45: THE REPRESENTATION OF TONE; CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS; CHAPTER 118: TURKISH VOWEL HARMONY; CHAPTER 123: HUNGARIAN VOWEL HARMONY), it became apparent from several of its most significant contributions that an adequate account of phonological patterns requires appeal not just to representations and rules, but also to constraints. The introduction of constraints can be found in such papers as Itô's (1989) theory of epenthesis, the extensive literature on the

Obligatory Contour Principle or concepts such as word-minimality in the work on prosodic morphology (cf. McCarthy 1979, 1981; McCarthy and Prince 1986).

The return to a role for constraints in phonological thinking naturally also triggered a return to relevance of the notion of conspiracies and constraints on derivations. Papers such as Myers (1991) on the notion of "persistent rules," the Theory of Constraints and Repairs developed in Paradis (1988), and ultimately Optimality Theory all found the issue identified in FUPR to be a significant one that needed to be addressed in phonological theory. We shall discuss the optimality-theoretic analysis of the conspiracy phenomena later, but at this point we would like to turn to specific examples of conspiracies.

3 Conspiracies in various domains of phonological research

The argument in FUPR is limited to a single synchronic phonological system, but the notion is relevant to most, if not all of the domains of phonological exploration: universals, variation, and dialectology (CHAPTER 92: VARIABILITY), language change (CHAPTER 93: SOUND CHANGE), acquisition (CHAPTER 101: THE INTERPRETATION OF PHONOLOGICAL PATTERNS IN FIRST LANGUAGE ACQUISITION), loanword phonology (CHAPTER 95: LOANWORD PHONOLOGY), etc. Space limitations do not permit an extensive discussion of all the domains where the notion of conspiracies is relevant, but we present some brief discussion of several domains: synchronic grammars, universal rules, phonological acquisition, and loanword phonology.

3.1 Conspiracies in synchronic grammars

We have already discussed at length the conspiracy in Yawelmani that revolves around the avoidance of complex syllable margins. There is a great variety of other examples of conspiracies that have been discussed over the past few decades. Here we will illustrate just two: hiatus avoidance conspiracies and conspiracies banning sequences of a nasal consonant followed by a voiceless obstruent.

Many languages do not allow onsetless syllables, particularly in word-medial position (cf. Casali 1997; CHAPTER 61: HIATUS RESOLUTION; CHAPTER 55: ONSETS for discussion). Traditionally, such languages are said to avoid hiatus (a succession of two vowels with no intervening consonant). There are, of course, several ways in which a VV sequence (hiatus) may be avoided. The first or the second vowel may be deleted. The first or the second vowel may undergo glide formation. A consonant may be inserted between the vowels. While some languages may choose to avoid hiatus by invoking a single anti-hiatus action, it is not at all uncommon to find a language invoking multiple actions according to the specific VV sequence.

Chicano Spanish provides a relevant example (cf. Hutchinson 1974; Reyes 1976; Baković 2007). Consider the data in (8), taken from Baković (2007). We have used orthographic representation for the input and retained that representation for the output, except that we put the surface form resulting from the hiatus avoidance rules inside brackets to highlight how the hiatus is dealt with.

(8)	a.	*tu uniforme*	*t*[u]*niforme*	'your uniform'	
		lo odio	*l*[o]*dio*	'hate (1SG him/it)'	
		era así	*er*[a]*sí*	'it was like that'	
		se escapó	*s*[e]*scapó*	'escaped (3SG)'	
		mi hijo	*m*[i]*jo*	'my son'	(*h* is silent)
	b.	*paga Evita*	*pag*[e]*vita*	'Evita pays'	
		la iglesia	*l*[i]*glesia*	'the church'	
		casa humilde	*cas*[u]*milde*	'humble home'	(*h* is silent)
		niña orgullosa	*niñ*[o]*rgullosa*	'proud girl'	
	c.	*mi obra*	*m*[jo]*bra*	'my deed'	
		mi ultima	*m*[ju]*ltima*	'my last one (FEM)'	
		mi hebra	*m*[je]*bra*	'my thread'	
		mi arbol	*m*[ja]*rbol*	'my tree'	
		tu epoca	*t*[we]*poca*	'your time'	
		tu alma	*t*[wa]*lma*	'your soul'	
		tu hijo	*t*[wi]*jo*	'your son'	(*h* is silent)
		su Homero	*s*[wo]*mero*	'your Homer'	(*h* is silent)
	d.	*me urge*	*m*[ju]*rge*	'it is urgent to me'	
		pague ocho	*pagu*[jo]*cho*	'that s/he pay eight'	
		porque aveces	*porqu*[ja]*veces*	'because sometimes'	
		como Eva	*com*[we]*va*	'like Eva'	
		tengo hipo	*teng*[wi]*po*	'I have the hiccups'	(*h* is silent)
		lo habla	*l*[wa]*bla*	'speaks it'	(*h* is silent)
	e.	*como uvitas*	*com*[u]*vitas*	'like grapes (DIM)'	
		se hinca	*s*[i]*nca*	'kneels'	(*h* is silent)

The data in (8) show that when two vowels are juxtaposed in Chicano Spanish, these VV sequences are not resolved in a single way. (7a) shows that when two identical vowels are adjacent to one another, the sequence is reduced to a single vowel. Thus in *tu uniforme*, a single [u] vowel is found. It is of course not readily apparent whether one or the other vowel is deleted or whether one should just say the two vowels coalesce.

(8b) demonstrates that a word-final low vowel *a* deletes before any vowel. Thus in *paga Evita*, the *a* at the end of the verb is absent in pronunciation. But the first vowel of the VV does not always delete. (8c) shows that if the first vowel is high, then it becomes the corresponding glide, whatever the second vowel might be. Thus *mi obra* yields [mjo] and *tu epoca* yields [twe]. On the other hand, if the first vowel is mid, the results are a bit more complex. (8d) shows that if the second vowel differs from the first with respect to either [±low] or [±back], then the initial vowel glides: *me urge* surfaces with [mju], *como Eva* results in [mwe] and *lo habla* (the *h* of the orthography is silent) becomes [lwa]. However, if the vowel that follows the mid vowel differs from it only with respect to the feature [±high], then there is no glide formation. Rather, as shown in (8e), the two vowels coalesce in a form identical to the second vowel: *como uvitas* results in [mu]. What this example from Chicano Spanish illustrates is that although a single strategy might avoid hiatus, languages may choose multiple means (for example vowel coalescence, vowel deletion, glide formation) to eliminate the problematic structures.

Combinations of a nasal consonant and a following voiceless consonant are disfavored in many languages (cf. Hayes and Stivers 2000 for discussion of the phonetic preference for voiced consonants following nasals; see also CHAPTER 8: SONORANTS). There are different ways in which these ill-formed consonant sequences could be avoided. The most common "repairs" are voicing the post-nasal consonant or deleting this consonant (while assimilating the nasal to the same point of articulation), or devoicing the nasal or even deleting it all together.

Pater (1999) points out that in various languages, nasal–voiceless stop sequences are avoided by the application of more than one rule (even though a single rule in principle could resolve the problem). For example, in Kwanyama, a Bantu language discussed in Steinbergs (1985), there is evidence to support a ban on nasal–voiceless consonant sequences. One piece of evidence for the ban is the absence of such sequences in morpheme-internal position. It is also the case that the sounds [k] and [g] are in complementary distribution. [k] occurs word-initially and intervocalically, and [g] appears only after nasals. This distributional pattern supports the proposition that there is a principle that voices a consonant after a nasal. Such a proposal is also supported by the treatment of English loanwords, as shown in (9) below.

(9) *Post-nasal voicing in Kwanyama loanwords*

[sitamba]	'stamp'
[pelenda]	'print'
[oinga]	'ink'

In these borrowings, English nasal–voiceless stop sequences are replaced by nasal–voiced stop sequences.

But voicing of the stop is not the only repair found in Kwanyama. A root-initial voiceless stop located after a nasal prefix requires assimilation of the nasal, but then elides from the representation.

(10) *Root-initial nasal substitution in Kwanyama*

/eːN + pati/	[eːmati]	'ribs'
/oN + pote/	[omote]	'good-for-nothing'
/oN + tana/	[onana]	'calf'

In two other Bantu languages, Umbundu and Luyana, the same ban on nasal–voiceless consonants can be found. In these languages we again find two different repairs. A nasal that comes to stand in front of a voiceless fricative elides, while a nasal in front of a stop will assimilate the stop's point of articulation, but the stop itself elides. Schadeberg (1982) illustrates from Umbundu that /N + tuma/ surfaces as [numa] 'I send', while /N + seva/ surfaces as [seva] 'I cook'. Givón (1970) shows that in Luyana /N + tabi/ surfaces as [nabi] 'prince', while /N + supa/ surfaces as [supa] 'soup'.

While Kwanyama, Umbundu, and Luyana use two distinct strategies to avoid a nasal–voiceless consonant sequence, other languages may opt for a uniform repair. According to Pater, in Indonesian a nasal assimilates the point of articulation of both a voiceless fricative and a voiceless stop, with the oral stop then eliding. In other languages, like Kelantan Malay, Venda, Swahili, and Maore, nasals delete before voiceless stops and fricatives alike. Although uniform avoidance strategies are possible, many languages are like Kwanyama, Umbundu, and

Luyana in that they opt for a (perhaps only superficially) more complex pattern of avoidance.

3.2 *Phonological conspiracies and universal grammar*

As mentioned earlier, the FUPR paper arose out of an attempt to make some sense out of the notion that there might be "universal" rules. The problem that confronted the researcher during the generative phonology period was that while one could easily find phenomena that seemed to reflect some universal principle, the rule-based descriptions were rarely uniform across languages in their details. Furthermore, just as in the case of conspiracies in synchronic grammars, sometimes formally unrelated rules in different languages could be involved in achieving the same outcome.

Let us take an example from Bantu tonal systems (CHAPTER 45: THE REPRESENTATION OF TONE; CHAPTER 114: BANTU TONE). Early on in the research on these systems, it was recognized that a *sequence* of H tones is not preferred in these languages. Meeussen (1963) observed that in Tonga, for example, a succession of two H tones is converted to HL. In Leben (1973) and Goldsmith (1976), this ban on successive H tones was seen as a natural consequence of the approach to phonology that came to be known as autosegmental phonology. Specifically, autosegmental models typically represent surface sequences of a feature value (e.g. H tone) as a single multiply linked autosegment. As a consequence, to the extent that this representation is maximized, sequences of the same autosegment will be unexpected. The proposed constraint against successive identical autosegments became known as the Obligatory Contour Principle (OCP). Although some support for the idea that the OCP constrains all features emerged in the phonological literature, there is no question that its tonal instantiation is by far the most robust evidence for the principle.

The notion of a ban on successive H tones depends on the ability to distinguish between *true* and apparent sequences of H tones. This distinction is captured in autosegmental phonology as follows. A true H tone is one located on the tonal tier, regardless of whether it is associated with one or more than one tone-bearing unit. Successive H-toned moras are not sequences of H tones if they are all linked to a single H tone on the tonal tier.

There are a number of ways in which the *HH principle can be manifested in a language. But before looking at these manifestations, two quite separates matters must be mentioned, both of which dramatically expand the diversity of the manifestations of the OCP. First of all, in the analysis of Bantu languages, it has sometimes been argued that phonologically there is just a contrast between H tone and the absence of tone. However, even in analyses that utilize inputs that lack L tones, rules have been proposed that derive L tones that then contrast with toneless moras. Consequently, we have some analyses where a H tone that violates the OCP may be simply deleted, and other analyses where it is changed to L. Formally, the rules are quite distinct, but in both versions the ban on successive H tones is satisfied. A second complication has to do with what it means for two H tones to be adjacent and thus in violation of the *HH ban. In some languages, adjacency of H tones on the tonal tier is the defining characteristic; in other languages, what is significant is that the H tones may not be associated with successive syllables; in yet other languages, what is critical is that the H tones not be linked to successive moras. Because of these differences in adjacency, there will be considerable variation from language to language with respect to which representations actually violate the ban on HH in those languages.

Let us now look at some of the different ways in which the *HH ban is implemented in different languages. One implementation has to do with the very nature of the underlying representations found in a given language (CHAPTER 1: UNDERLYING REPRESENTATIONS). For example, Cassimjee (1992) shows that in Venda (a Bantu language spoken in South Africa and adjacent parts of Zimbabwe) noun stems, there may be sequences of H-toned moras, but in every case there is evidence that these sequences consist of a single H tone on the tonal tier associated with multiple successive moras. There are no morphemes with successive true H tones. In other words, underlying representations are structured so as to avoid violations of the OCP. The evidence that these H-toned sequences are a single H tone comes from a morphophonemic phenomenon known as Meeussen's Rule (Goldsmith 1984), whereby a H tone that is immediately preceded by a H tone is changed to L. For example, a noun such as /góŋ´óŋ´ó/ 'bumblebee' will, when preceded by a verb ending in a H tone, change first of all to the intermediate form: */gòŋ´òŋ´ò/, due to Meeussen's Rule: i.e. all three syllables become L-toned (indicating that all three syllables started off linked to a single H tone that then changed to L). Subsequently, the preceding H tone spreads onto the first syllable forming a contour tone: */gô/. Falling tones in Venda, however, are only permitted on bimoraic vowels and bimoraic vowels occur only in the penultimate syllable of an Intonational Phrase. As a consequence, /gô/ surfaces simply as a H-toned syllable: [góŋ´òŋ´ò]. If /góŋ´óŋ´ó/ were analyzed as having a sequence of three H tones rather than having a single H-tone multiply linked, we would have to explain why Meeussen's Rule does not affect the underlying representation of this word (see Cassimjee 1992 for more detailed discussion).

Even if underlying representations are configured to avoid violations of the OCP (as in Venda), it still may happen that the juxtaposition of morphemes yields potential HH sequences. One very common reaction to this threat is the deletion/lowering of the rightmost of these adjacent H tones as shown above for Venda. Rules that target the rightmost H in a HH sequence are said to be instantiations of "Meeussen's Rule" (cf. Goldsmith 1984). In Venda it is necessary for Meeussen's Rule to change H to L in order to obtain the right results. In the Ikorovere dialect of Emakhuwa (spoken in southern Tanzania), Meeussen's Rule simply deletes the H tone. For example, an underlying form like /k-a-ho̲-ka̲viha/ 'I helped' (underlining indicates the location of input H tones) has a H-toned tense-aspect morpheme /ho/ followed by a stem /kaviha/, which has a H tone (predictably) associated with its first vowel. This morpheme sequence violates the OCP ban *HH. The second of these H tones deletes, but subsequently the first H tone doubles onto the second by a general High Tone Doubling rule that applies in a wide variety of circumstances in the language: [k-a-hó̲-ká̲viha]. Of course, one might ask: how do we know that the H tone on the stem-initial /ka/ has deleted if in fact this mora is pronounced on a H tone? The answer is simple. If the underlying H tone on /ka/ had not deleted, then it would have triggered doubling onto the next mora, resulting in the ill-formed *[k-a-hó̲-ká̲viha]. Notice that it is clear that the H tone on /ka/ is deleted and not the H tone on the preceding morpheme /ho/. If we had deleted the H tone from /ho/, then we would predict the incorrect output *[k-a-ho̲-ká̲víha], since the second H tone would double to its right. (See Kenstowicz and Kisseberth 1979 for a more extended discussion of Emakhuwa tone, albeit in a pre-autosegmental framework.)

Although deletion of the rightmost H is particularly common in Bantu, other responses to violations of the *HH constraint can be found. In some cases it is the leftmost of the two H tones that deletes or changes to L. For instance, in the Bantu language Rimi (Yukawa 1989), underlying H tones shift systematically one vowel to the right. Thus [u-t<u>e</u>ghéja] 'to understand' has a H-toned verb stem, where the H tone is underlying on the stem-initial vowel (underlined) but surfaces on the syllable [ghe]. In an example like [u-v<u>a</u>-ríghitja] 'to speak to them', the verb stem is toneless but the object prefix /va/ bears an underlying H tone that shifts onto the first syllable of the verb stem. When a H-toned object prefix precedes a H-toned verb stem, as in [u-v<u>a</u>-t<u>e</u>ghéya], we see that the object prefix loses its H tone and the syllable [te] retains its H tone, although this H tone does shift to the next vowel. Rimi thus differs from languages like Venda and Emakhuwa in that a violation of the OCP is repaired by deleting the leftmost H rather than the rightmost H.

In yet other cases, the two adjacent H tones are merged into a single H tone that is still linked to all of the moras that the original H tones were linked to. We can refer to this as H-tone fusion. The evidence for H-tone fusion is sometimes a bit indirect. The Bantu language Shambaa (Odden 1982) provides an interesting example, however, since in addition to the need for H-tone fusion, it also illustrates an entirely different means of avoiding *HH violations.

In Shambaa, whenever a sequence of H tones would be created (either within a word or across words), a downstep is inserted between the H tones. For example, the second H tone in each of the following examples is downstepped relative to the first (downstep is indicated by the downward arrow and an underlying H tone is indicated by underlining; the data also illustrate H tone spreading, but we do not discuss this aspect of the data): [at<u>é</u>-k↓<u>ó</u>má] 'he killed', [ang<u>é</u>-↓j<u>á</u>] 'he should have cooked', [<u>á</u>zak<u>ó</u>má nj↓<u>ó</u>ká] 'he killed a snake', and [n<u>í</u> k↓<u>ú</u>i] 'it is a dog'. These data suggest clearly that a sequence H!H does not count as a violation of *HH. It should be noted that while in some languages downstep may derive from a so-called "floating" L tone, this is not the case in Shambaa. We do not address here the issue of how downstep is represented, nor whether it is represented in the phonology or only in the phonetics.

The insertion of a downstep between successive H tones in Shambaa is a very general phenomenon, but there are some cases where successive input H tones are *not* separated by a downstep. For instance, there is no downstep between a H-toned object prefix and a H-toned verb stem: /ku-w<u>á</u>-k<u>ó</u>má/ 'to kill them'. Odden explains the failure of a downstep to be inserted at this juncture by proposing that the H tone of the object prefix and the H tone of the verb stem fuse into a single multiply linked H tone. As a consequence, there is a single H tone and insertion of downstep cannot occur (since downstep is used only between H tones).

In all the preceding examples, we have dealt with situations where an input would violate *HH and formally different rules operate to alter the representation so that it no longer has a H H sequence. The OCP ban *HH is manifested in other ways in the grammars of the world's languages. Recall from our discussion of Yawelmani how the ban on complex margins may serve to restrict the application of vowel deletion rules. The same thing can be observed in tonal systems: the ban on *HH may restrict the application of other tone rules. The most commonly observed phenomenon where *HH plays a restrictive role on another tonological process is in H-tone spreading. The precise formulation of H-tone spreading differs from language to language (at least in a rule-based model of phonology), but one overarching pattern is that spreading may be prevented from going onto a mora that itself is adjacent to a H-toned mora. It should be observed that in some languages, it is not spreading but shifting (i.e. spreading of a H followed by a delinking from all but the last mora in the spreading structure) that is blocked.

For example, in Isixhosa (cf. Cassimjee and Kisseberth 1998) the H tone on the subject prefix /bá-/ shifts to the following toneless syllable in [ba-yá-lwa] but is unable to do so in [bá-ya-bóna] due to the fact that the prefix /ya/ is followed in this case by a H-toned syllable. (As in our earlier examples, a mora that bears an underlying H tone is underlined.) Sometimes the adjacency of the H tones may be obfuscated. In Isixhosa the H on /bá-/ does not shift in [bá-ya-bonísa], even though at first glance it does not seem that the /ya/ is adjacent to a H-toned syllable. The problem in this example is that there is a H tone on the syllable /bó/ in the input, but in the output this H tone has shifted to the following syllable. What we observe here is the much discussed problem of phonological opacity: the H-toned nature of /bo/ serves to block spreading onto the syllable in front of it, even though in fact /bo/ is toneless on the surface.

The preceding discussion shows that if we look at *HH (an instantiation of the OCP) across a diverse set of languages, we find that it is entirely parallel to the constraint against complex margins in the synchronic grammar of Yawelmani. There is a "functional" unity that unites all these diverse ways of avoiding HH sequences: they are working towards the same end, representations that lack HH sequences. Any theory (such as generative phonology) that sees rules as devices that marry a structural change to a structural description will fail to express the universal principle *HH.

3.3 *Phonological conspiracies in phonological acquisition*

Much of the work on the child's acquisition of the phonology of a language assumes that the child accurately perceives (for the most part) the data to which s/he is exposed, but that various markedness principles (e.g. preferences for open syllables, preferences for oral over nasal vowels, preferences for stops over fricatives) restrict the child's attempt to produce an output faithful to that perception (CHAPTER 101: THE INTERPRETATION OF PHONOLOGICAL PATTERNS IN FIRST LANGUAGE ACQUISITION). As early as the extremely important work of Smith (1973), it was recognized that the notion of conspiracies is as relevant to child phonology as it is to adult phonologies. Smith proposed various rules to account for the fact that the child Amahl simplified consonant clusters in adult speech, but just as in Yawelmani, the unity found in the child's output was not reflected in the diversity of the rules that achieved this unity. Smith recognized this as a failing of his rule-based, derivational approach. Naturally, constraint-based approaches such as OT have gained currency in the field of phonological acquisition, at least in part because of their ability to capture conspiracies better in child language acquisition. It should be clear from the discussion throughout this chapter that "conspiracies" are first and foremost attempts to achieve an unmarked structure or avoid a marked structure. Since markedness principles serve to shape the child's outputs, it follows that we will expect conspiracies to be manifested.

Space considerations limit us to a single instance of a conspiracy in phonological acquisition. Pater (2002) and Pater and Barlow (2003) discuss the ban on fricatives, *FRICATIVE, which can be observed in the data on phonological acquisition (see also CHAPTER 28: THE REPRESENTATION OF FRICATIVES). The common repair of this constraint is for a fricative to be converted to the corresponding stop. Pater discusses a child LP65, aged 3:8 with a phonological delay, who in acquiring English lacked fricatives entirely from her output. However, adult English forms were repaired in two different ways: the deletion of the fricative or the stopping of the fricative. The choice of the repair was dependent on the adult input. If the fricative was part of a cluster, it was deleted. If it was not part of a cluster, it was converted to the corresponding stop. Thus *sneeze* became [ni:d], *three* became [wi], *drive* became [waɪb]. Two different repairs secure the absence of fricatives, one of the many examples of conspiracy in the acquisition literature.

3.4 Conspiracies in loanword phonology

When speakers of L1 adapt words from L2 for use in speaking their native language, some of these words may contain structures that violate a constraint operative in L1. If these words are fully nativized, then these structures will be altered so as to avoid violations of the constraint in question. If the notion of conspiracies is applicable to the phenomenon of loanword adaptation then we expect that there will be cases where a given constraint will be enforced by means of quite distinct adaptation strategies (see also CHAPTER 95: LOANWORD PHONOLOGY).

In the Australian Aboriginal language Gamilaraay, words must end either in a vowel or a coronal sonorant ([n j l rr]), according to McManus (2008). If Gamilaraay borrows an English word that ends in a consonant that is permitted word-finally, then that consonant will surface. Thus English *barrel* is realized as [baril], and *poison* as [baadjin]. When an English word ending in a labial or dorsal consonant is borrowed, an epenthetic vowel appears. This epenthetic vowel obviously functions to avoid a disallowed coda consonant. Some examples of epenthesis are given in (11).

(11)	baaybuu	'pipe'	nhaayba	'knife'
	dhuubuu	'soap'	yurraamu	'rum'
	milgin	'milk'	yurrugu	'rope'

However, when the English word ends in one of the coronal obstruents, which are not allowed in a word-final coda, an epenthetic vowel is *not* inserted; rather, the coronal is converted to a sonorant.

(12)	bulaang.giin ~ bulang.giin	'blanket'	burrgiyan	'pussy cat'
	marrgin	'musket'	yuruun ~ yurruun	'road'
	dhalbin	'tablet'		
	garaarr	'grass'	nhiigiliirr	'necklace'
	dhindirr	'tin dish'	maadjirr	'matches'
	yarrarr	'rice'	gabirr	'cabbage'

Gamilaraay employs both vowel epenthesis and the sonorization of a consonant to achieve an output where every output ends in a vowel or a sonorant coronal. Two formally distinct alterations yield an output that conforms to the same regularity.

4 The optimality-theoretic analysis of conspiracies

In the preceding section we provided some illustration of the relevance of the notion "conspiracy" in various domains of phonology. An examination of the phonological literature over the four decades since the publication of FUPR is rife with examples of the phenomenon. Although FUPR identified a facet of phonological structure that is no doubt of critical importance to the theory of phonology, it failed to offer a comprehensive solution to the problem that conspiracies pose. In addition, it failed to raise a significant question: *why* do conspiracies exist? Why do languages not employ a single device to achieve a preferred structure or to avoid a structure that is not preferred?

Optimality Theory provides the essential ingredients of both a comprehensive account of conspiracies and an explanation for why universal "rules" may not be reflected in the outputs of particular languages. By separating the constraints from the actions that repair potential violations of these constraints, it allows a constraint to both trigger some actions while preventing others. Thus *ComplexMargins can both trigger the appearance of an epenthetic vowel and block the elision of a vowel if elision would violate *ComplexMargins. OT is not a theory of actions, but rather a theory of how constraint interactions account for the pattern of observed actions. At the same time, the existence of highly ranked faithfulness constraints may prevent a constraint violation from being repaired at all in some languages.

Optimality Theory also explains why conspiracies occur, and can perhaps even be expected. Given a constraint such as *ComplexMargins, there are several different repairs that might avoid complex margins. However, each of these actions necessarily violates some other constraint (at the same time that they avoid a violation of *ComplexMargins). The constraints that the repair violates may be a faithfulness constraint, a markedness constraint, or some sort of morphological constraint. Since these other constraints have a particular ranking, this ranking will determine which repair to *ComplexMargins is optimal in a given situation.

Almost any optimality-theoretic description of a language will offer clear evidence that it has provided an insightful account of conspiracies. There is, however, a significant problem with the OT analysis of conspiracies. Specifically, for any constraint C, there are many logical "actions" that might repair a representation so that C is not violated. OT seems to claim that any of these actions could occur in some language. But is this in fact true? Various linguists have suggested that it is not true that all "repairs" for the violation of a constraint are in fact possible. This has been labeled the "Too Many Solutions" problem (cf. Steriade 2001, 2009). For example, Steriade considers the constraint that disfavors voiced obstruents at the end of a word. She notes that there are a considerable variety of phonological actions that might result in outputs that do not violate this constraint. Taking an input with a final /b/ as an example, the following repairs could avoid a violation of the constraint: (a) the devoicing of the /b/ to [p], (b) the nasalization of the /b/ to [m], (c) the lenition of /b/ to [w], (d) deletion of the /b/, (e) the insertion of an epenthetic vowel after /b/, and (f) the metathesis of the /b/ with a preceding consonant that does not violate the constraint, etc. Steriade argues, however, that in fact it is only devoicing that is utilized to repair violations of the ban on word-final voiced obstruent.

The "Too Many Solutions" problem is a critical issue for Optimality Theory, since it calls into question the foundation of the theory – i.e. the separation of phonological actions from the constraints on structure that trigger these actions. It was the need to find an account of conspiracies that led to an abandonment of "rules" in favor of a set of constraints whose ranking determines the optimal action. So it is natural that a challenge to its analysis of conspiracies is at the same time a challenge to its very foundations.

There have been various attempts to solve the "Too Many Solutions" problem in OT: the "P-map" proposal in Steriade (2001, 2009), the "targeted constraints" in Wilson (2001), and the appeal to procedural markedness principles ("implicational constraint principle") in Blumenfeld (2006) are examples. It is beyond the scope of this chapter to explore the adequacy of these different attempted solutions, but there is no question that the issue is a central one in the exploration of OT approaches to phonology. Despite the challenges to the OT account of conspiracies, there is no doubt that much of the motivation for Optimality Theory resides in the advances that it made in explaining conspiracies, and these advances have been considerable indeed.

5 Conclusion

In this chapter we have explained the notion "conspiracy" in phonology and have illustrated its relevance to several domains of phonological investigation: synchronic grammars, phonological universals, the acquisition of phonology, and loanword adaptation. We have attempted to explain the historical background out of which the notion emerged, specifically the attempt to find what is universal in systems of phonological rules, and the reasons why an adequate solution was not available. We concluded with the observation that while Optimality Theory goes a long way towards providing an insightful account of conspiracies, it must still deal with the Too Many Solutions problem.

REFERENCES

Archangeli, Diana. 1984. Underspecification in Yawelmani phonology and morphology. Ph.D. dissertation, MIT.

Baković, Eric. 2007. Hiatus resolution and incomplete identity. In Fernando Martínez-Gil & Sonia Colina (eds.) *Optimality-theoretic studies in Spanish phonology*, 62–73. Amsterdam & Philadelphia: John Benjamins.

Blumenfeld, Lev. 2006. Constraints on phonological interactions. Ph.D. dissertation, Stanford University.

Boersma, Paul. 1997. The elements of Functional Phonology. Unpublished ms., University of Amsterdam (ROA-173).

Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73. 493–533.

Cassimjee, Farida. 1992. *An autosegmental analysis of Venda tonology.* New York: Garland.

Cassimjee, Farida & Charles W. Kisseberth. 1998. Optimal Domains Theory and Bantu tonology: A case study from Isixhosa and Shingazidja. In Larry M. Hyman & Charles W. Kisseberth (eds.) *Theoretical aspects of Bantu tone*, 33–132. Stanford: CSLI.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. & John A. Goldsmith (ed.) 1984. *Autosegmental studies in Bantu tone.* Dordrecht: Foris.

Cole, Jennifer & Charles W. Kisseberth. 1995. Restricting multi-level constraint evaluations: Opaque rule interaction in Yawelmani vowel harmony. In Keiichiro Suzuki & Dirk Elzinga (eds.) *Proceedings of the 1995 Southwestern Workshop on Optimality Theory (SWOT)*, 18–38. Tucson: Department of Linguistics, University of Arizona.

Givón, Talmy. 1970. *The Si-Luyana language.* Lusaka: Institute for Social Research, University of Zambia.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT. Published 1979, New York: Garland.

Goldsmith, John A. 1984. Meeussen's Rule. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*, 245–259. Cambridge, MA: MIT Press.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology.* Oxford & Cambridge, MA: Blackwell.

Haiman, John. 1972. Phonological targets and unmarked structures. *Language* 48. 365–377.

Harris, Zellig S. 1944. Yokuts structure and Newman's grammar. *International Journal of American Linguistics* 10. 196–211.

Hayes, Bruce & Tanya Stivers. 2000. Postnasal voicing. Unpublished ms., University of California, Los Angeles.

Hockett, Charles F. 1967. The Yawelmani basic verb. *Language* 26. 278–282.

Hockett, Charles F. 1973. Yokuts as a testing ground for linguistic methods. *International Journal of American Linguistics* 39. 63–79.

Hooper, Joan B. 1973. Aspects of natural generative phonology. Ph.D. dissertation, University of California, Los Angeles.

Hooper, Joan B. 1979. Substantive principles in natural generative phonology. In Daniel A. Dinnsen (ed.) *Current approaches to phonological theory*, 106–125. Bloomington: Indiana University Press.

Hutchinson, Sandra P. 1974. Spanish vowel sandhi. In *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 10(2). 184–192.

Itô, Junko. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7. 217–259.

Kenstowicz, Michael & Charles W. Kisseberth. 1977. *Topics in phonological theory.* New York: Academic Press.

Kenstowicz, Michael & Charles W. Kisseberth. 1979. *Generative phonology: Description and theory.* New York: Academic Press.

Kiparsky, Paul. 1968. How abstract is phonology? Distributed by Indiana University Linguistics Club.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Kisseberth, Charles W. 1969. Theoretical implications of Yawelmani phonology. Ph.D. dissertation, University of Illinois.

Kisseberth, Charles W. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1. 291–306.

Kuroda, S.-Y. 1967. *Yawelmani phonology*. Cambridge, MA: MIT Press.
Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT.
McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.
McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12. 373–418.
McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.
McCarthy, John J. 1993. A case of surface constraint violation. *Canadian Journal of Linguistics* 38. 169–195.
McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.
McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.
McManus, Hope E. 2008. Loanword adaptation: A study of some Australian Aboriginal languages. Honours thesis, University of Sydney.
Meeussen, A. E. 1963. Morphotonology of the Tonga verb. *Journal of African Languages* 2. 72–92.
Mohanan, K. P. 1986. *The theory of Lexical Phonology*. Dordrecht: Reidel.
Mohanan, K. P. 1995. The organization of the grammar. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 24–69. Cambridge, MA & Oxford: Blackwell.
Myers, Scott. 1991. Persistent rules. *Linguistic Inquiry* 22. 315–344.
Newman, Stanley. 1944. *Yokuts language of California*. New York: Viking Fund.
Odden, David. 1982. Tonal phenomena in KiShambaa. *Studies in African Linguistics* 13. 177–208.
Paradis, Carole. 1988. On constraints and repair strategies. *The Linguistic Review* 6. 71–97.
Pater, Joe. 1999. Austronesian nasal substitution and other NC̦ effects. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*, 310–343. Cambridge University Press.
Pater, Joe. 2002. Form and substance in phonological development. *Proceedings of the West Coast Conference on Formal Linguistics* 21. 348–372.
Pater, Joe & Jessica Barlow. 2003. Constraint conflict in cluster reduction. *Journal of Child Language* 30. 487–526.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.
Reyes, Rogelio. 1976. Studies in Chicano Spanish. Ph.D. dissertation, Harvard University.
Schadeberg, Thilo C. 1982. Nasalization in Umbundu. *Journal of African Languages and Linguistics* 4. 109–132.
Shibatani, Masayoshi. 1973. The role of surface phonetic constraints in generative phonology. *Language* 49. 87–106.
Smith, Neil V. 1973. *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.
Sommerstein, Alan H. 1974. On phonotactically motivated rules. *Journal of Linguistics* 10. 71–94.
Stampe, David. 1973. A dissertation on natural phonology. Ph.D. dissertation, University of Chicago. Published 1979, Garland: New York.
Steinbergs, Aleksandra. 1985. The role of MSCs in OshiKwanyama loan phonology. *Studies in African Linguistics* 16. 89–101.
Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 219–250. San Diego: Academic Press.
Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Studies in honor of Paul Kiparsky*, 151–179. Cambridge, MA: MIT Press.

Vennemann, Theo. 1971. Natural generative phonology. Paper presented at the 45th Annual Meeting of the Linguistic Society of America, St Louis.

Vennemann, Theo. 1974. Phonological concreteness in natural generative grammar. In Roger Shuy & Charles-James N. Bailey (ed.) *Toward tomorrow's linguistics*, 202–219. Washington, DC: Georgetown University Press.

Wilson, Colin. 2001. Consonant cluster neutralization and targeted constraints. *Phonology* 18. 147–197.

Yukawa, Yasutoshi. 1989. A tonological study of Nyaturu verbs. In Yasutoshi Yukawa (ed.) *Studies in Tanzanian languages*, vol. 2, 451–480. Tokyo: ILCAA.

71 Palatalization

Alexei Kochetov

1 Introduction

The term "palatalization" denotes a phonological process by which consonants acquire secondary palatal articulation or shift their primary place to, or close to, the palatal region. This usually happens under the influence of an adjacent front vowel and/or a palatal glide (e.g. [ki] → [kʲi], [tja] → [tʃa]). As such, palatalization is a type of consonant–vowel interaction. The term may also refer to a phonemic contrast between consonants with secondary palatal articulation and their non-palatalized counterparts (e.g. [pʲa] *vs.* [pa]). The primary focus of this chapter will be on palatalization as a process, and particularly as a synchronic phonological process manifested in segmental alternations.

Palatalization has been typically viewed as a classic example of a "natural" phonological process – one that is widely attested across world languages and has a clear phonetic motivation, such as in consonant-to-vowel co-articulation (e.g. Hyman 1975; see also CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). However, many formal accounts of palatalization undertaken over the last 40 years have faced considerable challenges. These challenges partly stem from the fact that palatalization processes show a wide range of manifestations – across languages and within a given language. Many synchronic palatalization processes also exhibit complex phonological and morphological conditioning and pervasive opacity effects, reflecting complicated historical sound changes and paradigmatic restructuring. Given this, the difficulty faced by many theoretical approaches has been in providing an empirically adequate and uniform formal treatment of the phenomenon, capturing both cross-linguistic and language-particular generalizations. The goal of this chapter is to review some of the key formal accounts of palatalization, focusing particularly on the challenges posed by the phenomenon for generative theories of phonological representations.

The chapter is organized as follows. §2 presents some concrete examples of palatalization, followed by an overview of cross-linguistic patterns of the phenomenon. §3 examines how the feature system of the early generative phonology captures natural classes involved in palatalization. §4 focuses on two key approaches to palatalization within the feature geometry framework. §5 turns to treatments of palatalization within Optimality Theory, and §6 concludes with a review of some alternative proposals.

2 Background

2.1 *Some examples*

We will begin with some relatively well-known examples of palatalization. English has at least three kinds of alternations that fall under the general definition of palatalization processes. *Coronal palatalization* involves an alternation between alveolars [t d s z] and palato-alveolars [ʧ ʤ ʃ ʒ], as shown in (1). In these examples, the palato-alveolars occur before a palatal glide (in an unstressed syllable), while alveolars occur elsewhere. These alternations can be analyzed as a process – a change of alveolars to palato-alveolars in the context of [j] (Chomsky and Halle 1968; Borowsky 1986; among others).[1] The process is assimilatory in the sense that the consonants targeted by palatalization become more similar in place of articulation to the segment that triggers palatalization. Note that stops not only shift in place, but also undergo assimilation (becoming sibilant affricates). All the other features of the target consonants (e.g. continuancy, voicing, etc.) remain unchanged.

(1)	t – ʧ	*perpe*[t]*uity*	*perpe*[ʧ]*ual*
	d – ʤ	*resi*[d]*ue*	*resi*[ʤ]*ual*
	s – ʃ	*gra*[s]*e*	*gra*[ʃ]*ious*
	z – ʒ	*plea*[z]*e*	*plea*[ʒ]*ure*

The second process – *velar softening* – is exhibited by alternations between velar stops [k g] and coronal fricatives or affricates [s] and [ʤ], respectively. The coronal alternants are found before certain Latinate or Greek suffixes beginning with (mainly) front vowels; the velar alternants are found elsewhere (2). Given this, the alternations are usually analyzed as a palatalizing change of velars to coronals triggered by front vowels (Chomsky and Halle 1968; Borowsky 1986). Unlike coronal palatalization, this process is more complex, as it actually involves two non-identical changes – a shift of the voiced velar stop to the palato-alveolar affricate and a shift of the voiceless velar stop to the alveolar fricative. While the outputs of velar softening are not identical in terms of minor place of articulation and continuancy, they are both sibilant coronals.

(2)	a.	k – s	*medi*[k]*ation*	*medi*[s]*ine*
			criti[k]	*criti*[s]*ize*
	b.	g – ʤ	*analo*[g]	*analo*[ʤ]*y*
			pedago[g]*ue*	*pedago*[ʤ]*y*

The third process – *spirantization* – exhibits alternations between the alveolar stop [t] and the alveolar fricative [s] (or [ʃ] in conjunction with coronal palatalization). The latter segment occurs before suffixes with an unsyllabified /i/ (3), and this process is assumed to involve a change of stop to fricative before high front vowel (Borowsky 1986). As such, the process does not involve a change in place

[1] Similar, albeit optional and phonetically gradient, alternations are also exhibited across words, as in *go*[ʧ] *you*, *plea*[ʒ]*e yourself*, etc. (Zsiga 1993).

of articulation, but a change in continuancy and sibilancy (CHAPTER 28: THE REPRESENTATION OF FRICATIVES).

(3)	t – s	*secre*[t]	*secre*[s]*y*
		regen[t]	*regen*[s]*y*
		emergen[t]	*emergen*[s]*y*
		par[t]	*par*[ʃ]*ial*

The three palatalization processes manifested by the alternations in (1)–(3) differ in several respects. The targets of palatalization are anterior coronals (alveolars) in (1) and (3), and dorsals in (2). The outputs are posterior coronals (palato-alveolars) in (1) and (2b), and anterior coronals in (2a) and (3). The triggers are /j/ in (1), and high front vowels in (2) and (3). (The processes are also different in terms of their phonological or morphological conditioning: morpheme boundaries, particular suffixes, stress, etc.) What the processes have in common, however, is that they appear to be triggered by front vocoids and result in coronal segments, notably, all sibilants. As such, they are representative of the three general types of palatalization processes defined by Bhat (1978): "coronal raising" (1), "velar fronting" (2), and "spirantization," which may or may not be accompanied by change in place ((2) and (3) respectively), as discussed in the next section.

Another important type of palatalization, not exhibited by the English processes, is an addition of secondary palatal articulation, without a change in primary place or assibilation. As shown in (4a), Russian exhibits alternations between non-palatalized consonants of all places – labials, anterior coronals (dentals), and dorsals –and their palatalized counterparts (CHAPTER 121: SLAVIC PALATALIZATION). The palatalized segments in such alternations occur before front vowels (/e/ in (4a)), while non-palatalized consonants are unrestricted (Kenstowicz and Kisseberth 1979). These alternations can be straightforwardly analyzed as an assimilatory process involving a simple addition of secondary palatal articulation (the high front position of the tongue body) before front vowels. This process in Russian is fairly general, and is not restricted to particular morphological categories. Secondary palatalization may co-occur in a given language with "place-changing" palatalization. In Russian, non-palatalized dentals and velars also exhibit alternations with palato-alveolars, with the latter occurring before certain verbal suffixes (4b) (Lightner 1965; Kenstowicz and Kisseberth 1979). Note that the palatalizing suffixes may or may not begin with overt front vocoids, showing that place-changing palatalization in Russian is a more opaque, morphologically conditioned process. Note also that the relation between targets and outputs of palatalization in (4b) is less transparent than in (4a): for example, the same palato-alveolar output [ʧ] can result from two different target consonants, /t/ or /k/.

(4)	a.		*nom sg*	*dat sg*		
		$p – p^j$	trap-a	$trap^j$-e	'path'	
		$t – t^j$	s^jirat-a	$s^j irat^j$-e	'orphan'	
		$k – k^j$	sabak-a	$sabak^j$-e	'dog'	
	b.		*inf*	*3 pers sg*	*1 pers sg*	
		t – ʧ	$pr^j at$-at^j	pr^jaʧ-it	pr^jaʧ-u	'hide'
		k – ʧ	plak-at^j	plaʧ-it	plaʧ-u	'weep'

Palatalization in Russian produces outputs that are phonemic, since the language has palatalized consonants and posterior coronals whose occurrence is not conditioned by front vowels. The same can be said about the English palatalization processes. Allophonic palatalization is also quite common, however. In Nupe (5), for example, the velars have secondary palatal articulation before front vowels (5a), secondary labial articulation before round vowels (5b), and no secondary articulation before the back unrounded /a/ (5c). This pattern can be analyzed as an allophonic assimilatory process involving palatalization and labialization of phonemic plain velars before front and round vowels respectively (Sagey 1990; Archangeli and Pulleyblank 1994).

(5)	a.	/egi/	egʲi	'child'
		/ege/	egʲe	'beer'
	b.	/egũ/	egʷũ	'mud'
		/ego/	egʷo	'grass'
	c.	/ega/	ega	'stranger'

So far we have examined fairly representative examples of palatalization, those involving changes that are common across world languages. It is worth contrasting these examples with that from two Southern Bantu languages, Tswana and Swati, shown in (6). Here, labials alternate with (labialized) palato-alveolars in the context of the passive suffix /-wa/. Consonants of other places remain unaffected (cf. Tswana [ratʷa] /rat-wa/ 'love'; Swati [pʰegʷa] /pʰeg-wa/ 'cook'). These alternations are also considered to manifest palatalization (Halle 2005; Bateman 2007); however, the process is different from the cases above in several important respects. First, the targets are labial consonants to the exclusion of the other places, and the labials change their place of articulation, something that is usually restricted to coronals and dorsals. (Coronals, however, do undergo palatalization in other contexts.) Second, the trigger of the process, the suffix /-wa/, does not contain an overt or even an underlying front vocoid, but presumably develops it as a result of labial dissimilation (Kotzé and Zerbian 2008). Third, the process in Swati is not strictly local, as it can target labials occurring in non-adjacent syllables. All this makes palatalization in Tswana and Swati (and similar processes in other Southern Bantu languages) a relatively "unnatural" case in the typology of palatalization, as we will see below.

(6)		*non-passive*	*passive*		
a.	*Tswana*				
	p – ʧʷ	lɔpa	lɔːʧʷa	/lɔp-wa/	'request'
	pʰ – ʧʰʷ	tlʰʊpʰa	tlʰʊːʧʰʷa	/tlʰʊpʰ-wa/	'choose'
	b – ʤʷ	roba	rɔːʤʷa	/rob-wa/	'break'
b.	*Swati*				
	b – ʤ	hamb-	hanʤʷa	/hamb-wa/	'go'
	ɓ – ʧʼ	seɓenta	seʧʼentʷa	/seɓent-wa/	'work'
	pʰ – ʃ	sipʰula	siʃulʷa	/sipʰul-wa/	'uproot'

The examples from English, Russian, Nupe, Tswana, and Swati provide a snapshot of a vast range of variation found in palatalization processes both within and across languages. Of particular interest here are the variation and preferences in terms of featural composition of segmental classes involved in palatalization – its triggers, targets, and outputs. There are clearly many other theoretically important issues relevant to palatalization – including those of allophonic/phonemic status, morphological or lexical conditioning, locality, etc.; these, however, are beyond the scope of this chapter (on some of these topics, see CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION; CHAPTER 81: LOCAL ASSIMILATION; CHAPTER 88: DERIVED ENVIRONMENT EFFECTS; CHAPTER 121: SLAVIC PALATALIZATION).

2.2 *Typological patterns of palatalization*

To better understand the complexity of the phenomenon, it is useful to examine cross-linguistically more and less common patterns of palatalization. The following discussion is based on the author's survey of synchronic palatalization processes, with some reference to the earlier often-cited survey by Bhat (1978), and the more recent one by Bateman (2007).[2] (See also Chen 1973, Kochetov 2002, and Stadnik 2002 for surveys specific to certain geographic areas or phonological contrasts.)

The survey covers cases of palatalization as exhibited by segmental alternations (as opposed to phonotactic restrictions or historical changes) selected from theoretical phonological literature for the purposes of this chapter. Altogether, it contains data from 64 languages and dialects belonging to 17 language families and 25 genera. We will begin with observations about targets and outcomes of palatalization, and corresponding general types of palatalization processes. The focus will be on changes targeting labial, (anterior) coronal, and dorsal stops. Table 71.1 represents three general processes of palatalization: secondary palatalization

Table 71.1 Targets and outputs of palatalization (alternations only) and corresponding processes, and their relative frequency in world languages (based on numbers of language families and genera, given in square brackets; see the text for details)

Type	*Palatalization*		*labial*	*coronal*	*dorsal*
I	Secondary		$p \rightarrow p^j$ common [6, 9]	$t \rightarrow t^j$ common [6, 8]	$k \rightarrow k^j$ common [6, 7]
II	To a posterior coronal	a. to a non-sibilant	p → c rare [1, 1]	t → c common [7, 8]	k → c common [4, 6]
		b. to a sibilant	p → tʃ rare [1, 1]	t → tʃ common [9, 14]	k → tʃ common [4, 7]
III	To an anterior coronal	a. to a non-sibilant	p → t absent [0, 0]	n/a	k → t absent [0, 0]
		b. to a sibilant	p → ts rare [1, 1]	t → ts common [3, 6]	k → ts rare [2, 4]

[2] The latter two studies have certain limitations with respect to the goals of this chapter. Although quite extensive, Bhat's survey does not clearly distinguish between synchronic processes and historical sound changes, of which only the former are relevant here. Bateman's survey, while drawing on a genetically balanced language sample and focusing on synchronic processes, is restricted to only certain types of palatalization processes, leaving out, for example, place-changing palatalization of labials and changes resulting in anterior coronals. The latter types are important for our discussion, as these are the ones that have been most problematic for theoretical accounts.

(Type I), palatalization resulting in a posterior coronal (palato-alveolar or (alveolo-)palatal; Type II), and palatalization resulting in an anterior coronal (alveolar or dental; Type III). The Type II process can produce either non-sibilants or sibilants, resulting in two subtypes ((a), (b)). The same subdivision is given for the Type III process. Columns on the right schematically present typical (or expected) changes involved in each process, depending on the target consonant – labial, coronal, or dorsal. (To avoid cluttering the table, voiceless stops/ affricates stand here for segments regardless of their laryngeal specification, e.g. "ʧ" can include [ʧ˺], [ʧʼ], [ʧʰ], or [ʤ]; changes in continuancy are not noted, e.g. "k → ts" includes the changes to [ts] or [s].) To facilitate the comparison, each change is labeled as "common," "rare," or "absent," indicating its relative frequency in the sample, based on numbers of separate language families and genera (given in parentheses), rather than individual languages.[3] (See the online version of this chapter for examples of the processes.)

Note that the typology of processes in Table 71.1 is compatible with the typology of changes in Bhat (1978), whose terms are based on *SPE* (Chomsky and Halle 1968) feature terminology (see §3.1). Taking coronal targets as an example, Type I and Type IIa correspond to Bhat's process of alveolar "raising" without "spirantization" (i.e. [−high] → [+high] in *SPE* notation), Type IIb corresponds to "raising" accompanied by "spirantization" (i.e. [−high, −strident] → [+high, +strident]), and Type IIIb corresponds to "spirantization" without "raising" (i.e. [−strident] → [+strident]). Bhat's terms are not fully appropriate for our purposes, as they do not distinguish between secondary and place-changing palatalization, in addition to being tied to a specific feature framework. Place-changing processes (Types II and III) involving non-coronals have also been referred to as "coronalization" (Hume 1992; Flemming 2002).

What is interesting about the different types of palatalization shown in Table 71.1 is that certain targets and outputs (and the corresponding processes) are common, while others are rare or unattested. Overall, there is a tendency for place-changing palatalization to result in sibilants rather than non-sibilants. While both sibilants and non-sibilants are possible outputs for Type II palatalization, only sibilants are possible for Type III palatalization. Another important observation is that labials as targets of place-changing palatalization processes (Types II and III) are exceedingly rare, compared to coronals and dorsals. The only examples of labial place-changing palatalization (with stops as targets) in the sample are Southern Bantu languages (see (6)) and Moldova Romanian (e.g. [plop] 'poplar tree', [ploc] /plop-i/ (PLURAL); [drob] 'block (of salt)', [droɟ] /drob-i/ (PLURAL); Bateman 2007). Among the other two place categories, coronals as targets tend to occur overall more frequently than dorsals. Notably, the most common palatalization process is a change of alveolars to palato-alveolars (Type IIb), attested in nine language families and 14 genera. Further examination of the cases (Table 71.2) shows that in a given language, coronals and dorsals can be targeted

[3] As an example, Type IIb coronal palatalization in English (1) and Russian (4b) represents a single case of palatalization at the level of family (Indo-European), and two separate cases at the level of genus (Germanic and Slavic). This allows for an estimation of cross-linguistic frequency that is relatively conservative and less biased toward certain language families or genera. For expository reasons, "common" changes are defined as occurring in at least three or more families, and "rare" changes occur in one or two families. The study uses the language classification from Haspelmath *et al.* (2005).

Table 71.2 Targets of palatalization and their relative frequency in world languages

Target consonants	*Occurrence*
coronal only	common [13, 16]
dorsal only	common [4, 6]
labial only	absent [0, 0]
coronal and dorsal	common [3, 5]
coronal and labial	rare [1, 1]
dorsal and labial	absent [0, 0]
coronal, dorsal, and labial	common [6, 9]

by palatalization independently or together, while labials are targeted only when coronals (and dorsals) are targeted too (but see (6)). This suggests implicational relations among targets of palatalization, with place-changing palatalization of labials implying palatalization of coronals and dorsals (cf. Chen 1973; Foley 1977; Bhat 1978). The results of the survey are also indicative of a greater propensity of coronals to palatalization, compared to dorsals. This is consistent with some previous studies (cf. Bateman 2007), while in part contradicting those based on more limited language samples or mixed synchronic/diachronic data (Chen 1973; Foley 1977; see §3.2).

Turning to triggers of palatalization, Table 71.3 shows that these include front vocoids (vowels and glides) that differ in height and high vocoids that differ in backness. Non-high back vowels do not trigger palatalization. Among all the triggers, high front /i/ and /j/ are the most likely triggers, followed at a considerable distance by mid front vowels. Recall that these are the triggers in all examples shown above (leaving aside the more opaque Bantu cases). Examples of the rare types of palatalization triggered by front low vowels or high back vocoids include Slovak ([vnuːʧa] /vnuːk-æ/ 'grandson-DIM'; Hume 1992) and Lomongo ([konʤwá] /kond-wa/ 'cover with sand-PASS'; Kenstowicz and Kisseberth 1977). It is important to note that in a given language, low and mid front vowels

Table 71.3 Triggers of palatalization and their relative frequency in world languages

Trigger Backness	Height		*Occurrence*
front only	high only	i/j	common [17, 24]
front only	high and mid	i/j, e/ɛ	common [4, 5]
front only	high, mid, and low	i/j, e/ɛ, æ	rare [1, 1]
front only	mid/low	e/ɛ/æ	absent [0, 0]
front and back	high only	i/j, i/u/w	rare [1, 1]
back only	high only	i/ɯ/ɰ	absent [0, 0]
back	mid/low	ɑ/ə/o/ɔ	absent [0, 0]

apparently only trigger palatalization if high front vowels trigger it too.[4] Similarly, non-front high vocoids are triggers when front high vocoids are also triggers. These observations are indicative of implicational relations between vowel height/frontness and the ability to trigger palatalization (cf. Chen 1973; Foley 1977; Bateman 2007). Further, there are some interesting dependencies between triggers and targets. Coronals tend to be targeted by high vocoids, and especially by /j/, while dorsals are almost exclusively targeted by /i/ and other front vowels (see Bhat 1978: 52–56 for details).

In terms of directionality, palatalization processes can be regressive (right-to-left) or progressive (left-to-right). Both types are quite common, with regressive palatalization attested in eight families and sixteen genera, and progressive palatalization in nine families and nine genera (mainly in the Americas). Some languages show both types, although in different morphological/phonological contexts, as, for example, Chimalapa Zoque (7a) (Kenstowicz and Kisseberth 1977). The overwhelming majority of palatalization processes are local, triggered by immediately adjacent vocoids. In a few cases, however, palatalization can apply across a consonant, as in Barrow Inupiaq (7b) (Archangeli and Pulleyblank 1994), or across one or more syllables as in Harari (7c) (Rose 1997; cf. Swati in (6)).

(7) a. *Chimalapa Zoque*

tit̠ʲi	/tits-i/	'dry'	cf.	tɨts-pa	'is drying'
kuj ʧets-pa	/kuj tsets-pa/	'is carving wood'		tsets-pa	'is carving'

b. *Barrow Inupiaq*

isiẋsuq	/isiq-tuq/	'be smoky (3SG INT)'
isiʁʎuni	/isiq-luni/	(3SG IRREALIS)

c. *Harari*

kiʧəbi	/kitəb-i/	'write! (2SG FEM)'	cf. kitəb (2SG MASC)

2.3 *Summary*

Our examination of particular cases and cross-linguistic patterns of palatalization processes reveals a number of asymmetries involving targets, triggers, and outputs. Some of the key observations are restated in (8). First, certain places of articulation are better targets of palatalization than others (8a). Among the three places, the key difference is between coronals and dorsals on one hand and labials on the other. Another important difference is between coronals and non-coronals. Second, certain vowels/glides are better triggers of palatalization than others (8b). In particular, front vowels are considerably better triggers than non-front vowels, and among the former class, high vowels are the best triggers, and low vowels the worst. Third, there are important dependencies between the vowel height (and syllabicity) of triggers and the place of targets (8c). Fourth, outputs of palatalization are either palatalized consonants or coronals; among the latter, posterior coronals and/or sibilants are the preferred outputs (8d). In addition, all but a few cases of palatalization are local, triggered by immediately preceding or following vocoids.

[4] Bhat (1978) mentions some cases where mid front vowels palatalize velars to the exclusion of high front vowels. None of these cases, however, appears to involve alternations.

(8) *Main typological generalizations about palatalization*
(">" = "better than, more likely than")

a. *Target place asymmetry (place-changing palatalization)*
   i. coronal, dorsal > labial
   ii. coronal > dorsal, labial

b. *Trigger asymmetry*
   i. (high) front > high central/back
   ii. high front > mid front > low front

c. *Trigger–target dependencies*
   i. front vowels and dorsals
   ii. high vocoids (especially /j/) and coronals

d. *Output asymmetries*
   i. posterior > anterior
   ii. sibilant > non-sibilant

It should be noted that as a type of consonant–vowel interaction, palatalization is somewhat unique (see CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). Although other consonant-to-vowel assimilation processes commonly result in consonants with secondary vowel articulations (t → t^w / __ u), they hardly ever exhibit synchronic shifts in primary place of articulation (t → p or k / __ u) (cf. Ní Chiosáin and Padgett 1993). If such alternations are indeed observed, they appear to imply palatalization alternations, as in Ikalanga, where vowels (through gliding) trigger both place-changing velarization and palatalization (9) (Mathangwane 1996). In addition, unlike palatalization, hardly any other vowel–consonant interactions affect consonant manner features, in a way that, for example, produces sibilant affricates or fricatives from stops. Equally implausible are the processes that would produce the reverse effect of place-changing palatalization, for example, converting sibilant coronals to coronal or non-coronal stops (ʧ → t or k). All this underscores the seemingly unique place of palatalization in the typology of consonant–vowel interactions, and its highly asymmetric nature.

(9)

	plain	*diminutive*		
a.	ʒínó	ʒìŋʷáná	/ʒino-ana/	'tooth'
	ʃamú	ʃaŋʷáná	/ʃamu-ana/	'lash'
b.	báni	baɲáná	/bani-ana/	'forest'
	semé	seɲáná	/seme-ana/	'basket'

What makes palatalization so special? Why are some patterns of palatalization cross-linguistically more common, while other patterns are rare or unattested? It has long been known that the naturalness of many palatalization processes has its roots in phonetics – articulation and perception. As Hyman (1975: 171) noted in his discussion of velar palatalization, gradient fronting of a [k] before [i] is a phonetic process that is universal, shared by all languages. The two articulatory gestures – the tongue body backing for [k] and the tongue body fronting for [i] – simply cannot be co-produced without this co-articulatory adjustment. In this sense, the process is automatic, part of the "universal phonetics" (although the degree of velar fronting can be language-particular). Further, fronted velars or palatals tend to be produced with greater frication at the release, which makes them acoustically more similar to palato-alveolar affricates. Given this acoustic similarity, the former are often auditorily confused with the latter (but not the reverse), resulting in common historical shifts of velars to palato-alveolars (Guion 1996). The change [k] → [ʧ] before [i] is therefore motivated by both articulation and perception. Similar articulatory, and possibly perceptual, reasons underlie the change of [t] → [ʧ] before [j] or [i] – presumably arising due to overlap of the tongue tip and tongue body gestures, producing a more retracted laminal constriction with a turbulent sibilant-like release (Zsiga 1993). In contrast, the articulation of [p] before [j] or [i] presents no articulatory difficulties, as the two gestures – the lips and the tongue body – are physically uncoupled and therefore can be freely co-produced. Despite some frication at the release, [pʲ] is still quite acoustically different from [ʧ] and is thus less likely to be confused with the former. This suggests that unlike dorsal and coronal palatalization, labial palatalization is phonetically much less plausible, and therefore phonologically less natural. Indeed, comparative historical evidence suggests that cases of labial palatalization have arisen through "telescoping" – a series of historical changes involving glide strengthening and cluster simplification (Hyman 1975; Bateman 2007). In fact, different stages of these developments are often reflected in closely related languages or dialects, as the case with Tswana and Moldova Romanian (10) (Udler 1976; Kotzé and Zerbian 2008). Finally, the lack of phonetic motivation can explain some of the asymmetries in triggers of palatalization (e.g. high front vocoids *vs.* low and back) and the unnaturalness of changes reverse to palatalization (e.g. [t̢ʃ] → [t] or [k]).

(10)	a.	*Tswana*		*Northern Sotho*	*Lobedu*	
		-gaʧʷa	/-gap-wa/	-gapʃa	-habja	'request (PASS)'
	b.	*Moldova Romanian dialects*				
		Standard		*Northern Bukovina*	*Chernovcy*	
		aric	/arip-i/	aripc	aripʲ	'wing (PL)'

While the diachronic phonetic sources of palatalization have rarely been debated, most phonologists would agree that at least some of the common patterns and important asymmetries in palatalization in (9) (or any phonological process) require synchronic explanation (but see §6). Further, regardless of historical changes, it is commonly agreed that synchronic grammars should have ways of modeling palatalization alternations (as in English or Russian) or allophonic variation (as in Nupe). Yet the question of how to represent the process synchronically while capturing relevant significant generalizations has proven to be difficult, if not impossible. It is remarkable that almost 40 years after the first generative account of English velar palatalization in Chomsky and Halle (1968), Halle (2005: 23) concedes that "to this time there has been no proper account of palatalization that would relate it to the other properties of language, in particular, to the fact that it is found most commonly before front vowels." This is despite the fact that palatalization has received extensive treatment in early generative phonology, autosegmental phonology, and more recently Optimality Theory. The goal of this chapter is to review some of the influential theoretical treatments of palatalization as a synchronic process, while focusing particularly on distinctive features and feature geometry representations as ways of capturing the naturalness of common palatalization processes. As we will see, some of the problems encountered by formal models of palatalization can be attributed to the complexity of the phenomenon; other difficulties, however, seemingly stem from the reliance on a universally fixed, closed set of rigid, unidimensional representations. We will also examine other formal ways of capturing relevant generalizations using constraints and constraint hierarchies or more phonetically detailed representations in Optimality Theory, and conclude with a brief review of some recent alternative proposals that challenge traditional generative assumptions.

3 Palatalization in early generative phonology

3.1 *Distinctive features and marking conventions of* SPE

One possible way of capturing the naturalness of phonological processes is through stating natural classes of segments involved (as triggers, targets, and outputs), using distinctive features (CHAPTER 17: DISTINCTIVE FEATURES). The concept of natural classes encoded by a universal set of features has been an important part of generative phonology since Chomsky and Halle (1968; *SPE*). The distinctive features in *SPE* were exclusively articulatorily based (unlike the auditorily based features of Jakobson *et al.* 1952). One proposal that has important theoretical consequences for our discussion is the use of features [±high], [±back], and [±low] for both vowels and consonants. Among the latter, these features are used as "a natural manner to characterize subsidiary consonant articulations such as palatalization, velarization, and pharyngealization" (*SPE*: 305), which are defined as [−back, −low], [+back, −low], and [+back, +low], respectively. This proposal was intended to capture the fact that secondary articulations tend to occur before vowels of the same qualities, for example, palatalized consonants before front vowels (cf. Nupe (5)). The feature specification thus allowed one to state these restrictions as "an obvious case of regressive assimilation" (*SPE*: 308). In addition, the proposal captured a cross-linguistic observation that the three types of secondary articulations are mutually exclusive, since, for example, palatalized consonants cannot be simultaneously velarized or pharyngealized. These feature specifications also helped in the formulation of typical vowel raising and fronting changes in the environment of palatalized consonants as a simple case of assimilation. An example from Russian is shown in (11a), where underlying vowels /e/ and /a/ shift to [i] when occurring after a palatalized consonant in an unstressed syllable (Kenstowicz and Kisseberth 1979). An *SPE*-style rule capturing the process is stated in (11b).

(11)	a.	*1st plural*	*1st singular*	
		ˈpʲiʃim	pʲiˈʃu	'write'
		ˈmʲetʃim	mʲiˈtʃu	'throw'
		ˈvʲaʒim	vʲiˈʒu	'bind'
		ˈmaʃim	maˈʃu	'wave'
	b.	[+syll, −high] → [+high, −back] / [−syll, +high, −back] __		

Note that in this respect the *SPE* feature system is a step forward compared to Jakobson *et al.*'s (1952) system, where consonants with secondary articulations and corresponding vowels did not share the same feature values. For example, palatalized consonants were specified for [+sharp] (and [+grave] or [−grave]), while (high) front vowels were [−sharp] and [−grave].

The *SPE* specification for [+high, −back] was not limited to high front vowels and palatalized consonants, but also extended to palatals and postalveolars. This is important, because it naturally grouped together common triggers and outputs of palatalization processes (Types I and II). The feature system, however, treated palatals as non-coronals and grouped them with velars: both are [−coronal, −anterior, +high] and differ in [±back]. Notably, palatalized velars were not distinguished from palatals (both are [−coronal, −anterior, +high, −back]). These two specifications created certain problems. First, Type I (secondary) palatalization was represented as two different featural processes: a change from [−high] to [+high] (raising) for coronals and labials (which are [±coronal, +anterior, −high, −back]), and a change from [+back] to [−back] (fronting) for velars. This also predicted – partly incorrectly – that palatalization of velars should be triggered by front vowels only (regardless of height), while secondary palatalization of coronals and labials should be triggered by high vowels only (regardless of frontness/backness).

Another seemingly non-trivial problem is revealed by the treatment of place-changing palatalization processes (Types II and III). Consider the rules proposed to account for English coronal palatalization (12a) and velar softening (12b) (see (1) and (2)). As Chomsky and Halle note (*SPE*: 424), palatalization is an intrinsically assimilatory process. Nothing in the rules below, however, captures its assimilatory nature. In fact, specifications of triggers and outputs in each of the rules do not share a single feature.

(12) a.
$$\begin{bmatrix}\text{−son}\\ \text{+cor}\end{bmatrix} \rightarrow \begin{bmatrix}\text{−ant}\\ \text{−strid}\end{bmatrix} / \_\_ \begin{bmatrix}\text{−back}\\ \text{−voc}\\ \text{−cons}\end{bmatrix}\begin{bmatrix}\text{−cons}\\ \text{−stress}\end{bmatrix}$$

b.
$$\begin{bmatrix}\text{−cont}\\ \text{−ant}\\ \langle\text{−voice}\rangle\end{bmatrix} \rightarrow \begin{bmatrix}\text{+cor}\\ \text{+strid}\\ \langle\text{+ant}\rangle\end{bmatrix} / \_\_ \begin{bmatrix}\text{−back}\\ \text{−low}\\ \text{−cons}\end{bmatrix}$$

Acknowledging this and other problems arising from the excessively permissive rule notation mechanism as a "fundamental theoretical inadequacy" (*SPE*: 400), Chomsky and Halle propose to supplement rules and feature specification with a substantive component – a theory of markedness consisting of a list of "marking conventions." They illustrate the application of these conventions in rules representing historical palatalization processes in Slavic. The so-called "first velar palatalization" in Slavic (Type IIa; [k g x] → [ʧ ʤ ʃ]) can be stated as a "simple assimilation rule" (*SPE*: 400) by which velars ([−anterior]) acquire the [−back] value from following front vowels (13a). The change of stops to strident coronal affricates and fricatives ([+coronal, +delayed release, +strident]) is not an assimilatory effect, but is due to an application of relevant marking conventions (13b). According to these conventions, a postalveolar affricate [ʧ] is less marked than a palatal stop [c] or a palato-alveolar stop [ṯ], and therefore "when velar obstruents are fronted, it is simpler for them also to become strident palato-alveolars with delayed release" (*SPE*: 423). Thus, the unmarked value of the feature [coronal] ([ucoronal]) for [−back, −anterior] consonants is [+coronal], and the unmarked values for the other two features of posterior coronals are [+delayed release] and [+strident] (CHAPTER 12: CORONALS; CHAPTER 22: CONSONANTAL PLACE OF

ARTICULATION). Similar assimilation rules and marking conventions were proposed for the Slavic "second velar palatalization" and language-specific realizations of "dental palatalization" (Types IIa, IIb, and IIIb).

(13) a. [−ant] → [−back, +cor, +del rel, +strid] /__ [−cons, −back]
b. [*u*cor] → [+cor] / [__, −back, −ant]
[*u*del rel] → [+del rel] / [__, −ant, +cor]
[*u*strident] → [+strident] / [__, +del rel, +cor]

The combined use of rules and marking conventions made it possible to formulate palatalization as an assimilatory process. Yet it remains unclear when marking conventions should be invoked in general, and there are questions about the appropriateness of some of the conventions for particular cases of palatalization. For example, while it is true that postalveolar affricates are less marked (at least less cross-linguistically common) than palatal or postalveolar stops, the same is true, even more so, about the unmarked status of alveolar or labial stops (both [−anterior, −back]) – the segments that are never produced by palatalization. Further, the account views all place-changing palatalization processes as consisting of two consecutive stages – the fronting or secondary palatalization followed by simplification (e.g. [k] → [c] → [ʧ]; [t] → [tʲ] → [ʧ]). While these stages may correctly recapitulate the historical development of some palatalization processes (as in Slavic), it can be argued that they are simply unnecessary as statements of the synchronic rules of a language.

3.2 Naturalness and phonological rules

Questions about alternative ways of constraining the excessively powerful rule machinery of *SPE* were central to the theoretical debate in the late 1960s and the 1970s (see Hyman 1975 for a review). Why, for example, is the palatalization rule in (14a) cross-linguistically common and natural, while the exact reverse of it (14b) is highly unlikely and unnatural? From the point of view of computational simplicity, both rules are equally simple, involving the same number of features. The fact that the formal theory had no way of distinguishing between natural and unnatural rules was seen by some phonologists as highly problematic. In response to this, Schachter (1969) proposed to encode naturalness directly into phonological rules, introducing the feature specification *n*, marking feature values that are "natural" for a given process. Given this, the rule of velar palatalization can be rewritten as (14c), stating that the natural value of the feature [±back] before front vowels is [−back]. Features marked as natural are not counted by the rule simplicity metric, thus rendering the rule in (14c) less "costly" than the rule in (14b). Taking this idea further, Chen (1973) proposed to formalize the target place and trigger height asymmetries of palatalization (8a) and (8b.ii) as part of special meta-rules – language-specific rules supplemented with universal constraint statements. For example, his meta-rule in (14d) states that consonants become palatalized before front vowels ([1 back] = {i, e, æ}), however, with certain implicational relations: (i) if a consonant of a given point *m* along the backness scale undergoes palatalization, so does the consonant higher on the scale (i.e. [p] implies [t] and [k], and [t] implies [k]; cf. (8a), but see (8b)); (ii) if a consonant undergoes palatalization by a vowel of a given point *n* of the height scale, it also does so before any vowel higher on that scale (i.e. [æ] implies [e] and [i], and [e] implies [i]; cf. (8b.ii)).

In a related proposal, Foley (1977) formulated "synchronic truth statements" – implicational relations among triggers and targets of palatalization – and provided detailed calculations of relative probability of palatalization depending on the target place and trigger height (among other factors), as shown in (14e).

(14) a. C[+back] → [−back] / __ V[−back] (k → ʧ / __ [i, e, æ])
b. C[−back] → [+back] / __ V[+back] (ʧ → k /__ {u o a})
c. C[+back] → [*n*back] / __ V[−back]
d. C[αback] → palatalized / V[1back, βhigh]
language-universal constraints: $\alpha \geq m$, $\beta \geq n$, where
C backness scale: 1 [p], 2 [t], 3 [k]; V backness scale: 1 [i e æ], 2 [u o ɑ];
V height scale: 1 [æ], 2 [e], 3 [i]
e. Relative probability scale of palatalization
kj > ki, tj > ke, ti, pj > kæ, te, pi > tæ, pe > pæ

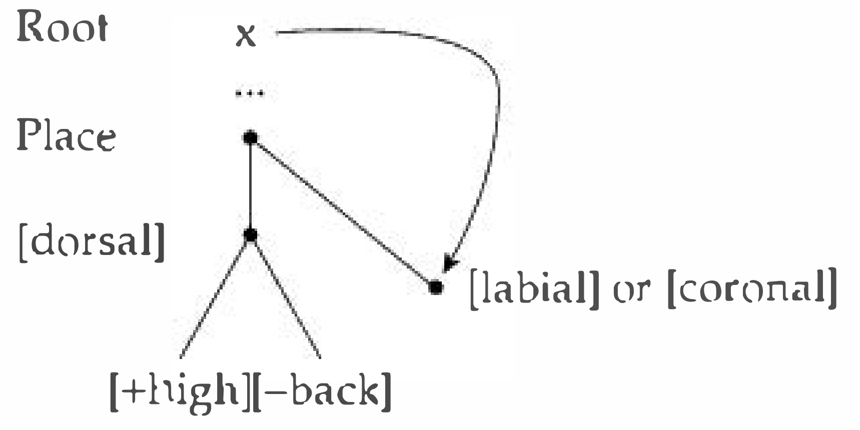
These proposals, despite some empirical inadequacies, are interesting as the first relatively systematic attempts to directly incorporate substantive factors into the formal computational mechanism. The use of phonetic naturalness as a formal phonological criterion, however, did not receive much support in mainstream generative phonology at the time, as it was difficult to reconcile with the fact that languages allow both natural and unnatural rules, seemingly without preference for the former. If naturalness considerations were part of the grammar, why would some languages maintain such phonetically implausible rules as labial palatalization (cf. Hyman 1975 on the Bantu rule [p] → [s] / __ [i])?

4 Feature geometry

New ways of constraining the application of phonological rules were brought by the framework of feature geometry (Clements 1985; Sagey 1986; among others; see also CHAPTER 27: THE ORGANIZATION OF FEATURES). More elaborate, geometrically organized autosegmental featural representations were expected to delimit the typology of phonological rules, distinguishing between possible, natural and impossible, unnatural processes. Within feature geometry, it is useful to distinguish two main approaches to palatalization. Both view palatalization as an assimilatory phenomenon, but differ in the feature specification of the main triggers of the process – front vocoids. The first approach treats front vocoids as [dorsal], essentially following the *SPE* tradition. The second approach specifies front vowels as [coronal], in an attempt to state some of the generalizations missed by the *SPE*-style featural accounts. Palatalization is thus modeled as either spreading [dorsal] or spreading [coronal].

4.1 *Palatalization as spreading* [dorsal]

One key proposal of this approach, initially developed in Sagey's dissertation (1986), is that vowels, glides, dorsal consonants, and secondary articulations like palatalization and velarization are characterized by the [dorsal] node with features [±high, ±back, ±low]. In contrast to *SPE*, labials and coronals are not specified for these features, but are characterized by [labial] and [coronal] nodes respectively. The feature [anterior] in the new approach is limited to coronals only, being specified under the [coronal] node. Palatalized consonants in this system are viewed as complex segments with temporally unordered primary [labial] or [coronal] nodes and the secondary [dorsal] node specified for [+high, −back] (with "designated articulators" – primary nodes – marked diacritically with a pointer from the Root node):

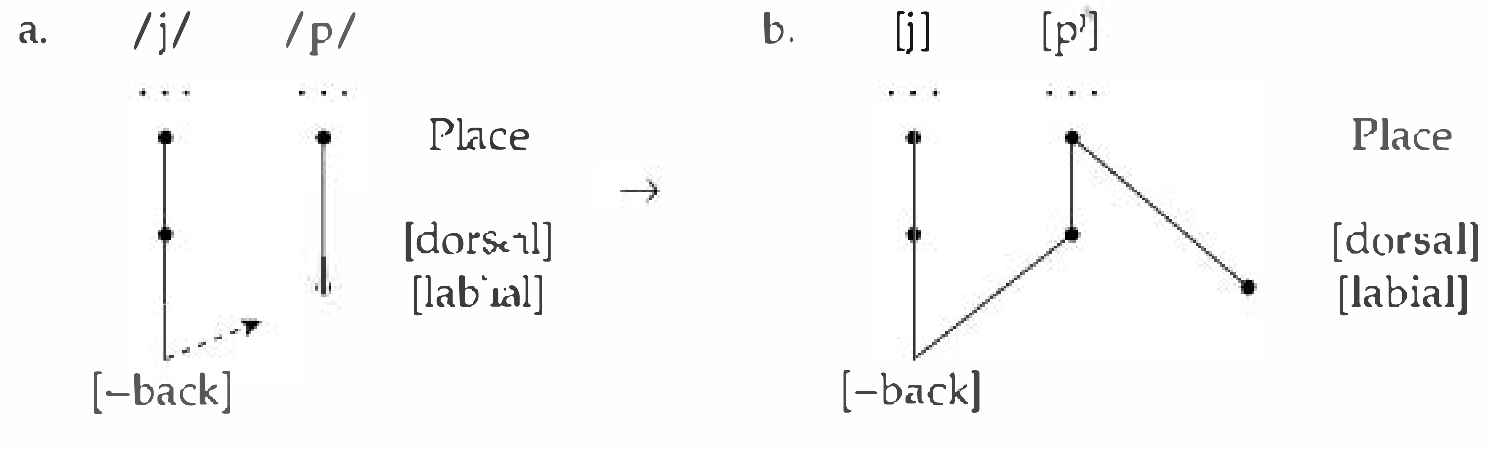
(15)


As in *SPE*, this system does not allow for palatalized velars distinct from palatals: both are designated as [kʲ], [gʲ], etc. and represented as simple segments having Place[dorsal[+high, −back]].

Given these feature specifications, palatalization processes are treated as spreading [+high, −back], either with or without the [dorsal] node. Sagey presents data from Zoque to illustrate the application of palatalization. As shown in (16), palatal glide /j/ triggers palatalization of following consonants by adding a secondary palatal articulation to labials and velars, and changing anterior coronals to posterior coronals ([t̠] = a [−anterior] alveolo-palatal stop).

(16)	a.	/j-pata/	pʲata	'his mat'
	b.	/j-kama/	kʲama	'his cornfield'
	c.	/j-tatah/	t̠atah	'his father'
		/j-tsʌhku/	ʧʌhku	'he did it'

Sagey analyzes palatalization as spreading [−back] from the preceding /j/ to all places (together with [dorsal] for labials and coronals), with a subsequent deletion of the glide. This analysis is illustrated for labials in (17). Spreading [−back] produces the required results for labials and dorsals, but not for coronals. For the latter, the addition of [dorsal[−back]] to [coronal[+anterior]] results in palatalized alveolars instead of the expected posterior coronals. What is necessary here, Sagey argues, is an additional process that would simplify the complex coronal structure Place[coronal[+anterior]]*[dorsal[−back]] to a posterior articulation, [coronal[−anterior]]. She refers to this process as "fusion," by which coronal and dorsal nodes are fused to produce a simple posterior coronal.

(17)


In sum, while the model provides a relatively simple and intuitively appealing account of secondary palatalization, its treatment of place-changing coronal palatalization as a two-stage process is arguably problematic for the same reasons as the *SPE* solution. Even more problematic, critics would argue, is the analysis of place-changing velar palatalization (Types IIa and IIb) (cf. Lahiri and Evers 1991; Hume 1992). Such an analysis is not worked out by Sagey, but would presumably involve a subsequent change of palatalized velars/palatals (Place[dorsal[−back]]) to posterior coronals (Place[coronal[−anterior]]). This, however, cannot be motivated by structural complexity, as palatalized velars (or palatals) are assumed to be simple articulations. Given this, fusion is not an option. Nor is it clear why this "simplification" should result in a posterior coronal, as opposed, for example, to an anterior coronal or a labial. Further, the model remains silent about the role of sibilancy in outputs of palatalization, and does not predict palatalization to anterior coronals (Type III) as a possible option.

Despite these and other limitations (see Kenstowicz 1994), the approach to palatalization as spreading [dorsal] has been relatively successful when dealing with contrastive palatalization, particularly with systems having both palatalized and velarized consonants. For example, both in Irish and in Russian consonants in clusters assimilate to following consonants in secondary palatalization or velarization, and vowels get fronted or backed by adjacent palatalized and velarized consonants. These facts can be easily stated as spreading the [dorsal] node with either [−back] or [+back] (Ní Chiosáin 1991; Rubach 2000). The use of the binary feature [±back] is also useful when it comes to stating morphological "exchange rules" – reversing palatality of consonants ([αback] → [−αback]) to mark certain morphological categories, as in Kildin Saami and Scots Gaelic (18) (Kert 1971; MacAulay 1992).

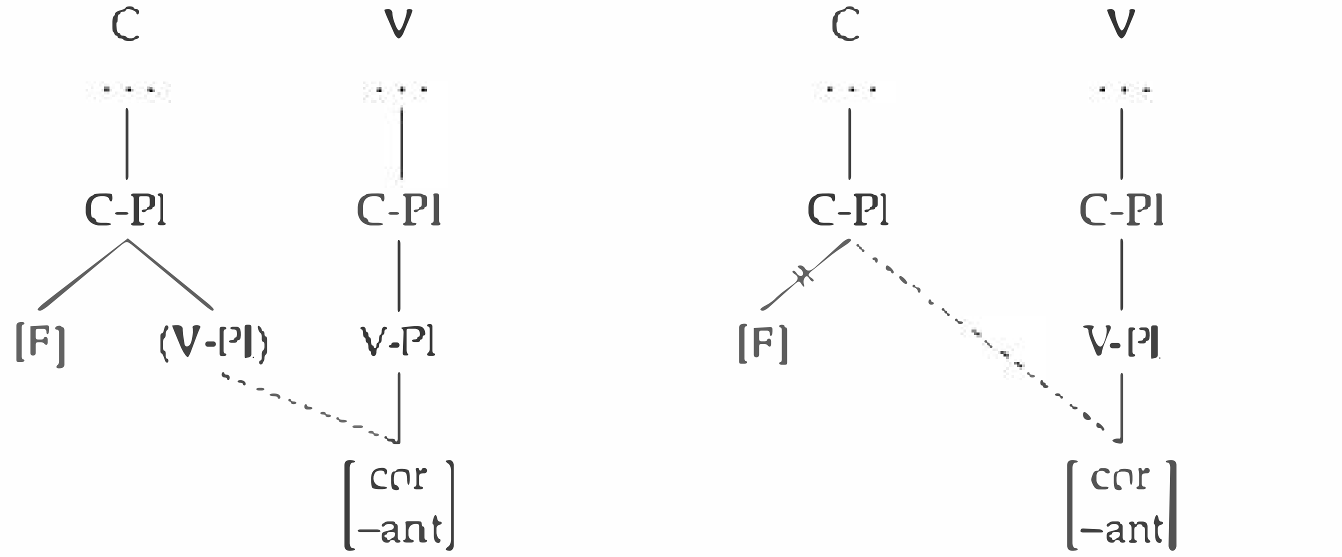
(18)	a.	kɛbb	'illness (NOM SG)'	kɛbʲbʲɛ	(DAT/ILLAT SG)
		kobʲbʲ	'pit (NOM SG)'	kobba	(DAT/ILLAT SG)
	b.	maːɫ̪	'rent (NOM SG)'	maːlʲ	(GEN SG/NOM PL)
		ahɔrʲ	'father (NOM SG)'	ahɔr	(GEN SG/NOM PL)

4.2 *Palatalization as spreading* [coronal]

The approach to palatalization as spreading [coronal] was advanced to remedy some of the inadequacies of the [dorsal] spreading model. It develops the original insight of Clements (1976) that palatalization and coronality are related, and that front vowels and coronals should form a natural class. While treatments of palatalization as spreading [coronal] were advocated in a number of works (Mester and Itô 1989; Broselow and Niyondagara 1990; Clements 1991; Lahiri and Evers 1991), the most extensive development of the idea was presented in Hume's dissertation (1992).

Hume's feature geometry model builds on Clements's (1991) proposal to use distinct tiers for consonant and vowel places, C-Place and V-Place nodes. These separate tiers were introduced for reasons largely independent of modeling palatalization – to allow for cross-consonantal assimilatory effects (such as vowel harmony and umlaut). These structures also made it possible to represent consonants with secondary articulation as having both C-Place and V-Place nodes. The V-Place node of vowels included the features [±coronal] and [±dorsal], with front vowels being [+coronal[−anterior]]. Height features were represented under a separate Stricture node, the property that will be relevant to our further discussion. Unlike the binary V-Place features, C-Place features were assumed to be primitive: [labial], [coronal], and [dorsal]. The [+anterior] under the [coronal] nodes referred to dentals and alveolars (as in Sagey's framework), while [−anterior] referred to various posterior coronal articulations, crucially including palatals. Despite some formal inconsistency in the use of binary and primitive features, the model allows for representing front vowels, coronal consonants, and palatalized consonants as a natural class – all sharing [coronal], specified either at the V-Place or C-Place. This is clearly a considerable advance in the theoretical modeling of palatalization, as both secondary palatalization (Type I) and place-changing palatalization (Types IIa and IIb) can be stated as assimilatory processes, virtually involving a single step. According to this analysis, secondary palatalization is triggered by spreading V-Place[coronal[−anterior]] from a front vowel or glide to the consonant. In the case of place-changing palatalization, this spreading is accompanied by delinking the original C-Place and promoting V-Place to the position of the former. Changes in other features, such as stridency or continuancy, are not considered to be part of the assimilation process per se, being specified as a rule parameter (the "constriction status change"). Hume's analysis of the two general processes is illustrated in (19).

(19) a. *Constriction status change: No* b. *Constriction status change: Yes*



The key insight of the [coronal] spreading approach is that secondary palatalization and place-changing palatalization (also known as "coronalization" with non-coronal targets) are essentially the same general process. Hume notes that both are cross-linguistically common, and in fact may optionally apply under the same phonological conditions in a given language, as in Acadian French (20).

(20) a. ki ~ kʲi ~ ʧi 'who' cf. [ka] 'case', [kut] 'cost', [kote] 'side'
kɛ ~ kʲɛ ~ ʧɛ 'quay'
b. tjed ~ tʲed ~ ʧed /tied/ 'lukewarm' cf. [dyp] 'dupe', [typ] 'type'

While these data nicely illustrate the similarity between secondary palatalization and place-changing palatalization, they also highlight some problems with the model. The triggers of the processes involving dorsals (20a) and coronals (20b) are different: front vowels in the first case and [j] in the second case. The target–trigger dependencies (see (8c)) are therefore not predicted by the model (cf.

Kenstowicz 1994). While the same is true for Sagey's and Chomsky and Halle's models, Hume's model has a structural limitation – stricture features are assumed to be independent of place features, and therefore statements of such dependencies are not possible. In fact, the model predicts that vowel height is not a factor in the process and any front vocoid can equally well palatalize a consonant of any place of articulation. While correctly capturing the important role of front vowels in palatalization processes, Hume's model does not allow for finer-grained frontness/height distinctions and rules out some of the attested processes. Among such processes are vowel raising next to palatalized consonants (as in Russian (11)) and a shift of velars to anterior coronals – the phenomena that could be relatively straightforwardly handled in the *SPE* approach.

4.3 *Further developments*

Subsequent work in the framework of feature geometry included attempts to resolve some of the problematic aspects of either approach, or to combine the insights of both. Lahiri and Evers (1991) propose to revise the [coronal] spreading approach by simplifying the two-tier place system and dispensing with the tier promotion mechanism used by Clements (1991) and Hume (1992). While maintaining the treatment of place-changing palatalization as due to spreading [Coronal[–anterior]], they analyze secondary palatalization as spreading [+high] (specified under the Tongue Position node) – a representationally elegant, yet arguably empirically problematic approach (Hume 1992; Jacobs and van de Weijer 1992). Calabrese (1993) uses alternative feature geometry representations and markedness filters in an attempt to address some of the issues largely overlooked in the Sagey and Hume approaches. Among these are the propensity of palatalization to produce sibilant affricates and fricatives, and the possibility of anterior coronals as outputs of the process. Jacobs and van de Weijer (1992) propose that front vowels are complex articulations, having both [coronal] and [dorsal] nodes (cf. Halle 2005). Palatalization may involve spreading only dorsal features, as in the case of velar fronting ([x] → [ç]), or both coronal and dorsal features, as in the case of place-changing palatalization of velars. This specification is also intended to characterize the class of coronals and dorsals as common targets of palatalization, as opposed to labials. While the move to specify front vocoids for both features adds flexibility to analyses of palatalization, its implications for analyses of other processes, such as vowel harmony and consonant harmony, and the interactions of these processes with palatalization, still remain to be explored. For example, is the patterning of front vowels in palatalization (as triggers) consistent with their patterning in backness vowel harmony (as targets or transparent vowels)? Do palatalized consonants always block backness harmony (as in Turkish: Kenstowicz and Kisseberth 1979)? Why do front vocoids fail to block coronal consonant harmony in some languages (Sanskrit: Calabrese 1993), while triggering it in other languages (Rundi: Broselow and Niyondagara 1990)? Finally, none of the above reviewed approaches seems to address the important question of why palatalization is special among consonant–vowel interactions – that is, why front vowels systematically displace consonant primary place of articulation, while other vowels hardly ever do so (cf. Ní Chiosáin and Padgett 1993).

To conclude this section, palatalization has played an important role in the development of a feature geometry framework, serving as a testing ground for competing proposals. While the rigidly constrained featural representations combined with a set of simple operations have contributed to a more empirically adequate account of cross-linguistic patterns of palatalization, it became clear that the same representations have often stood in the way of further empirical coverage of the phenomenon (and sometimes created problems for accounts of other phenomena). This particularly applies to cases of palatalization that can be considered less phonetically natural, such as place-changing processes resulting in anterior coronals or involving labial consonants. Ironically, some of the processes that could be easily stated in the *SPE*-style approach (although not always in a natural and insightful way) could no longer be stated in the feature geometry approach without making some ad hoc stipulations. At the same time, the discrete and binary feature geometry representations have also turned out to be incapable of capturing finer-grained, presumably phonetically motivated, scalar phenomena and trigger–target dependencies. This subsequently led some phonologists to (at least partly) revise the traditional view of representations as fixed and universal, and to explore ways of capturing cross-linguistic generalizations and variability in phonological processes through underspecification (Steriade 1995), contrastive specification (Avery and Rice 1989), contrastive feature hierarchies (Dresher 2009), or a system of parameterized rules (Archangeli and Pulleyblank 1994); see CHAPTER 2: CONTRAST; CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION).

5 Constraints and representations in Optimality Theory

The advent of Optimality Theory (OT; Prince and Smolensky 1993) brought back phonetic substance into phonology, now in the form of violable markedness constraints. While feature geometry-style representations and feature spreading assumptions have continued to play an important role in most OT accounts of palatalization, the task of capturing relevant feature asymmetries was partly relegated to constraints and constraint hierarchies. For example, the labial/non-labial target asymmetry could now be formalized as a universally fixed hierarchy of constraints prohibiting palatalized labials, dorsals, and coronals (21a) (Chen 1996; Rose 1997), while the trigger height asymmetry was represented as a fixed hierarchy of PALATALIZE (spread V-Place) constraints indexed for vowel height (21b) (Rubach 2003). Meshing these two hierarchies and combining them with different rankings of other markedness (e.g. AFFRICATION and POSTERIORITY; Rubach 2000) and faithfulness constraints can generate a restrictive factorial typology of palatalization patterns (cf. (14)), to some extent approximating the actual typology of palatalization (see §2). However, as the objects of constraint manipulation were the same inviolable feature geometry representations, some of the earlier noted problems persisted into OT analyses.

(21) a. *[lab]/VPL[cor] >> *[dors]/VPL[cor], *[cor]/VPL[cor]
b. PAL/j, PAL/i >> PAL/e >> PAL/æ

One possible solution to these problems was sought in the use of more detailed, phonetically realistic representations. Chen (1996), for example, uses articulatory gestures in conjunction with the traditional feature geometry representations to analyze palatalization in Japanese, Polish, and Swati. All palatalization processes are assumed to involve spreading V-Place[coronal] from front vocoids (following Hume 1992) and resulting in abstract complex segments with a secondary place and the original primary place (e.g. [Dorsal]/V-Place[Coronal]). The cross-linguistic diversity in outputs of palatalization arises, according to Chen, from language-particular phonetic implementation via articulatory gestures (Browman and Goldstein 1989). Bateman's (2007) OT analysis of cross-linguistic patterns of palatalization fully replaces the traditional feature geometry representations with articulatory gestures. She models secondary palatalization and place-changing palatalization as resulting from two different gestural coordination strategies: the coordination of the vowel gesture at the release of the consonant gesture or at the center of it – producing either consonants with secondary articulation or simple articulations of intermediate constriction location and degree, respectively. The appeal of both proposals is in the use of independently motivated, physically concrete representations and a simple mechanism of gestural overlap. Problems arise, however, as before, with treatments of articulatorily less natural cases of dorsal and labial place-changing palatalization. In Chen's analysis, velar palatalization results in the abstract phonological structure [Dorsal]/V-Place[Coronal], which can be phonetically interpreted as [kʲ], [c], [ʧ], or [tɕ], depending on the language. Yet it is not clear how this mapping would work in languages with more than one velar palatalization process (as Russian (4)). In Bateman's analysis, the process [k] → [ʧ] cannot be reasonably analyzed as resulting from gestural blending only (which would give only [kʲ] or [c]), and requires additional markedness stipulations (cf. Chomsky and Halle 1968 on [c] → [ʧ]). While Chen analyzes labial palatalization in Swati as a case of phonological neutralization to the default coronal place, this is not an option for the gestures-only framework of Bateman. As gestural blending is technically impossible between the mechanically uncoupled gestures of the lips and the tongue body, labial palatalization is in principle ruled out by the model. Bateman contends that the few attested cases of labial–coronal alternations (as in Southern Bantu and Moldova Romanian) can be explained diachronically. Yet, arguably, these cases still require a synchronic analysis.

A different approach – exploring phonetically detailed, scalar auditory features – is taken by Flemming (2002). He analyzes palatalization as a process driven primarily by constraints requiring perceptual enhancement of phonological contrasts (as part of his Dispersion Theory; see CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). In this analysis, secondary palatalization is an optimizing strategy, as it enhances the contrast of the vowel with other vowels by extending the span of its second formant (F2) to the preceding consonant (e.g. Nupe [egʲe] *vs.* [ega] compared to [ege] *vs.* [ega]). The change of the fronted velar or palatal stop to a palato-alveolar affricate is yet another step in the enhancement of the contrast (e.g. [edʒe] *vs.* [ega]), by which the duration of frication and its loudness are increased, while the contrast with the non-palatalized counterpart in F2 remains relatively large. Thus, sibilants as outputs of palatalization are fully expected, as affrication is part of contrast enhancement: "It is easier to enhance a contrast by exaggerating a difference that would be present anyway as an articulatory side-effect, rather than attempting to reverse the articulatorily motivated pattern" (2002: 106). The same kind of enhancement through affrication is also possible for coronals, but is unlikely for palatalized labials, since the production of these does not involve as much frication. While the actual implementation of this analysis of palatalization is relatively complex, it does capture some important generalizations about palatalization processes that have evaded many previous analyses. The Dispersion Theory approach thus provides an interesting insight into how palatalization may arise through auditory enhancement of phonological contrasts (cf. Padgett 2001, 2003). It remains to be seen, however, how the approach can model synchronic alternations, and particularly more complex cases of morphophonological palatalization.

6 Recent alternatives

Despite the greater flexibility and apparent naturalness provided by violable substantive constraints in OT, some of the problems with the formal modeling of palatalization have not been resolved. In part, these difficulties appeared to stem from a more fundamental problem – the persistent use of traditional featural representations (with some modifications), which were assumed to be inviolable, universal, and innate. These assumptions about representations were clearly important in the development of generative phonology, as the universal set of features provided a simple formal tool to state phonological rules and to capture significant cross-linguistic generalizations about natural classes of segments. Yet the basis for these assumptions has hardly been questioned or systematically investigated until recently. As Mielke's (2008) survey of phonological processes shows, unnatural classes are widespread in languages, with some of them being more common than typical natural classes. As traditional feature theories are incapable of characterizing many of these classes, the usefulness of maintaining the assumptions about feature universality and innateness is in serious doubt. Mielke's proposal is that features are not innate but emergent, arising from language learners' phonetic generalizations (cf. Hayes and Steriade 2004 on OT constraints). If features, and phonological representations in general, are indeed emergent, this has some wide-ranging implications for phonological theory, and for formal modeling of phonological processes. Specifically with respect to palatalization, languages may be expected to vary in how they define features and natural classes involved in the process, while at the same time showing many similarities, given the similar articulatory and acoustic properties of alternating consonants and vowel triggers. One may also expect that featural representations are not immutable within a given language, but possibly reflect local generalizations, specific to certain morphological domains or lexical strata (as, for example, in cases of multiple palatalization processes targeting the same consonants). However, these and many other implications for analyses of palatalization have not yet been explored.

Another notable recent development reflects a resurgence of interest in diachronic explanation of synchronic phonological patterns. This approach is most systematically represented by Blevins's (2004) Evolutionary Phonology, where cross-linguistically common, "natural" sound patterns are explained exclusively diachronically – as a by-product of recurrent phonetically motivated sound changes. Given the well-established phonetic motivation for palatalization in co-articulation and auditory misperception (see §2.3), synchronic patterns of palatalization can be interpreted as arising from sound changes involving these phonetic factors. As such, these patterns arguably do not require synchronic explanation – either structural or substantive (cf. Kochetov 2002 on the phonotactics of palatalization contrasts). Taking velar place-changing palatalization as an example, the unidirectional nature of this change ([ki] → [ʧi], *[ʧi] → [ki]) and its common result (a postalveolar affricate) has little to do with phonological grammar per se, as it can be attributed to common errors in the perception of fronted velars (Guion 1996). The same applies to the asymmetry between high and non-high front vowels as triggers – listeners simply make more errors of the type [ki] → [ʧi] than [ke] → [ʧe]. By the same token, listeners rarely make errors like [pi] → [ʧi], unless under some specific phonetic conditions (see Ohala 1978) – the fact that explains the labial/non-labial asymmetry in palatalization. If most or all the cross-linguistic generalizations about palatalization in (8) can be accounted for by phonetically based sound changes, the goal of synchronic grammar becomes much simpler – to state language-particular generalizations about the patterning of segments in alternations or their phonotactic distribution. What specific form these language-particular synchronic grammatical generalizations would take, however, is not clear, and has not been sufficiently explored by the proponents of Evolutionary Phonology. One interesting implication of the approach is that synchronic patterns of palatalization alternations should mirror sound changes involving the process. Whether this is true, however, is subject to further typological research. Another related question is how to reconcile the substance-free grammar envisioned by Evolutionary Phonology with apparent evidence that speakers possess some phonetic knowledge and seem to use it to make higher-level grammatical generalizations (Hayes and Steriade 2004). An interesting relevant case is provided by the cross-linguistically common use of palatalization in baby talk and diminutive sound-symbolism – presumably reflecting bottom-up generalizations, grammaticalized associations between the phonetics of palatalized consonants and the meaning of smallness and childishness (Kochetov and Alderete, forthcoming). Whether phonetic knowledge plays a role in phonological generalizations, and specifically whether phonetic naturalness considerations are part of the grammar, are important questions that could possibly be answered through systematic psycholinguistic experimentation and computer simulations (see some relevant work by Wilson 2006). The challenge for future work is, therefore, to tease apart synchronic phonological and phonetic knowledge of palatalization and historical influences shaping cross-linguistic patterns of palatalization over time.

ACKNOWLEDGMENTS

The chapter has benefited greatly from the insightful comments and helpful suggestions provided by two anonymous reviewers and the editors, Marc van Oostendorp and Beth Hume.

REFERENCES

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Avery, Peter & Keren Rice. 1989. Segment structure and coronal underspecification. *Phonology* 6. 179–200.

Bateman, Nicoleta. 2007. A crosslinguistic investigation of palatalization. Ph.D. dissertation, University of California, San Diego.

Bhat, D. N. S. 1978. A general study of palatalization. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 2: *Phonology*, 47–92. Stanford: Stanford University Press.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Borowsky, Toni. 1986. Topics in the lexical phonology of English. Ph.D. dissertation, University of Massachusetts, Amherst.

Broselow, Ellen & Alice Niyondagara. 1990. Feature geometry and Kirundi palatalization. *Studies in the Linguistic Sciences* 20. 71–88.

Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6. 201–251.

Calabrese, Andrea. 1993. On palatalization processes: An inquiry about the nature of a sound change. Unpublished ms., Harvard University.

Chen, Matthew Y. 1973. Predictive power in phonological description. *Lingua* 32. 173–191.

Chen, Su-I. 1996. A theory of palatalization and segment implementation. Ph.D. dissertation, New York State University, Stony Brook.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1976. Palatalization: Linking or assimilation? *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 12. 96–109.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. 1991. Place of articulation in consonants and vowels: A unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5. 77–123.

Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.

Flemming, Edward. 2002. *Auditory representations in phonology*. London & New York: Routledge.

Foley, James. 1977. *Foundations of theoretical phonology*. Cambridge: Cambridge University Press.

Guion, Susan. 1996. Velar palatalization: Coarticulation, perception, and sound change. Ph.D. dissertation, University of Texas.

Halle, Morris. 2005. Palatalization/velar softening: What it is and what it tells us about the nature of language. *Linguistic Inquiry* 36. 23–41.

Haspelmath, Martin, Matthew S. Dryer, David Gil, Bernard Comrie & Hans-Jörg Bibiko (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press.

Hayes, Bruce & Donca Steriade. 2004. Introduction: The phonetic bases of phonological markedness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 1–33. Cambridge: Cambridge University Press.

Hume, Elizabeth. 1992. Front vowels, coronal consonants and their interaction in nonlinear phonology. Ph.D. dissertation, Cornell University. Published 1994, New York: Garland.

Hyman, Larry M. 1975. *Phonology: Theory and analysis*. New York: Holt, Rinehart & Winston.

Jacobs, Haike & Jeroen van de Weijer. 1992. On the formal description of palatalisation. In Reineke Bok-Bennema & Roeland van Hout (eds.) *Linguistics in the Netherlands*, 125–135. Amsterdam & Philadelphia: John Benjamins.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kenstowicz, Michael & Charles W. Kisseberth. 1977. *Topics in phonological theory*. New York: Academic Press.

Kenstowicz, Michael & Charles W. Kisseberth. 1979. *Generative phonology: Description and theory*. New York: Academic Press.

Kert, G. M. 1971. *Saamskii iazyk (Kildinskii dialekt): Fonetika, morfologiia, sintaksis*. Moscow: Nauka.

Kochetov, Alexei. 2002. *Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization*. New York & London: Routledge.

Kochetov, Alexei & John Alderete. Forthcoming. Patterns and scales of expressive palatalization: Typological and experimental evidence. *Canadian Journal of Linguistics*.

Kotzé, Albert E. & Sabine Zerbian. 2008. On the trigger of palatalization in the Sotho languages. *Journal of African Languages and Linguistics* 29. 1–28.

Lahiri, Aditi & Vincent Evers. 1991. Palatalization and coronality. In Carole Paradis & Jean-François Prunet (eds.) *The special status of coronals: Internal and external evidence*, 79–100. San Diego: Academic Press.

Lightner, Theodore M. 1965. Segmental phonology of Contemporary Standard Russian. Ph.D. dissertation, MIT.

MacAulay, Donald. 1992. The Scottish Gaelic language. In Donald MacAulay (ed.) *The Celtic languages*, 137–248. Cambridge: Cambridge University Press.

Mathangwane, Joyce T. 1996. Phonetics and phonology of Ikalanga: A diachronic and synchronic study. Ph.D. dissertation, University of California, Berkeley.

Mester, Armin & Junko Itô. 1989. Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language* 65. 258–293.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Ní Chiosáin, Máire. 1991. Topics in the phonology of Irish. Ph.D. dissertation, University of Massachusetts, Amherst.

Ní Chiosáin, Máire & Jaye Padgett. 1993. Inherent VPlace. Report LRC-93-09, Linguistics Research Center, University of California, Santa Cruz.

Ohala, John J. 1978. Southern Bantu *vs.* the world: The case of palatalization of labials. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 4. 370–386.

Padgett, Jaye. 2001. Contrast dispersion and Russian palatalization. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 187–218. San Diego: Academic Press.

Padgett, Jaye. 2003. Contrast and post-velar fronting in Russian. *Natural Language and Linguistic Theory* 21. 39–87.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Rose, Sharon. 1997. Theoretical issues in comparative Ethio-Semitic phonology and morphology. Ph.D. dissertation, McGill University.

Rubach, Jerzy. 2000. Backness switch in Russian. *Phonology* 17. 39–64.

Rubach, Jerzy. 2003. Polish palatalization in Derivational Optimality Theory. *Lingua* 113. 197–237.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT. Published 1990, New York: Garland.

Schachter, Paul. 1969. Natural assimilation rules in Akan. *International Journal of American Linguistics* 35. 342–355.

Stadnik, Elena. 2002. *Die Palatalisierung in den Sprachen Europas und Asiens: Eine arealtypologische Untersuchung*. Tübingen: Gunter Narr Verlag.

Steriade, Donca. 1995. Underspecification and markedness. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 114–174. Cambridge, MA & Oxford: Blackwell.

Udler, R. Y. 1976. *Dialektnoe chlenenie moldavskogo iazyka.* Chişinău: Shtinica.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30. 945–982.

Zsiga, Elizabeth C. 1993. Features, gestures, and the temporal aspects of phonological organization. Ph.D. dissertation, Yale University.

72 Consonant Harmony in Child Language

Clara C. Levelt

1 Introduction

Consonant harmony (CH) in child language production data has attracted a great deal of attention in the phonological literature. It has been defined as an "assimilation-at-a-distance" process between consonants (Vihman 1978), in which consonants affect other, non-adjacent consonants. The assimilating features in child language CH are mostly primary place of articulation features, like Labial and Dorsal, but cases where other features are involved have also been observed. Over time, several analyses of this phenomenon have been proposed in the literature, the nature of the analysis changing with the theoretical state of the art: a phonological rule (Smith 1973), autosegmental spreading (Menn 1978; McDonough and Myers 1991; Stemberger and Stoel-Gammon 1991; Levelt 1994), a connectionist account (Berg and Schade 2000), and constraint interaction (Goad 1997; Pater and Werle 2003; Fikkert and Levelt 2008).

Consonant Harmony as such is not a phenomenon specific to child language (see CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS). However, the nature of CH in child language differs from that in the languages of the world in an important way: unlike in the world's languages, in child language it appears that *primary* place of articulation features can assimilate between non-adjacent consonants. This constitutes a challenge for a phonological account, because it clearly violates the principle of locality (see e.g. Archangeli and Pulleyblank 1987). According to this principle, only segments that are adjacent at some level of analysis can interact.[1] For primary place of articulation features (CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION) of consonants that are not string-adjacent such a level can only be assumed under special circumstances, such as planar segregation (McCarthy 1989; see §3 below). Of course, CH in child language would violate locality only if a strong form of continuity is assumed, i.e. if the phonological systems of language learners and adults make use of the same units and obey the same set of principles. Therefore, as we will see, some accounts of child language

[1] For an overview of the different definitions of locality, and the ways in which accounts of CH in the world's languages deal with this principle, see CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS.

CH refer to a child-language-specific aspect of the developing phonological system, which allows for a local interpretation of the interaction. Alternatively, the locality problem can be circumvented by denying that the data are the result of an assimilation-at-a-distance process. Instead, some – mostly child-specific – form of feature licensing is invoked to account for the data.

This interesting and sometimes controversial topic will be discussed in the following way. First, in §2 the main similarities and differences between CH in adult language and in child language will be pointed out. Because in child language CH involving primary place of articulation features is the most salient and systematic phenomenon, and forms a challenge for most phonological theories, I will focus on this type of harmony in the remainder of the chapter. In §3, different accounts of CH in child language will be discussed, focusing on the way in which phonological theory has highlighted different – problematic – aspects of the phenomenon. In §4, a view on CH in child language will be presented that sets it apart from CH in adult languages, and it will be discussed how such a child-specific phenomenon can come about during language acquisition. §5 concludes this chapter.

2 Consonant harmony in adult and child language

The prevalent view on consonant harmony is that it is a widespread phenomenon in child language, while it is rare in the world's languages. This view is not totally supported by facts, however. Hansson (2001) describes different types of CH processes in 127 languages, while descriptions of systematic CH processes in child language can be found for only a handful of children (Smith 1973 and Goad 1997 (Amahl); Cruttenden 1978 (one child); Menn 1978 (Daniel); Berg and Schade 2000 (Melanie); Levelt 1994 and Fikkert and Levelt 2008 (Eva, Robin); Rose 2000 (Clara); Pater and Werle 2001, 2003 (Trevor)). Some studies discuss larger groups of children (Vihman 1978 (13 children); Stemberger and Stoel-Gammon 1991 (69 children)). However, it is unclear from Stemberger and Stoel-Gammon's study how many of these children actually had CH productions or to what extent the phenomenon occurred systematically in the data. Vihman studied CH forms in vocabularies that contained between 109 and 372 words. In only four of the 13 children did she find a relatively high number of CH productions, i.e. between 18 and 32 percent of the productions in the vocabulary studied. Of the remaining nine children, three scored around 10 percent and the other six scored between 1 percent and 5 percent. In the study, almost half of all the CH forms were provided by two of the children, Amahl and Virve. All in all, if we base ourselves on facts, i.e. cases reported in the literature, then "rare in languages of the world" is supported, but "widespread in child language" is less obvious. It is widespread in the sense that an occasional CH form will probably show up in the speech of many children. Systematic CH patterns, however, i.e. CH forms that show up predictably for a longer period of time, have, up to now, only been described for a handful of children.

2.1 *Features involved in harmony*

Most CH processes, in child and adult language alike, involve place of articulation features. In the languages of the world, CH always concerns *secondary* place of articulation features. Feature-geometrically speaking, these features are usually dependents of the coronal node: [anterior], [distributed], and [strident]. CH of features that are dependents of the labial or dorsal node exists too, but is very rare. Place of articulation harmony in the world's languages occurs mostly between consonants that are already very similar: within a word, the feature value for [anterior] is shared between sibilants, or the feature values [anterior] and [distributed] are shared between stops. For a more extensive review see CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS and Hansson (2001).

In child language, however, the CH phenomenon that is discussed most often concerns *primary* place of articulation features, specifically Labial and Dorsal. The existence of a systematic Coronal harmony process in child language is less evident. For one thing, neutralization to a coronal place of articulation, as in Velar Fronting, where /k/ is replaced by [t], often occurs as an independent process during phonological development (CHAPTER 12: CORONALS). Utterances with multiple coronal consonants are thus often the result of neutralization rather than assimilation. Below, in §3.2.2, we will see that underspecification of Coronal in the lexicon has been invoked to account for the absence of coronal harmony in child language. Unlike in adult CH, the consonants involved are not necessarily highly similar in other respects: primary place of articulation features are shared between any combination of nasals, fricatives, and stops.

Concerning place of articulation, then, the two groups of speakers appear to have almost contrasting sets of features that are active in CH: adult speakers only show CH involving *secondary* place of articulation features, mostly dependents of Coronal, while children show CH involving *primary* place of articulation features, most commonly Labial and Dorsal. Typical examples of these two types of harmony are given in (1).

(1) *Place of articulation harmony*

Adult speakers

a. *Sibilant* [anterior] *harmony in Ineseño Chumash*
(Applegate 1972, cited by Hansson 2001)
/k-su-ʃojin/ [kʃuʃojin] 'I darken it'
/s-api-ʧʰo-it/ [ʃapiʧʰolit] 'I have a stroke of good luck'

b. *Coronal* [anterior] *harmony in Päri*
(Andersen 1988, cited by Hansson 2001)
[dè:l] 'skin' [dè:nd-á] 'my skin'
[t̪ùol] 'snake' [t̪úon̪d̪-à] 'my snake'

Child speakers

c. *Dorsal harmony* (English)
(Trevor at 1;5: Compton and Streeter 1977, cited by Pater and Werle 2003)
dog [gɔg]
bug [gʌg]
coat [kok]

d. *Labial harmony* (Dutch)
(Robin at 1;10: Levelt 1994)
tafel /tafəl/ [pafy] 'table'
zeep /zep/ [fep] 'soap'
neef /nef/ [mef] 'cousin'

In (1a) we observe that two sibilants that underlyingly carry different feature values for [anterior] at the surface both show up as [−anterior]. In (1b) the feature value for [anterior] of the stop consonant in the stem is shared with the prenasalized stops in the derived forms. The examples in (1c) from child language show Dorsal CH in data from an English-speaking child. Both underlying Labial and Coronal consonants show up as Dorsal consonants on the surface. In (1d) Labial CH is illustrated with examples from a Dutch-speaking child. Here we see that the interacting consonants do not necessarily agree in their manner features.

Harmony involving other features is quite rare, both in languages of the world and in child language. In languages of the world, systematic patterns of long-distance assimilation have been attested for laryngeal features, nasality, and continuancy. In Hansson's (2001) overview we also find languages that show some form of liquid harmony. This form of harmony, or rather lateral harmony, is the only other type of consonant harmony that apparently, occurs in a systematic way in the speech of a child, Amahl, discussed in Smith (1973). In Amahl's case a target word with a combination of an /r/ or /j/ and a lateral results in a production with two laterals, as illustrated in (2).

(2)	*really*	[li:li:]
	lorry	[lɔli]
	yellow	[lɛlɔu]

Some less evident forms which have been listed as CH forms, involving other features, are given in (3) (from Spanish child language; Vihman 1978).

(3)	*comiendo*	[kabiendo]	'eating'	nasal is assimilated to stop
	llorando	[ɲrdardɲo]	'crying'	nasal harmony
	telefono	[ɸweɸwʌno]	'telephone'	continuant harmony

However, no systematic patterns of CH in child language have been described that involve these features. For forms like in (3), which are probably just produced once, there is no predictable relation between the form of the adult target and the resulting production. More examples of occasional CH productions are given in (4), from the speech of Jiři (Czech; Vihman 1978).

(4)	*balonek*	[babo:nek]	'ball'
	ježek	[ʒeʒek]	'hedgehog'
	gramofon	[gagafo:n]	'gramophone'

According to Vihman (1978), forms like these seem analogous to speech errors or alliterations in adult speech. Up to now, the reported set of occasional CH forms has been so diverse and fragmented that it has been impossible to come up with a comprehensive analysis.

Since CH in child language involving features other than place of articulation features presents such an unclear picture, the remainder of the chapter is concerned solely with primary place of articulation harmony.

2.2 *Directionality of the process*

Consonant harmony comes in two varieties in the languages of the world. Most commonly it takes the form of a morpheme structure constraint (CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS): certain combinations of consonants within a stem are allowed, while others are disallowed (Hansson 2001). In these cases, determining the direction of assimilation is not always evident, since underlying and surface forms are identical with respect to these harmonic consonants. The direction can sometimes be reconstructed from diachronic and cross-linguistic comparisons, which have shown that the default direction of harmony is anticipatory, i.e. right to left. Furthermore, Hansson (2001) establishes a Palatal Bias effect in adult CH. An underlying – or former – alveolar–palatal combination is likely to become a palatal–palatal combination. The morpheme structure constraint can be accompanied by a morphological harmony rule, i.e. a productive harmony process. Here, harmony is most commonly stem-controlled: the consonant of an affix will carry a certain feature value, depending on the feature value of consonants in the stem (CHAPTER 104: ROOT–AFFIX ASYMMETRIES).

In child language, the CH phenomenon is usually present in the period before any productive morphology has been developed. We therefore do not find CH forms that can be analyzed as being either stem- or affix-controlled. Rather, certain combinations of sounds within a stem appear to be disallowed on the surface. In this sense, CH in child language seems to take the form of a morpheme structure constraint. However, under the assumption that children's underlying forms are similar to the adult target forms, and given that these adult target forms can contain the disallowed combination of sounds, child CH is often assumed to be an active process. As in adult CH, the default direction of the process is right to left. Instead of the Palatal Bias effect that was found in languages of the world, in child language we find a strong Labial or Dorsal Bias effect: if the C_2 in a target $C_1VC_2(V)$ combination is Labial, the C_1 will end up being Labial too, or if C_2 is Dorsal, C_1 will be Dorsal.

2.3 *Summary*

What can be concluded from the above comparison between adult CH and child CH? Are the phenomena similar or different? Hansson (2001) pulls together adult CH, child CH, and speech errors. He states that the underlying source for CH must lie in the domain of phonological encoding for speech production. Both speech errors and CH show a default right-to-left directionality, and, as in adult CH, assimilatory speech errors are more likely to occur between segments that are already very similar. According to Hansson, then, CH in languages of the world is a phonologized form of speech error. With a little twist, this could also apply to CH in child language. The occasional forms are speech errors (as proposed in Vihman 1978) and in some cases a systematic, i.e. phonologized, type develops. According to Hansson, the difference in place of articulation bias between adult CH and child CH is caused by the nature of the sound inventory. In child language, the sound inventory is much smaller, and minor place of articulation features do not yet play a role. This impoverished inventory also puts "similarity between consonants" in a different perspective. Pairs of segments that are judged as very different by adult speakers, like /t/ *vs.* /k/, could be judged as relatively similar by children. This would account for the fact that major place harmony is child-language-specific, and is not found in languages of the world. In this view, then, CH in the world's languages and CH in child language are of the same kind, and the different surface appearance can be attributed to the impoverished segment inventory in the developing phonological systems of young children. Although this is an elegant perspective, which makes it possible to view phonological development as being continuous, by invoking and adhering throughout to identical principles and processes, a different account of the child-specificness of the phenomenon will be proposed in §4. Taking into account the developing place of articulation structure of young children's entire vocabulary, it appears that CH in child language is of a very different nature than CH in the world's languages.

In the remainder of this chapter we will concentrate on consonant harmony in child language, starting in §3 below with an overview of the accounts of CH that have been proposed in the literature.

3 Theoretical approaches to consonant harmony in child language

Several grammatical accounts have been presented of child language CH. These accounts are in terms of rules, autosegmental representations, activation spreading, or constraints. I will pay special attention to the way the account deals with the issue of *locality*: how is the intervening vowel dealt with, and which part of the process, if any, is deemed child-language-specific?

3.1 Consonant harmony as the result of a phonological rule

Smith (1973), working in the tradition of *SPE* (Chomsky and Halle 1968), presents a series of "realisation rules" that derive the consonant harmony forms of his son Amahl from "English Standard Pronunciation," i.e. the adult forms. In fact, Smith argues that one of the general functions of realization rules is to implement both consonant and vowel harmony, and he suspects that it is universal in child language. Of the eight realization rules that have consonant (and vowel) harmony as their motivation, the one in (5) results in labial and dorsal harmony:

(5) $$[+\text{coronal}] \rightarrow \begin{bmatrix} -\text{coronal} \\ \alpha\text{anterior} \end{bmatrix} / \_\_ \ [+\text{syllabic}] \begin{bmatrix} -\text{coronal} \\ \alpha\text{anterior} \end{bmatrix}$$

This rule initially applied systematically before velars, but was optional if $\alpha = +$, i.e. before labials. The intervening vowel, i.e. the [+syllabic] element in the rule, does not play any role and is not considered to be an obstacle to the process. The rule undergoes a couple of changes over time, capturing the fact that fewer and fewer coronal segments are affected. As a first change, for example, the rule split into two parts, operating in the original way in case of [−anterior], but applying only to nasals and continuants when [+anterior]. In its different forms, the rule is operative in Amahl's system from stage (1), when the data collection started at 2 years and 60 days, until stage (14), when Amahl was 2 years and 247 days. The rule can be considered child-language-specific, in that it disappears from Amahl's system over time. However, according to Smith, it is a genuine phonological rule in the sense that the formal properties of the realization rules are the same as those of phonological rules in mature grammars.

3.2 Consonant harmony in the autosegmental framework

3.2.1 Output templates

The first account in an autosegmental framework (CHAPTER 14: AUTOSEGMENTS) is presented by Menn (1978). She views CH as one of the child's strategies to comply with a general constraint on his or her output structure. The proposed, child-language-specific, output constraint is termed a "consonant harmony constraint" and states that consonants within a word should be of one place type. There are different ways to comply with this constraint. If an adult target word contains consonants with different place features, the child can either delete all but one consonant, or the child can render all the consonants in a word of one place type. This perspective integrates CH in the child's phonological system as a whole, instead of treating it as an isolated phenomenon: CH is just one of the possible ways of satisfying the constraint. In §4 a similar integrated view is elaborated. Menn posits the following "output lexical entry" for the words *stuck*, *duck*, and *truck*, which are all produced [gʌk] by the child Daniel:

(6) *Output lexical entry for* stuck, duck, *and* truck

tier 3	stop position	#		velar		#
tier 2	fricative	#		Ø		#
			C	V	C	
tier 1	word structure	#	+voice	low-mid	−voice	#

The child has the output representation in (6), resulting from the rule: "If an entry in the recognition lexicon contains a velar, then select [velar] as the stop-position specification for the corresponding entry in the output lexicon" (1978: 167). As in Smith's account, above, the underlying stored form is altered if conditions apply; in case a stored form contains [coronal] or [labial] and [velar], only the feature [velar] will end up being linked to the consonant positions in the word. Although this seems to be a classic case of autosegmental spreading, in this account the intervening vowel is not perceived as posing a potential problem.

A slightly different templatic approach to CH is taken by Iverson and Wheeler (1987). Following Moskowitz (1971), among others, who posited that words appear to be unanalyzed units, they argue that many phonological phenomena in child language are the result of the association of features with suprasegmental constituents, like words, syllables, and rhymes. The child's output representations are viewed as well-formedness templates, which characterize and filter the set of permissible words in the child's language (1987: 249). The well-formedness template that would result in the CH productions [kok] for *coat* and [gag] for *dog* is given in (7):

(7) *Output template for* coat *and* dog

WORD
[–anterior]
C V C

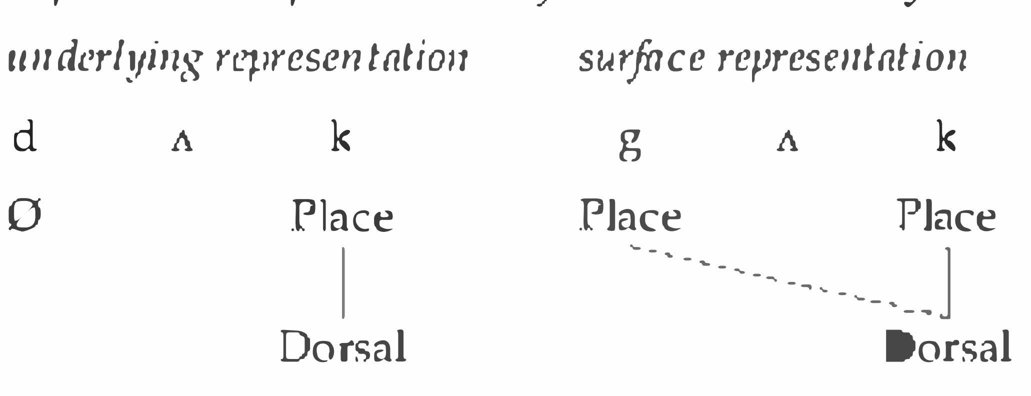
For the child in question, any word having a [–anterior] consonant will be associated with the above word structure, resulting in the harmonized forms. In Iverson and Wheeler's view, CH is actually not a derivational process, linking an adult-like input representation to a harmonized child output representation. The output form actually represents the child's knowledge of the phonological system of his or her target language, and it thus equals the underlying representation. What the child needs to learn, then, is that features should be associated with segments instead of larger units like syllables or words. The child-language-specific aspect of the account is the fact that features link to entire words, rather than segments.

Apart from the fact that the proposed constraint behaves like a morpheme structure constraint rather than as a constraint on output forms only, the account is almost exactly parallel to Menn's account: there is a child-specific template and a floating place feature that will be linked to the C-slots in the template (CHAPTER 54: THE SKELETON). Although the notion of association line is appealed to in Iverson and Wheeler's paper, the intervening vowel is still not viewed as potentially problematic for the account. As long as [anterior] is the feature being associated, the intervening vowel will not disrupt the linking process, since vowels are not normally specified for [anterior]. In this case, the association of [–anterior] with the vowel will simply have no effect. In this account, then, the locality problem is circumvented because only consonant-specific features are used.

3.2.2 *Underspecification*

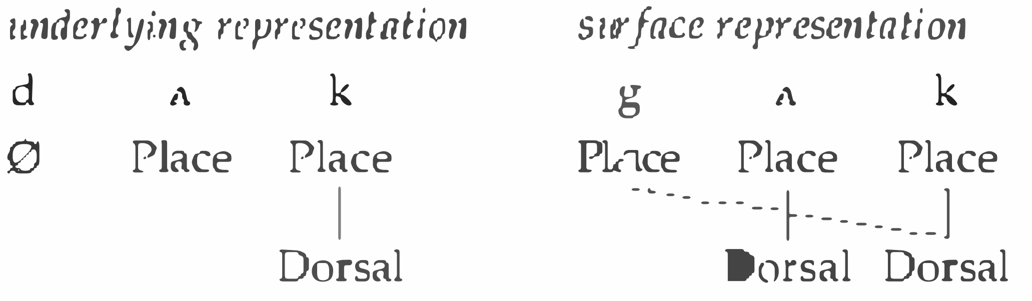
A more theoretically detailed autosegmental analysis of CH is presented in Stemberger and Stoel-Gammon (1989, 1991) and in Stoel-Gammon and Stemberger (1994). Here, CH is considered to be an "unconscious" process, caused on the one hand by underspecified consonants in the child's inventory, and on the other hand by a tendency for unmarked segments to assimilate to marked segments. CH is thus viewed as a feature-filling process, whereby a place feature spreads from a consonant specified for place, to a consonant unspecified for place. This is illustrated for the form [gʌk] for *duck* in (8):

(8) *A procedural representation of consonant harmony*



However, in the model of feature representation adopted by Stemberger and Stoel-Gammon and that of Sagey (1986), vowels also have a Dorsal place specification. This entails that spreading Dorsal from /k/ to the initial consonant position would lead to crossing association lines, as in (9), which is ruled out by the Line Crossing Prohibition (Goldsmith 1976).

(9) *Consonant harmony and crossing association lines*

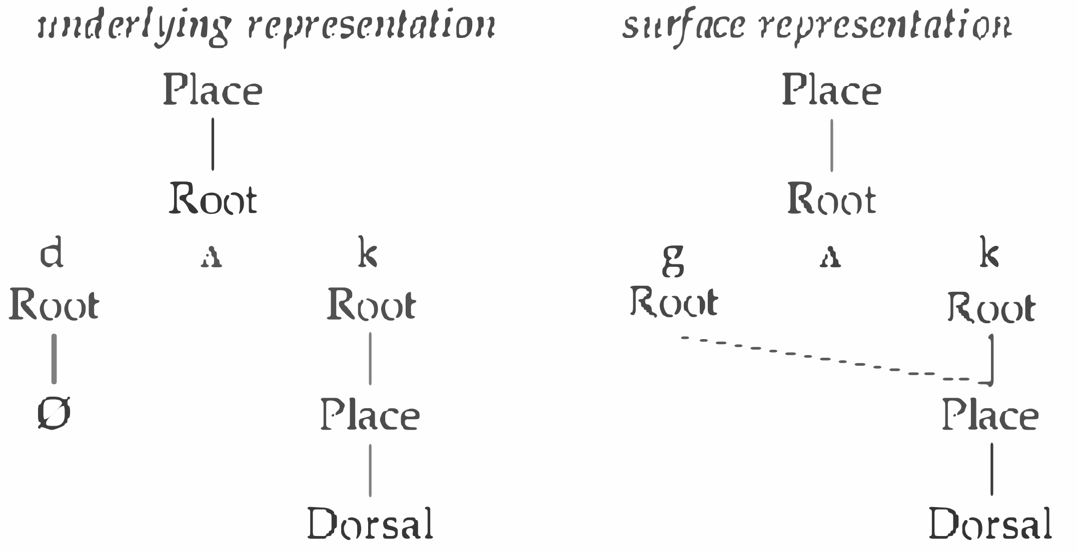


Stemberger and Stoel-Gammon recognize this locality problem. They argue that since these intervening vowels apparently do not block the harmony process, they should be transparent in one way or another. To achieve this, consonants and vowels should either reside on different planes when the process takes place, i.e. there is planar segregation (McCarthy 1989), or vowels and consonants should have different sets of place features (see also CHAPTER 105: TIER SEGREGATION; CHAPTER 27: THE ORGANIZATION OF FEATURES). They opt for the last solution and turn to the feature model proposed by Clements (1985), where consonants and vowels are partially segregated. Place features are divided into a "primary" place tier containing consonantal place features, and a "secondary" place tier containing vocalic place features. In this model, place features of consonants can be spread across vowels, and place features of vowels can be spread across consonants. Consonant harmony can thus be characterized as feature spreading, affecting only the primary place node. Interference with vowels, specified for place on the secondary place node, is avoided. In Clements's later elaboration of his feature model (Clements 1991), however, both vowels and consonants now have this primary consonant-place node. One of Clements's arguments for the change is precisely to exclude the possibility of consonants spreading their place features across vowels, which does not occur in the world's languages.

3.2.3 Planar segregation

McDonough and Myers (1991) take the planar segregation option seriously. Planar segregation can only be invoked if the relative order of consonants and vowels is predictable (McCarthy 1989). According to McDonough and Myers, many children at this stage of development have quasi-templatic constraints on the structure of words, and they conclude that therefore consonant–vowel planar segregation can be assumed. Their representation of CH is as in (10) below, and involves spreading a specified place node onto an adjacent root node unspecified for place on the consonant plane.

(10) *Consonant harmony and planar segregation*



The problem for this account is the background assumption, namely that CH is present in child language at the stage in development where the order of consonants and vowels in a word is predictable. Although children initially often do reduce the syllable structure of adult target words to simple consonant–vowel sequences, this does not necessarily happen at the time they have CH productions, as the Dutch examples in (11) show:

(11) *CV sequences and Robin's (1;9.21) consonant harmony forms*

a. CV *structure*

CVC	*niet*	/nit/	[nit]	'not'
VCC	*eend*	/ent/	[ɪnt]	'duck'
CVCC	*fiets*	/fits/	[fits]	'bicycle'
VCV	*auto*	/oto/	[oto]	'car'

b. *Consonant harmony*

schommelen	/sχɔmələ/	[vomə]	'to swing'
Grover	/χrovər/	[fofɔ]	(name)
stoep	/stup/	[fup]	'sidewalk'

The data in (11) show that the position of the vowel *vis-à-vis* the consonant is not predictable at the stage where CH forms are produced. Planar segregation can thus not be invoked either at the segmental level or at the feature level.

Locality is clearly a serious problem for accounts of CH in child language. In the literature discussed below the problem is dealt with in different ways.

3.3 *Consonant harmony as the result of a speech-processing problem*

There is no locality problem in the connectionist account of Berg and Schade (2000), since it is not a representational, but a local connectionist processing account. CH is viewed as a mispronunciation – i.e. a speech error – due to a speech plan that is carried out imperfectly. It is not, however, a low-level articulatory plan, precisely because the harmony is not co-articulatory, but involves units at a distance. The basic idea is that the level of activation differs between segments. Depending on their developmental status, links between phoneme-like units and their constituting features can be stronger or weaker. Weak links lead to hypo-activation, and hypo-activated features can be too weak to be available for production. This problem is then solved by inspecting activation levels in the network of nodes constituting a word, and picking out the element that has the highest activation level. In production, the hypo-activated feature is thus replaced by a more strongly activated feature in the word network, and this is one way in which consonant harmony can result. This can be seen as the processing version of the representational underspecification account of Stemberger and Stoel-Gammon, discussed above in §3.2.2. The other way is when a certain feature is hyper-activated because excessive weight has been attributed to the link between this feature and a segmental unit. A hyper-activated node in the network can mask the less activated nodes, leading to consonant harmony. Direction of harmony – which is usually right to left – is accounted for by self-inhibition. As soon as an onset consonant is selected, the activation level is temporarily set to zero, due to self-inhibition. The onset is thus unable to interfere with a following consonant. When production of an onset is eminent, however, the following consonant is already active due to parallel activation. Both hypo- and hyper-activated states are characteristic for a developmental system, accounting for the fact that consonant harmony is typical for child language.

Although there is no locality problem for this account formally speaking, the question remains what the effect of hyper- and hypo-activity levels of intervening vowels would be. Unfortunately, the intervening vowels are completely ignored in this account. The locality problem is formally circumvented, but in practice it is still there.

3.4 Consonant harmony as the result of an optimality-theoretic constraint

3.4.1 *Agreement*

Pater and Werle (2003) give a detailed account of the CH pattern in the longitudinal data of Trevor (Compton and Streeter 1977) within Optimality Theory (OT: McCarthy and Prince 1993; Prince and Smolensky 1993). According to Pater and Werle, consonant harmony in child language is related to place agreement in consonant clusters in adult languages (see CHAPTER 81: LOCAL ASSIMILATION). Both phenomena are due to a constraint AGREE, which requires two successive consonants to be homorganic. The domain of application differs for children and adults: in child language the successive consonants can be separated by a vowel, in adult languages the process is strictly local and only applies to adjacent consonants. Development in this view consists of narrowing down the domain in which the constraint applies to this strictly local domain.

The fact that we usually find labial and dorsal consonant harmony is independently regulated by a universal faithfulness hierarchy for place, whereby FAITH[Dors] and FAITH[Lab] are ranked above FAITH[Cor]. That is, if in order to comply with AGREE one place feature from the input form needs to be left out in the output form, it will be coronal rather than dorsal or labial. Examples from Pater and Werle illustrating this are given in (12):

(12) *Interaction of* AGREE *and* FAITH

/dɔg/	AGREE	FAITH[Dors]	FAITH[Cor]
☞ a. [gɔg]			*
b. [dɔd]		*!	
c. [dɔg]	*!		

/tap/	AGREE	FAITH[Lab]	FAITH[Cor]
☞ a. [pap]			*
b. [tat]		*!	
c. [tap]	*!		

However, since this general AGREE constraint would lead to both progressive and regressive consonant harmony, a more specific form of AGREE is invoked to capture regressive harmony, namely AGREE-L, which mentions the direction of agreement. In addition, when the regressive harmony has a specific trigger, like dorsal in the case of Trevor, this feature is mentioned in the directional AGREE constraint. In (13) the working of AGREE-L[Dors] is shown (from Pater and Werle 2003). While the general AGREE constraint would lead to harmony both in the case of *dog* and *coat*, the specific AGREE constraint affects only *dog*:

(13) AGREE-L[Dors] *for* dog *and* coat

/dɔg/	AGREE-L[Dors]	FAITH[Cor]
☞ a. [gɔg]		*
b. [dɔg]	*!	

/kot/	AGREE-L[Dors]	FAITH[Cor]
☞ a. [kot]		*
b. [kok]	*!	

Trevor's data show a developmental pattern, where Dorsal consonant harmony is initially both progressive and regressive, and Labial triggers both progressive and regressive harmony when other target consonants are coronals. Later, there is only regressive dorsal harmony. This development is captured by the demotion of markedness constraints below faithfulness constraints, the general way in which developmental changes are captured in OT (Gnanadesikan 2004; for an overview see Boersma and Levelt 2003). In this case, AGREE is demoted below FAITH[Cor], and AGREE-L[Dors] is demoted below FAITH[Lab].

A strictly local version of AGREE-L[Dors] also plays a role in Korean, where labials and coronals assimilate only regressively to dorsals, as can be seen in (14):

(14) AGREE-L[Dors] *in Korean* (de Lacy 2002, cited in Pater and Werle 2003)

a.	/əp+ko/	→	[əkko]	'bear on the back+CONJ'
	/kamki/	→	[kaŋki]	'a cold/influenza'
b.	/pat+ko/	→	[pakko]	'receive+CONJ'
	/han+kaŋ/	→	[haŋkaŋ]	'the Han river'
c.	/kot+palo/	→	[koppalo]	'straight'
	/han+bən/	→	[hambən]	'once'
d.	/paŋ+to/	→	[paŋto]	'room as well'
	/kuk+pap/	→	[kukpap]	'rice in soup'

This gives support to the analysis. However, there are also some concerns with an analysis in terms of agreement between two (non-adjacent) consonants. First, the domain of the constraint AGREE has to change in the course of development. Pater and Werle (2001) suggest that the domain for AGREE changes from "Word" in childhood to "string-adjacent consonants" in adulthood. It is not obvious how this domain change of AGREE would come about, however. A concern is the extra set of rerankings the learner would need to perform. Because of the initial Word-domain, the child's grammar has to go through different rankings in order to get the different FAITH[Place] constraints in higher positions in the constraint hierarchy than the AGREE constraint, which will over time rule out CH candidates. However, at the point where the child domain of AGREE changes into the string-adjacent adult domain, the grammar needs to undo all the rerankings of the FAITH[Place] constraints with respect to AGREE, in order to allow, or rather force, string-adjacent consonants to agree in primary place of articulation – as will often be the case in the target adult language. A second concern is the child-language-specific, non-local domain for AGREE itself. In this domain, the intervening vowel forms no obstacle for agreement between the non-adjacent consonants. This seems to imply that the solution to the locality problem of the proposal comes down to assuming that there is no locality requirement in child language.

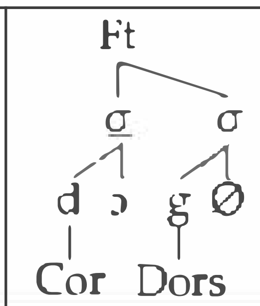
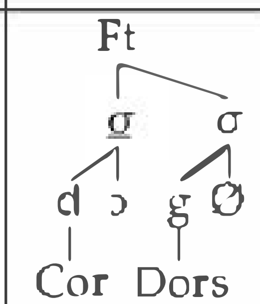
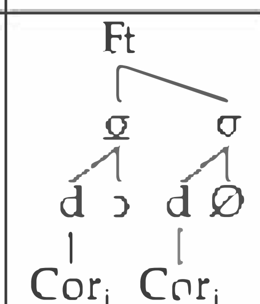
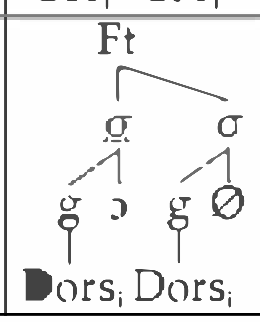
3.4.2 Licensing

Rose (2000) and Goad (1997, 2001, 2003) view CH as resulting from the relation between features and prosodically strong positions. CH is a consequence of the requirement that place features within the domain of a foot should be licensed by the foot head (CHAPTER 40: THE FOOT). Place features in prosodically weak positions can surface through being associated with, and therefore licensed by, prosodically strong positions. According to Goad, the directionality of CH follows from prosodic structure; in English, word onsets of trochaic words are prosodically strong positions and consequently they can license marked features that cannot be licensed in weak prosodic positions. A marked feature in a weak prosodic position, i.e. a coda or an intervocalic consonant, needs to be licensed by this strong onset position, resulting on the surface in regressive harmony. In Rose (2000), where both English and French child language data are analyzed, high-ranked faithfulness constraints on input place of articulation features can force the direction of harmony to go from head to dependent.

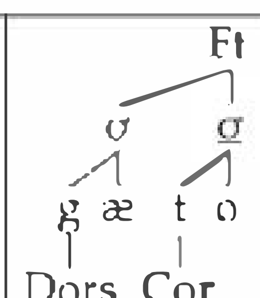
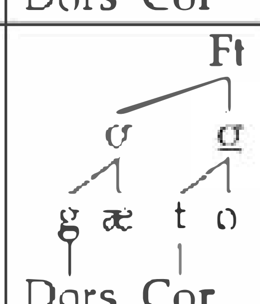
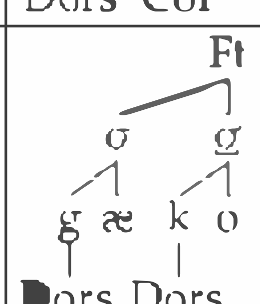
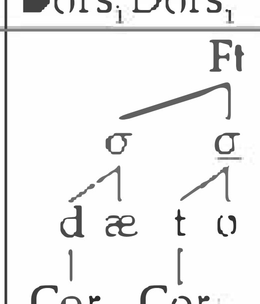
In order to circumvent the problem of crossing association lines with an intervening vowel, association with the strong position is accomplished by *melody copy*, instead of spreading: a new instance of the harmonic feature is inserted in the harmonizing position. According to Goad (2001), this makes the process similar to reduplication in mature grammars. The drive to copy a melody is different in the two systems. In languages of the world, reduplication is morphologically driven: there is a reduplicative affix that needs melodic content. In child language, however, CH is usually present before morphology kicks in, and melody copy here is driven by prosodic licensing.

A prediction is that languages with different prosodic structures exhibit different types of CH. This claim is defended by Rose (2000). English children have CH in both C_1VC_2 and C_1VC_2V words, because C_1 and C_2 are within the domain of the foot in both types of words. In contrast, for French the prediction is that CH only occurs in CVCV words, because of the claim that in French the second consonant of CVC words lies outside the foot, it being extra-prosodic. It will therefore not be involved in CH patterns. Indeed, in the data of the French subject Clara, a word like *goutte* 'drop (N)' does not undergo CH even though it has the same sequence of features, Dorsal . . . Coronal, as *gâteau* 'cake', which does undergo harmony. This is illustrated in (15)–(17), with three examples from Rose (2001):

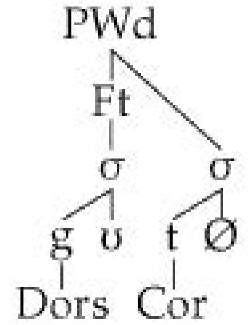
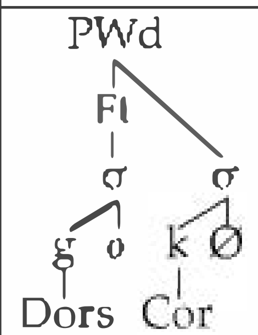
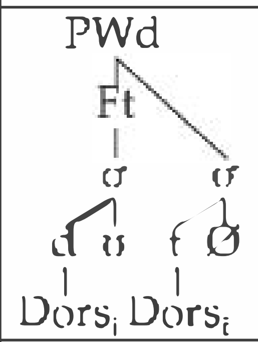
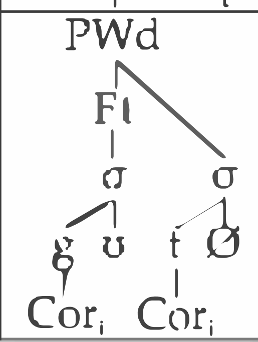
(15) *Dependent-to-head consonant harmony in English* (Coronal . . . Dorsal)

dog		MAX [Dors]	LIC (Dors,Ft)	MAX [Lab]	LIC (Lab,Ft)	MAX [Cor]	LIC (Cor,Ft)
a. [dɔg]			*!				
b. [dɔd]		*!					
☞ c. [gɔg]						*	

(16) *Head-to-dependent consonant harmony in French* (Dorsal . . . Coronal)

gâteau		MAX [Lab]	LIC (Dors,Ft)	LIC (Cor,Ft)	MAX [Cor]	MAX [Dors]	LIC (Lab,Ft)
a. [gæto]			*!				
b. [gæko]					*!		
☞ c. [dæto]						*	

(17) *No consonant harmony in CVC words in French* (Dorsal . . . Coronal)

goutte		MAX [Lab]	LIC (Dors,Ft)	LIC (Cor,Ft)	MAX [Cor]	MAX [Dors]	LIC (Lab,Ft)
a. [guk]					*!		
b. [dut]						*!	
☞ c. [gut]							

In (15) Dorsal is in the weak position of the foot, and needs to be licensed by the head position. This can be done by copying Dorsal into the head position, replacing input Coronal. Since faithfulness to input Coronal is low-ranked, this solution is optimal, and the output shows regressive Dorsal harmony. In (16), from French, Dorsal is again in the weak position and needs a licensor in the strong position of the foot. In this case, however, faithfulness to input Coronal is ranked higher than faithfulness to input Dorsal. The optimal candidate therefore does not show progressive Dorsal harmony, like the English example in (15), but regressive Coronal harmony. Finally, in (17), the second consonant is not in the foot. The licensing constraints do not apply in this case, and the optimal output candidate shows no harmony.

Although the idea of a licensing requirement seems attractive, in the end it does not seem to work. In practice, the combination of licensing and faithfulness constraints leads to a situation in which in English there is always regressive, dependent-to-head CH, while in French there is always regressive, head-to-dependent CH. The constant factor in CH, then, appears to be the regressive direction, rather than the licensing requirement. In addition, it remains unclear why the constraint ranking leading to CH forms in child language is not found in mature grammars. According to Goad, the drive to copy is different in developing and mature grammars, but as far as I can see nothing would preclude mature grammars from having a prosodic licensing drive. Finally, other French children do appear to have CH in both CVC and CVCV words (Wauquier-Gravelines 2003).

3.5 *Summary*

We have now seen that it is hard to convert the view of consonant harmony as agreement, spreading, or harmony between two non-adjacent consonants into a sound theoretical account.

We have seen three types of approach to the child-language specificness of the data: (i) the account has no child-language-specific aspects (Stemberger and Stoel-Gammon; Goad; Rose), leaving the issue unsolved; (ii) the account presents a child-language-specific aspect that could in principle also be present in mature grammars, i.e. planar segregation (McDonough and Myers), or a specific rule (Smith). Why mature grammars do not have this specific rule, or why a mature language with planar segregation probably does not have the specific type of CH we find in child language, still needs to be resolved; (iii) a formal change takes place in the grammar, which from then on precludes it from outputting CH forms (Menn; Pater and Werle; Iverson and Wheeler). This approach resolves the issue, but also introduces a new one: how does this formal change come about?

The intervening vowel, leading to the locality problem, is treated in four different ways: (i) it is not acknowledged as a problem (Menn; Smith; Iverson and Wheeler; Berg and Schade; Pater and Werle). In some cases this is because in the theoretical framework of the time consonants and vowels had different sets of place features; (ii) separate sets of place features are assumed for consonants and vowels (Stemberger and Stoel-Gammon); (iii) planar segregation is invoked to make CH a local process (McDonough and Myers); (iv) feature copying is assumed instead of feature speading (Goad; Rose). Only this last solution is able to circumvent the problem of crossing association lines in child language CH. However, crossing association lines is assumed to preclude the spreading of primary place of articulation features of consonants across vowels in mature languages. The copy solution thus solves the problem for child language, but creates a problem for the account of the absence of primary place of articulation CH in mature languages.

4 A re-analysis of consonant harmony in child language

It is clear that researchers have struggled to find an account that can explain the presence of primary place of articulation CH data exclusively in child language, without compromising the available theoretical tools.

While not abandoning these principles, there are advantages in assuming that the learner's developing phonological system differs from the adult system in certain respects (CHAPTER 101: THE INTERPRETATION OF PHONOLOGICAL PATTERNS IN FIRST LANGUAGE ACQUISITION). Levelt (1994) and Fikkert and Levelt (2008) stress the fact that consonant harmony in child language is not an isolated phenomenon (see Menn 1978 and Iverson and Wheeler 1987 for a similar view). The goal is therefore not to come up with an exclusive account of CH data, but to come up with a comprehensive account of developing place of articulation patterns in child language. When looked at in this way, it turns out that data that could be branded as instances of consonant harmony are present in child language at two different developmental stages. Since the data are, in both stages, clearly the result of a grammatical state specific to development, the fact that similar data are not found in adult languages is no longer puzzling.

In the remainder of this section, an overview of this approach is presented, focusing on the stages where so-called consonant harmony data are produced, and illustrated with longitudinal data from children acquiring Dutch (data from the CLPF database[2] and *Phon* (Rose *et al.* 2006)).

4.1 Place of articulation features

The place of articulation (PoA) features that play a role in this account are Labial, Coronal, and Dorsal. These features are monovalent and refer to both consonants and vowels (Clements 1991; Lahiri and Evers 1991; Hume 1992; Clements and Hume 1995). Thus Labial refers to both labial consonants and rounded vowels, Coronal to coronal consonants and front vowels, and Dorsal to dorsal consonants and back vowels (CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). Front rounded vowels like /y/ thus have a complex specification [Coronal, Labial], and back rounded vowels like /u/ have a complex specification [Dorsal, Labial]. The assumption here is that these vowels initially have a non-complex specification: Coronal for front vowels, and either Dorsal or Labial for back vowels. Front rounded vowels are acquired late, and are often replaced by front unrounded vowels. The low vowels /a/ and /ɑ/ are not specified for place of articulation – the idea being that place of articulation cannot be expressed in low vowels.[3]

4.2 Stage I: One word, one feature

In the CLPF corpus, in the first sets of meaningful words no combinations of different PoA features were found within words (CHAPTER 51: THE PHONOLOGICAL WORD), i.e. consonants and vowels within a word are all produced with the same PoA feature. The low vowels /a/ and /ɑ/ can be combined with either coronal, labial, or dorsal consonants. This is illustrated by data from Robin in (18):

(18) *Robin's (1;5.11) initial vocabulary*

a.	*die*	/di/	[ti]	'that one'
	huis	/hœys/	[hœys]	'house'
	niet	/nit/	[n̩t]	'not'
	thuis	/tœys/	[tœs]	'home'
	zes	/zɛs/	[sɛs]	'six'
	tiktak	/tɪktɑk/	[tita]	'tick-tock'
	aan	/an/	[an]	'on'
	daar	/dar/	[ta]	'there'
b.	*pop*	/pɔp/	[pɔ]	'doll'
	mamma	/mɑma/	[mɑma]	'mommy'
	aap	/ap/	[ap]	'monkey'

[2] CLPF database: data collected by Levelt (1994) and Fikkert (1994) of 12 children acquiring Dutch as their first language. Recordings were made every other week over a 12-month period. The database contains over 20,000 spontaneous utterances.

[3] In Dutch vowels also have a tense/lax specification, distinguishing /a e i o u/ from /ɑ ɛ ɪ ɔ u/. This specification is not relevant here.

The productions in (18a) all consist of coronal consonants (or placeless /h/) and coronal or low vowels, while the productions in (18b) have labial consonants and round or low vowels. A salient aspect of these data is that the adult target words have this same pattern. New words produced by Robin in the next two recording sessions also follow this pattern, as can be seen in (19):

(19) *New words produced by Robin (1;5.21–1;6.9)*

a.	*Coronal forms*			
	deze	/dezə/	[tis]	'this one'
	televisie	/teləvisi/	[zizi]	'television'
	trein	/trɛin/	[tin]	'train'
	ijs	/ɛis/	[æis]	'ice-cream'
	sesamstraat	/sesɑmstrat/	[zisa]	'Sesame Street'
	uit	/œyt/	[œyt]	'out'
b.	*Labial forms*			
	boom	/bom/	[bom]	'tree'
	mooi	/moi/	[bo:i]	'beautiful'
	bal	/bɑl/	[bɑo]	'ball'

This initial stage, then, can be characterized as "one word, one PoA feature." According to Levelt (1994) and Fikkert and Levelt (2008), this is caused by the fact that the initial unit for specification of PoA in the child's phonological system is the unsegmentalized word (see also, among others, Moskovitz 1971; Waterson 1971; Iverson and Wheeler 1987; de Boysson-Bardies and Vihman 1991). In Levelt (1994) early productions, like the ones in (18) and (19), are therefore represented as {WORD, Coronal} and {WORD, Labial}.

The data in (18) and (19) clearly do not resemble consonant harmony data at all. This is because the adult target words can be characterized by the same whole-word representations. Robin thus appears to select words for production that fit his phonological system. However, the data from Eva in (20) illustrate what happens in this whole-word stage when no selection takes place:

(20) *Whole-word stage: Eva (1;4.12)*

Coronal words

a.	*bed*	/bɛt/	[dɛt]	'bed'
b.	*kijk*	/kɛik/	[tɛĭt]	'look!'
c.	*prik*	/prɪk/	[tɪt]	'injection'
d.	*beer*	/ber/	[dɛ]	'bear'
e.	*dicht*	/dɪχt/	[dɪ]	'closed'
f.	*neus*	/nøs/	[nɛs]	'nose'
g.	*sleutel*	/sløtəl/	[høtœẙ]	'key'
h.	*trein*	/trɛin/	[tɛin]	'train'
i.	*eend*	/ent/	[en]	'duck'
j.	*eten*	/etə/	[etɪ]	'to eat'
k.	*konijn*	/konɛin/	[tæĭn]	'rabbit'
l.	*teen*	/ten/	[ten]	'toe'
m.	*vlinder*	/vlɪndər/	[ɪnə]	'butterfly'
n.	*auto*	/ɑuto/	[ɑŭtɑŭ]	'car'
o.	*patat*	/pa'tɑt/	[tɑt]	'French fries'

Labial words

p.	*brood*	/brot/	[bop]	'bread'
q.	*buik*	/bœyk/	[bop]	'stomach'
r.	*poes*	/pus/	[puf]	'cat'
s.	*sloffen*	/slɔfə/	[pɔfə]	'slippers'
t.	*schoenen*	/sχunə/	[umə]	'shoes'
u.	*oma*	/oma/	[oma]	'grandma'
v.	*op*	/ɔp/	[ɔp]	'on'
w.	*open*	/opə/	[opə]	'open'
x.	*aap*	/ap/	[ap]	'monkey'

Here we see that the productions both of adult target words that fit and of adult target words that do not fit a whole-word representation are represented as either {WORD, Coronal} or {WORD, Labial} in the child's phonological system. The productions in (20a)–(20c) and (20p)–(20t), of target adult forms that do not fit a whole-word representation, are forms that resemble CH forms consonants within a word that have different PoA features in the target form have identical PoA features in the produced form. However, with the possible exception of the account of Iverson and Wheeler (1987) discussed in §2.2.3, none of the above accounts can account for these forms. First, there is no fixed direction of assimilation, as illustrated below in (21a) and (21b). In (21a) a target Labial C–Coronal C combination becomes Labial C–Labial C, while the same target combination in (21b) leads to a Coronal C–Coronal C production. Second, sometimes the two "harmonized" consonants share a feature that is not present in the target form: the target Labial C–Dorsal C form in (21c) results in a Coronal C–Coronal C production.

(21) *Problems for a CH account*

a.	*brood*	/brot/	[bop]	'bread'
b.	*bed*	/bɛt/	[dɛt]	'bed'
c.	*prik*	/prɪk/	[tɪt]	'injection'

This is where the role of the target vowel becomes evident: the harmonizing feature does not originate with one of the target consonants, but with the target vowel. In case the target vowel is coronal, the child's production ends up being coronal throughout. In case the target vowel is labial, the child's production ends up being labial throughout. In (20a)–(20m) and (21b)–(21c), the target vowels are all coronal, and so are the consonants in the child's production. In (20p)–(20w) and in (21a), the target vowel is labial, and so are the produced consonants. This also accounts for the form in (20t), where the target word *schoenen*, containing the labial vowel /u/ but no labial consonants, is produced as [umə], with a labial [m]. The same applies to the form *beer* in (20d), which contains the coronal vowel /e/, but no coronal consonants, and is produced as [dɛ], with a coronal [d]. Low vowels, which do not carry a PoA specification, can be combined with either coronal consonants (20n), (20o) or labial consonants (20x).

The idea put forward in Levelt (1994) and Fikkert and Levelt (2008) is that at this developmental stage, the PoA specification of the representational unit WORD is provided by the target, non-low, vowel. In case the vowel is low, the PoA specification is provided by a consonant. This could have a perceptual background. The age group in which these homorganic productions are found, between 14 and 17 months, is the same age group that has been found to have difficulty discriminating similar-sounding words, e.g. *bin vs. din*, in perception studies (Stager and Werker 1997; Fikkert *et al.* 2003; Pater *et al.* 2004). Children in this age group have just started to build up their lexicon. It appears that at this point, the PoA information of perceptually salient segments, like vowels, can be mapped successfully onto a lexical representation, overriding the PoA information of less salient segments (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION).

To summarize, apparent CH data can be found in the initial vocabularies of children. As argued by Levelt (1994) and Fikkert and Levelt (2008), however, these data are not due to any interaction between non-adjacent consonants – hence "apparent CH" – but to the way in which early representations appear to be structured. The initial unit of specification is word-sized and carries a single PoA feature provided by the most salient segment of the target adult model, usually the vowel.

4.3 *Stage II: Overgeneralization*

After the initial stage in which words are unsegmentalized units, segmentalization gradually takes place, and at some point consonants and vowels within a word can be independently specified for place of articulation. This does not entail that all combinations of PoA features are immediately possible. Focusing on consonants, Labial + Coronal, Labial + Dorsal, and Coronal + Dorsal sequences are possible for quite some time, while Coronal + Labial and Dorsal + Labial sequences are absent from the data. It is at this stage that "typical" cases of CH start to appear in the data of some children. In this case too, it will be argued that the data are not due to a harmonic relation between non-adjacent consonants – hence the use of the quotation marks above.

(22) *"Typical" cases of consonant harmony: Robin (1;10.7)*

a.	*sop*	/sɔp/	[fɔp]	'suds'
	sloffen	/slɔfə/	[bɔfə]	'slippers'
	tafel	/tafəl/	[pafy]	'table'
	neef	/nef/	[mef]	'cousin'
	zeep	/zep/	[fep]	'soap'
b.	*klimmen*	/klɪmə/	[pɪmə]	'to climb'

At this stage in Robin's development, the PoA make-up of a produced word, given the PoA sequence of the adult target form, is completely predictable: any target word with a consonant sequence Coronal + Labial (22a) or Dorsal + Labial (22) will be produced with a consonant sequence Labial + Labial.

Interestingly, Fikkert and Levelt (2008) observe, in the longitudinal data of Robin and four other children, that these typical cases of consonant harmony only occur after a period in which attempted, non-homorganic sequences of consonants are exclusively of the types Labial + Coronal/Dorsal and Coronal + Dorsal. These attempted adult target words are always rendered faithfully with respect to the PoA structure in the children's productions. Adult target words with Coronal/

Dorsal + Labial sequences are simply not attempted in earlier stages. CH is thus an emerging phenomenon.

In Robin's data, for example, Fikkert and Levelt observe that around the age of 1;7.15 more and more adult target words with a Labial + Coronal consonant sequence, such as *bed* 'bed', *boot* 'boat', and *maan* 'moon', are attempted. Words with this PoA structure are highly frequent in Dutch. These targets are produced faithfully, i.e. with the same PoA structure. In Robin's production, then, the consonant sequence Labial + Coronal occurs frequently. One month later, in the recording at 1;8.12, the first cases of Labial CH appear. According to Fikkert and Levelt, there is a relation between these two events.

Their proposal is that, as soon as words can be segmentalized, after the initial stage in which words are unanalyzed wholes, the Dutch language learner analyzes his own active vocabulary and observes that Labial consonants are always found at the left edge of the word.[4] On the basis of this observation the learner overgeneralizes that Labial should always at least be aligned with the left edge of a word.

In OT terms this overgeneralization results in the emergence of a high-ranked constraint in the learner's grammar, requiring Labial to be aligned with the left edge of the word. This constraint is termed [Labial (Levelt 1994; Fikkert and Levelt 2008). The definition of the constraint is that an output word containing the feature Labial should always have an instance of Labial aligned with the left edge of the word. At this point, unfaithful, CH-like output forms for attempted input forms where there is no left-aligned Labial are deemed optimal by the grammar. This grammar is illustrated in (23):

(23) *Developmental grammar with emerged* [Labial

a. *poes* 'cat'

/pus/	[Labial	Faith
☞ i. pus		
ii. puf		*!

b. *tas* 'bag'

/tɑs/	[Labial	Faith
☞ i. tɑs		
ii. pɑs		*

c. *soep* 'soup'

/sup/	[Labial	Faith
☞ i. fup		*
ii. sup	*!	

[4] Even the few words in the Dutch learner's vocabulary with a Labial consonant at the right edge always have a Labial consonant at the left edge too: *pop* 'doll', *boom* 'tree', *mamma* 'mommy', *pappa* 'daddy'.

In (23a), the faithful candidate [pus] is the optimal candidate: a left-aligned instance of Labial is available without jeopardizing faithfulness. In (23b) the faithful candidate [tɑs] does not contain Labial, vacuously satisfying high-ranked [LABIAL. In (23c) the faithful candidate contains Labial, but there is no instance of a left-aligned Labial. The faithful candidate thus fatally violates high-ranked [LABIAL. The "CH" candidate [fup] does contain a left-aligned instance of Labial, at the cost of a faithfulness violation. However, faithfulness is ranked below [LABIAL, and [fup] is the winning candidate.

The very general constraint FAITH in (23) needs to be elaborated to show why other possible output candidates for input /sup/ that do not violate [LABIAL, like [sus], or metathesized [pus], are blocked (CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). In (24), a more detailed version of (23c), FAITH is split up into FAITH[Lab], requiring faithfulness to input Labial, a lower-ranked FAITH[Cor], requiring faithfulness to input Coronal, and LINEARITY, requiring faithfulness to the linear order of input segments.

(24) *Interaction of* [LABIAL *and Faithfulness constraints*

/sup/	[LABIAL	LINEARITY	FAITH[Lab]	FAITH[Cor]
☞ a. fup				*
b. sup	*!			
c. pus		*!		
d. sus			*!	

The hypothesis that CH data are due to an alignment requirement rather than to a harmony requirement is strengthened by data from other children, who appear to use metathesis in order to comply with the alignment requirement, as shown in (25):

(25) *Metathesized forms in child language*

Dutch	*kip*	/kɪp/	[pɪk]	'chicken'	(Noortje; Fikkert and Levelt 2008)
English	*sheep*		[piç]		(Alice; Jaeger 1997)
	TV		[piti]		
Spanish	*sopa*	/sopa/	[pwɔta]	'soup'	(Si; Macken 1979)
	libro	/libro/	[pɪtɔ]	'book'	

For these children, the faithfulness constraints LINEARITY and FAITH[Cor] are ranked in the opposite order from the one in (24), i.e. FAITH[Cor] >> LINEARITY. Of the output candidates complying with [LABIAL, this grammar prefers a metathesized form over a CH form. The relation between consonant harmony and metathesis is also noticed by Goad (2001) and Rose and dos Santos (2004).

To summarize, according to Levelt (1994) and Fikkert and Levelt (2008), the apparent CH data arising at this stage of development are, again, not the result of an assimilation process between non-adjacent consonants. Instead, they appear to result from the overgeneralization of a frequent PoA pattern in the active vocabulary of the learner, where Labial segments are either exclusively – or additionally, in case of Labial + Labial target words like *mommy* – found at word onsets. The pattern in the learner's vocabulary, in turn, reflects a highly frequent PoA pattern in the surrounding language.

This makes a testable prediction, namely that, depending on the distribution of PoA patterns in the language to be acquired, language learners will show different types of PoA alignment. Highly frequent patterns in the language enter the developing vocabulary first, and overgeneralization – if present – will be based on these language-specific frequent patterns. Fikkert *et al.* (2003) show that different distributions of PoA patterns in Dutch and English can indeed account for the different types of PoA-alignment data in Dutch and English child language. In English child language data, both Labial and Dorsal alignment occur, while in Dutch child language data only Labial alignment has been attested. This promising initial result should be elaborated.

Detailed studies of the development of PoA in longitudinal data of children acquiring languages other than Dutch are now needed to test the validity of the perspective discussed in this section.

5 Conclusions

Hansson (2001) tries to pull together consonant harmony in languages of the world and in child language, by pointing out that the source of these data is probably found in speech processing; consonant harmony productions are in fact phonologized speech errors. According to Hansson, the difference in place of articulation bias between adult CH and child CH is caused by the nature of the sound inventory. The sound inventory of language learners is small, and doesn't require the secondary place features that harmonize in adult languages to distinguish sounds.

Based on a detailed study of the longitudinal development of PoA patterns in child language productions, Levelt (1994) and Fikkert and Levelt (2008) come to a different conclusion. Consonant harmony forms in child language are not phonologized speech errors, but products of an immature phonological system. In the earliest stages of word production, CH-like forms result from an as yet unanalyzed representational unit Word, which can be specified for a single place of articulation feature. At a later stage, when words have segmental representations, CH-like forms result from the overgeneralization of a place of articulation pattern in a small vocabulary.

Instead of pulling together the two types of data, i.e. child language data and cross-linguistic data, then, it might be better to pull them apart, by providing them with different terms. *Consonant harmony* is a phenomenon that can be found in the world's languages. In child language there is an initial stage where words are *homorganic*, while *PoA alignment* occurs in later stages.

With this new perspective on child language data, it is no longer necessary to wonder why neither homorganic data nor PoA-alignment data are found in languages of the world: both types of data are due to specific developmental states of the phonological system that, because of fundamental changes, are transient and therefore no longer accessible in mature grammars. One developmental state is characterized by the initial unsegmentalized word-sized unit in early phonological representations. Later, these word-sized units are, irrevocably, replaced by segment-sized units. In the other developmental state the learner generalizes over his own – still small – vocabulary. The PoA structure of words in this developing vocabulary reflects a highly frequent pattern in the surrounding language. PoA-alignment data occur when this pattern is overgeneralized, leading to the emergence of a constraint in the developing grammar. Overgeneralization is typical for child language, and disappears when enough experience with other data is gathered. With the expansion of the learner's vocabulary, more and more evidence will be available that Labial is not necessarily always left-aligned. On the basis of this counterevidence, the learner will conclude that [Labial should not be an active constraint in the grammar. The constraint will disappear, or be demoted to the lowest ranks of the grammar, and candidates formerly called consonant harmony forms will never again be optimal.

ACKNOWLEDGMENTS

I would like to thank the editors for their invitation to contribute to the *Companion to Phonology*, and two anonymous reviewers for their detailed and helpful comments on the manuscript. Preparation of this chapter was made possible by a grant from the Netherlands Organization for Scientific Research (NWO Vidi grant 276-75-006).

REFERENCES

Andersen, Torben. 1988. Consonant alternation in the verbal morphology of Päri. *Afrika und Übersee* 71. 63–113.

Applegate, Richard B. 1972. Ineseño Chumash grammar. Ph.D. dissertation, University of California, Berkeley.

Archangeli, Diana & Douglas Pulleyblank. 1987. Maximal and minimal rules: Effects of tier scansion. *Papers from the Annual Meeting of the North East Linguistic Society* 17. 16–35.

Berg, Thomas & Ulrich Schade. 2000. A local connectionist account of consonant harmony in child language. *Cognitive Science* 24. 123–149.

Boersma, Paul & Clara C. Levelt. 2003. Optimality Theory and phonological acquisition. *Annual Review of Language Acquisition* 3. 1–50.

Boysson-Bardies, Bénédicte de & Marilyn Vihman. 1991. Adaption to language: Evidence from babbling and first words in four languages. *Language* 67. 297–319.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. 1991. Place of articulation in consonants and vowels: A unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5. 77–123.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 245–306. Cambridge, MA & Oxford: Blackwell.

Compton, Arthur & Mary Streeter. 1977. Child phonology: Data collection and preliminary analyses. In Eve T. Clark & Pamela Tiedt (eds.) *Papers and Reports on Child Language Development* 7. Stanford: Department of Linguistics, Stanford University.

Cruttenden, Alan. 1978. Assimilation in child language and elsewhere. *Journal of Child Language* 5. 373–378.

de Lacy, Paul. 2002. The formal expression of markedness. Ph.D dissertation, University of Massachusetts, Amherst.

Fikkert, Paula. 1994. On the acquisition of prosodic structure. Ph.D. dissertation, University of Leiden.

Fikkert, Paula & Clara C. Levelt. 2008. How does Place fall into place? The lexicon and emergent constraints in children's developing grammars. In Peter Avery, B. Elan Dresher & Keren Rice (eds.) *Contrast in phonology: Theory, perception, acquisition*, 231–268. Berlin & New York: Mouton de Gruyter.

Fikkert, Paula, Clara C. Levelt & Joost van de Weijer. 2003. Input, intake and phonological development: The case of consonant harmony. Paper presented at the Generative Approaches to Language Acquisition conference, Utrecht.

Gnanadesikan, Amalia E. 2004. Markedness and faithfulness constraints in child phonology. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in phonological acquisition*, 73–108. Cambridge: Cambridge University Press.

Goad, Heather. 1997. Consonant harmony in child language: An optimality-theoretic account. In S. J. Hannahs & Martha Young-Scholten (eds.) *Focus on phonological acquisition*, 113–142. Amsterdam & Philadelphia: John Benjamins.

Goad, Heather. 2001. Assimilation phenomena and initial constraint ranking in early grammars. In Anna H.-J. Do, Laura Domínguez & Aimee Johansen (eds.) *Proceedings of the 25th Annual Boston University Conference on Language Development*, 307–318. Somerville, MA: Cascadilla Press.

Goad, Heather. 2003. Licensing and directional asymmetries in consonant harmony. Poster presented at the Child Phonology Conference, University of British Columbia.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT. Published 1979, New York: Garland.

Hansson, Gunnar Ólafur. 2001. Theoretical and typological issues in consonant harmony. Ph.D. dissertation, University of California, Berkeley.

Hume, Elizabeth. 1992. Front vowels, coronal consonants and their interaction in non-linear phonology. Ph.D. dissertation, Cornell University. Published 1994, New York: Garland.

Iverson, Gregory K. & Deirdre Wheeler. 1987. Hierarchical structures in child phonology. *Lingua* 73. 243–257.

Jaeger, Jeri J. 1997. How to say "Grandma" and "Grandpa": A case study in early phonological development. *First Language* 17. 1–29.

Lahiri, Aditi & Vincent Evers. 1991. Palatalization and coronality. In Paradis & Prunet (1991), 79–100.

Levelt, Clara C. 1994. On the acquisition of place. Ph.D. dissertation, University of Leiden.

Macken, Marlys A. 1979. Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua* 49. 11–49.

McCarthy, John J. 1989. Linear order in phonological representation. *Linguistic Inquiry* 20. 71–99.

McCarthy, John J. & Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University.

McDonough, Joyce & Scott Myers. 1991. Consonant harmony and planar segregation in child language. Unpublished ms., University of California, Los Angeles & University of Texas, Austin.

Menn, Lise. 1978. Phonological units in beginning speech. In Alan Bell & Joan B. Hooper (eds.) *Syllables and segments*, 315–334. Amsterdam: North-Holland.

Moskowitz, Marlene. 1971. The acquisition of phonology. Ph.D. dissertation, University of California, Berkeley.

Paradis, Carole & Jean-François Prunet (eds.) 1991. *The special status of coronals: Internal and external evidence*. San Diego: Academic Press.

Pater, Joe & Adam Werle. 2001. Typology and variation in child consonant harmony. In Caroline Féry, Antony Dubach Green & Ruben van de Vijver (eds.) *Proceedings of HILP 5*, 119–139. Potsdam: University of Potsdam.

Pater, Joe & Adam Werle. 2003. Direction of assimilation in child consonant harmony. *Canadian Journal of Linguistics* 48. 385–408.

Pater, Joe, Christine Stager & Janet F. Werker. 2004. The perceptual acquisition of phonological contrasts. *Language* 80. 384–402.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Rose, Yvan. 2000. Headedness and prosodic licensing in the L1 acquisition of phonology. Ph.D. dissertation, McGill University.

Rose, Yvan. 2001. Licensing and feature interaction processes in child language. *Proceedings of the West Coast Conference on Formal Linguistics* 20. 484–497.

Rose, Yvan & Christophe dos Santos. 2004. The prosodic basis of consonant harmony and metathesis. Paper presented at the 3rd Phonological Acquisition Workshop, Nijmegen.

Rose, Yvan, Brian MacWhinney, Rod Byrne, Gregory Hedlund, Keith Maddocks, Philip O'Brien & Todd Wareham. 2006. Introducing Phon: A software solution for the study of phonological acquisition. In David Bamman, Tatiana Magnitskaia & Colleen Zaller (eds.) *Proceedings of the 30th Annual Boston University Conference on Language Development*, 489–500. Somerville, MA: Cascadilla Press.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.

Smith, Neil V. 1973. *The acquisition of phonology: A case study.* Cambridge: Cambridge University Press.

Stager, Christine & Janet F. Werker. 1997. Infants listen for more phonetic detail in speech perception than in word learning tasks. *Nature* 388. 381–382.

Stemberger, Joseph P. & Carol Stoel-Gammon. 1989. Underspecification and consonant harmony in child phonology. Unpublished ms., University of Minnesota.

Stemberger, Joseph P. & Carol Stoel-Gammon. 1991. The underspecification of coronals: Evidence from language acquisition and performance errors. In Paradis & Prunet (1991), 181–199.

Stoel-Gammon, Carol & Joseph P. Stemberger. 1994. Consonant harmony and phonological underspecification in child speech. In Mehmet Yavas (ed.) *First and second language phonology*, 63–80. San Diego: Singular Publishing Group.

Vihman, Marilyn M. 1978. Consonant harmony: Its scope and function in child language. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 2: *Phonology*, 281–334. Stanford: University Press.

Waterson, Natalie. 1971. Child phonology: A prosodic view. *Journal of Linguistics* 7. 179–211.

Wauquier-Gravelines, Sophie. 2003. Troncation et reduplication: Peut-on parler de gabarits morphologiques dans le lexique précoce!? In Bernard Fradin, Georgette Dal, Nabil Hathout, Françoise Kerleroux, Michel Roché & Marc Plénat (eds.) *Les unités morphologiques*, 76–84. Lille: University of Lille III.

73 Chain Shifts

Anna Łubowicz

1 What is a chain shift?

This chapter provides background on chain shift mappings. In a phonological chain shift, underlying /A/ maps onto surface [B] and underlying /B/ maps onto surface [C] in the same context but, crucially, underlying /A/ does not become surface [C]. Thus a chain shift has a standard representation A → B → C (see Ultan 1970; Kenstowicz and Kisseberth 1979; Labov 1994; Kirchner 1996; Parkinson 1996; Gnanadesikan 1997; Dinnsen and Barlow 1998; McCarthy 1999; Moreton and Smolensky 2002; amongst others).

The Finnish vowel shift (McCawley 1964; Lehtinen 1967; Keyser and Kiparsky 1984; Anttila 1995, 2002a, 2002b; Karlsson 1999; Harrikari 2000) provides an example. In Finnish, before the plural suffix *-i* (and before the past tense marker *-i*), long low vowels shorten (/aa/ → [a]), short low vowels undergo rounding and raising (/a/ → [o]), and short round vowels surface unchanged (/o/ → [o]). Thus we have the following chain shift effect:

(1) *Finnish chain shift*

aa → a → o ↻

Some examples are given in (2).

(2)			*sing nom*	*plural essive*	
	a.	/aa/ → [a]	*maa*	*ma-i-na*	'earth'
			vapaa	*vapa-i-na*	'free'
	b.	/a/ → [o]	*kissa*	*kisso-i-na*	'cat'
			vapa	*vapo-i-na*	'fishing rod'
	c.	/o/ → [o]	*talo*	*talo-i-na*	'house'
			pelko	*pelko-i-na*	'fear'

The key issue is that in Finnish, forms with underlying long low vowels shorten but do not round (/aa/ → [a], *[o]), but forms with underlying short low vowels undergo rounding in the same context (/a/ → [o]).

Chain shifts have been found in diverse areas, including dialectal variation (Labov *et al.* 1972; Labov 1994; Labov *et al.* 2006), language acquisition (Smith 1973; Braine 1976; Macken 1980; Dinnsen and Barlow 1998; Dinnsen *et al.* 2001; Jesney 2007), synchronic phonology (Rubach 1984; Hayes 1986; Clements 1991; Kirchner 1996; Parkinson 1996; McCarthy 1999; Moreton 2004; van Oostendorp 2004), and diachronic phonology (Bauer 1979, 1992; Lass 1999; Schendl and Ritt 2002; Minkova and Stockwell 2003; Ahn 2004).

This chapter focuses on synchronic chain shifts. A compendium of synchronic chain shifts has been compiled by Moreton (2004), and a corpus of synchronic chain shifts can be also found in Moreton and Smolensky (2002). It is important to note that some diachronic chain shifts have also been given a synchronic analysis, as in Miglio and Morén (2003).[1]

Many types of chain shifts have been described in the literature (see references in §2). There is debate on whether the description of chain shift types is accurate. This chapter will address the typology of chain shift mappings in the context of various theoretical proposals. It will describe the types of chain shifts found in the literature, and address the validity of this typology under different analyses.

Two background assumptions are made in this chapter. First, it is assumed that to describe the typology of chain shifts, it is important not only to provide an empirically correct analysis of a chain shift, but also to explain *why* chain shifts exist in phonology. That is, uncovering the motivation for chain shift mappings is essential to gain a full understanding of the genesis and acquisition of chain shifts.

The second background assumption is that the typology of chain shifts can be (largely) described in terms of markedness, where markedness motivates a phonological process (see §4). Since, as will be shown, analyses based solely on markedness fail to account for the attested types of chain shifts, it is suggested that something in addition to markedness drives chain shifts. A possible solution to this problem is provided in the form of an analysis with contrast (see §3.3), where a phonological process can take place to preserve contrast as well as satisfying markedness. This idea gives rise to potential avenues of productive research on chain shifts and phonology in general.[2]

The rest of this chapter is organized as follows. §2 introduces the typology of chain shifts. §3 describes several analyses of chain shift mappings. Finally, §4 explores implications of the various analyses for the typology of chain shifts.

2 The typology

Chain shift mappings can be categorized by type of segment (see §2.1) and mapping (see §2.2) involved in the shift.[3]

[1] For a discussion of diachronic processes in OT, see Holt (2003). See also CHAPTER 93: SOUND CHANGE.

[2] I would like to thank an anonymous reviewer for comments on this point.

[3] There are other possible ways to characterize chain shifts that are not discussed here: the number of steps involved in the shift (Gnanadesikan 1997), the trajectories of changes (Labov 1994), the extent of mergers (near mergers *vs.* full mergers) (Parkinson 1996), and the location of the mapping (the segment *vs.* the environment) (Łubowicz 2004).

2.1 *Segment type*

Both vowels and consonants can be involved in a shift. Some examples of vowel height chain shifts come from Bassá (Bantu; Schmidt 1996), Gbanu (Niger-Congo; Bradshaw 1996), Kikuria (Bantu; Chacha and Odden 1998), Lena Spanish (Hualde 1989), Nzɛbi (Clements 1991), and Servigliano Italian (Kaze 1989); see also CHAPTER 21: VOWEL HEIGHT; CHAPTER 110: METAPHONY IN ROMANCE. These are mostly mappings involving raising (Parkinson 1996). Some examples of consonantal chain shifts come from Southern Paiute (Sapir 1930; McLaughlin 1984), Toba Batak (Hayes 1986), Estonian (Ultan 1970), Finnish (Ultan 1970), and Irish (CHAPTER 117: CELTIC MUTATIONS; Ní Chiosáin 1991). These are mostly mappings involving lenition along either the voicing or consonantal stricture scale (Gnanadesikan 1997) (see also CHAPTER 13: THE STRICTURE FEATURES; CHAPTER 65: CONSONANT MUTATION; CHAPTER 66: LENITION). See (3) and (4), respectively.

(3) *Vowel shifts*

a. New Zealand English (Labov 1994) æ → e → i → ɨ
b. Nzɛbi (Bantu; Clements 1991) a → ɛ → e → i; ɔ → o → u

(4) *Consonantal shifts* (Ultan 1970)

a. Southern Paiute (Uto-Aztecan; Sapir 1930) pp → p → v
b. Toba Batak (Austronesian; Hayes 1986) np → pp → ʔp

A comprehensive account of vowel and consonant shifts is provided by Gnanadesikan (1997). The crux of Gnanadesikan's proposal is ternary scales that explain the type of segments involved in a chain shift mapping.

2.2 *Mapping type*

Another criterion for categorizing chain shifts, and the one that is the focus of this chapter, is the type of the mapping involved. This includes pull shifts, push shifts, circular shifts, and what I will refer to as regular shifts (these will be discussed in §2.2.4). All four kinds of chain shifts have been proposed in the literature. They are described below.

2.2.1 *Push shifts*

Assume that /A/ maps onto [B] (A → B) and /B/ maps onto [C] (B → C), but, crucially, /A/ does not become [C]. Thus there is a chain shift effect of the form A → B → C. One type of a chain shift is a push shift. In a phonological push shift, the latter step in the shift, /B/ → [C], is a consequence of the prior step, /A/ → [B], and not an independently motivated phonological process (see Martinet 1952, 1955; Labov 1994; Schendl and Ritt 2002; Miglio and Morén 2003; Ahn 2004; Maclagan and Hay 2004; Hsieh 2005; Barrie 2006; amongst others).

Some examples of push shifts described in the literature include the Swedish shift (Benediktsson 1970; Labov 1994): a → aː → ɔː → oː → uː → ʉː; the New Zealand shift (Bauer 1979, 1992; Trudgill *et al.* 1998; Gordon *et al.* 2004; Maclagan and Hay 2004): æ → e → i → ei/ɨ (but see Labov 1994); the Northern Cities shift (Labov 1994): ɛ → ʌ → ɔ; the Great Vowel Shift in English (Luick 1914; Jespersen

1949; Miglio and Morén 2003; Minkova and Stockwell 2003): ɛː → eː → iː → ai; ɔː → oː → uː → au; the Short Vowel Shift in Early Modern English (Lass 1999; Schendl and Ritt 2002): u → o → ɔ → ɒ; a → æ; and tone sandhi in Xiamen, a dialect of the Min language of the Sino-Tibetan family (Barrie 2006; Chen 1987) (see §4).

The basic observation is that the latter mapping takes place as a consequence of the prior mapping (hence the *push* effect). Push shifts are further described in §4.1, using the example of tone sandhi. §4.1 also evaluates various theoretical proposals with respect to push shifts.

2.2.2 *Pull shifts*

Another kind of shift found in the literature is a pull shift, also known as a drag shift (King 1969). Pull shifts are the opposite of push shifts. In a pull shift, the prior mapping in the shift, /A/ → [B], takes place as a consequence of the latter mapping, /B/ → [C], and is not an independently motivated phonological process (hence the *pull* effect).

Some examples of pull shifts reported in the literature include diachronic changes involved in the Lettish and Lithuanian Chain Shift (Endzelin 1922; Labov 1994: 134): eː → iə → iː → ij and oː → uə → uː → uw; the North Frisian Chain Shift (Labov 1994: 136): iː → i → a; uː → u → a; the Middle Korean Vowel Shift (Labov 1994: 139): e → ə → ɨ → u → ɔ → a; and the Northern Cities Shift (Labov 1994: 195): ɔ → ɑ → æ → ɪə; i → e → æ.

The arguments for pull shifts often involve historical evidence, whereby the latter mapping in the shift, /B/ → [C], historically precedes the prior mapping, /A/ → [B] (Labov 1994). It remains a question whether pull shifts are possible synchronically. This chapter will show that synchronic pull shifts are not admitted under any of the theoretical proposals to be described (see §4.2).

2.2.3 *Circular shifts*

In a circular shift, mappings form a circle or a semi-circle. Exchange processes (Anderson and Browne 1973) are examples of circular shifts: /A/ → [B] and /B/ → [A]. Circular shifts are often seen as morphologically conditioned (Anderson and Browne 1973; Moreton 1996; Alderete 2001a, 2001b; Horwood 2001).

Some examples of circular shifts, or exchange mappings, are consonantal polarity in Luo (Gregersen 1972; Okoth-Okombo 1982; Alderete 2001a), plural formation in Diegueño verbs (Langdon 1970; Walker 1970), vowel shift in Brussels Flemish (Zonneveld 1976; but see Moreton 1996), and tonal circles, also known as tone sandhi, in Xiamen (Cheng 1968, 1973; Yip 1980; Chen 1987; Moreton 1996; Hsieh 2005; Barrie 2006).[4]

It remains a question whether circular shifts have been described accurately in the literature. For example, it has been proposed that some circular chain shifts are conditioned morphologically, rather than being phonological (also CHAPTER 103: PHONOLOGICAL SENSITIVITY TO MORPHOLOGICAL STRUCTURE, but see Crowhurst 2000). Circular shifts are further discussed in §4.3.

2.2.4 *Regular shifts*

Finally, in what I will refer to as a regular shift, both mappings in the shift take place independently in the language, but form a chain shift when put together.

[4] For additional examples see Moreton (1996: 21).

For example, in Sea Dayak (Kenstowicz and Kisseberth 1979) there is a chain shift that includes cluster simplification (consonant deletion) and vowel nasalization: ŋga → ŋa → ŋã. Both nasalization and cluster simplification are found independently in the language, but nasalization fails in the context of consonant deletion. A similar process interaction is found in Ulu Muar Malay (Hendon 1966; Lin, forthcoming). Regular shifts are discussed further in §4.4.

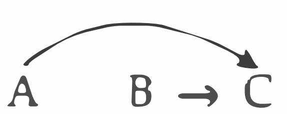
The following sections will address the validity of the typology of chain shifts described above.

3 Theoretical accounts of chain shifts

Chain shifts are a type of opaque process (Kiparsky 1973; Rubach 1984). Chain shifts are opaque because in a phonological chain shift one (or more) of the phonological processes simply do not apply as they should, based on the phonology of the language. Depending on the perspective taken by the analyst, in a chain shift mapping a phonological process underapplies (it does not apply where expected) or overapplies (it applies where it should not). Both underapplication and overapplication have been shown to be examples of opaque processes (Benua 1997; McCarthy 1999, 2003a, 2003b). Underapplication and overapplication are also known as counterfeeding and counterbleeding opacity, respectively (Kiparsky 1973). In the following discussion, opaque processes are considered with respect to synchronic phonologies.

Assuming a phonological chain shift A → B → C, the key property of chain shifts is that in the same context /A/ and /B/ map onto different outputs: /A/ goes to [B] and /B/ goes to [C]. This is unexpected. One might expect the two inputs to map onto the same output in one and the same grammar.[5]

(5) *Expected mapping*



A theoretical account of chain shifts has been sought for a long time. The rest of this section describes several accounts of chain shifts found in the literature, such as rule ordering (§3.1), local conjunction (§3.2), and Optimality Theory with contrast (§3.3). The predictions of these approaches will be compared in §4.[6]

3.1 Rule ordering

In rule-based phonology (Kenstowicz and Kisseberth 1979; Rubach 1984), chain shifts are most commonly accounted for by rule ordering. In a phonological chain shift A → B → C, it is proposed that /B/ → [C] precedes /A/ → [B]. Thus, derived [B]s do not map onto [C]. This is illustrated by the following derivation. There are two rules that apply in a fixed order: B → C / __ D precedes A → B / __ D. This ordering results in a chain shift effect.

[5] Another possibility would be for /B/ → [C] not to occur, so both /A/ and /B/ would map onto [B].
[6] Chain shifts have also been accounted for by underspecification in rule-based phonology. The underspecification account of chain shifts is not discussed here but see Kiparsky (1993).

(6) *Chain shift in rule ordering*

Input	/AD/	/BD/
Rule 1: B → C / __ D	does not apply	CD
Rule 2: A → B / __ D	BD	does not apply
Output	[BD]	[CD]

In Sea Dayak (Kenstowicz and Kisseberth 1979: 298, 308), there is a rule of nasalization which specifies that a vowel nasalizes immediately after a nasal, and a rule of nasal cluster simplification that deletes an obstruent following a nasal. Kenstowicz and Kisseberth propose that the rule of nasalization precedes the rule of nasal cluster simplification. Therefore nasalization fails in forms where the consonant deletes. Due to the specific rule ordering, the form /naŋga/ 'set up a ladder' maps onto [nãŋaʔ], while /naŋa/ 'straighten' maps onto [nãŋãʔ], with glottal insertion in both forms.

(7) *Sea Dayak in a rule ordering analysis*

Input	/naŋga/	/naŋa/
Vowel nasalization	nãŋga	nãŋã
Nasal cluster simplification	nãŋa	does not apply
Output	[nãŋaʔ]	[nãŋãʔ]

The predictions of rule ordering for the typology of chain shifts will be discussed in §4.

3.2 Local conjunction

There are different ways to account for chain shifts in Optimality Theory (OT) (Prince and Smolensky 1993; Kager 1999; McCarthy 2002). This section describes a common way of accounting for chain shifts in OT, "local constraint conjunction." Other OT approaches to chain shifts include more general accounts of opacity in OT, such as Sympathy Theory (McCarthy 1999, 2003a), Stratal or Derivational OT (Kiparsky 2000; Rubach 2003; Bermúdez-Otero 2007), output–output correspondence (Benua 1997; Burzio 1998), targeted constraints (Wilson 2001), comparative markedness (McCarthy 2003b), turbidity (Goldrick and Smolensky 1999), gestural coordination theory (Lin, forthcoming), and candidate chain theory (McCarthy 2007).

According to Gnanadesikan (1997) and Kirchner (1996), the solution to chain shift mappings in OT lies in an enriched theory of faithfulness (CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). Both researchers propose special types of faithfulness constraints that block two-step movements like /aai/ → [oi], thereby accounting for the discrepancy in phonological mappings between identical derived and underlying segments. Kirchner uses locally conjoined faithfulness constraints (Smolensky 1993, 1997), whereas Gnanadesikan distinguishes between classical IDENT-type constraints and novel IDENTADJACENT-type constraints on some scale of similarity.

Local conjunction is defined in (8) (Smolensky 1993).

(8) The Local Conjunction of C_1 and C_2 in domain D, $C_1 \&_D C_2$, is violated when there is some domain of type D in which both C_1 and C_2 are violated.

If the domain of local conjunction is a segment, both C_1 and C_2 cannot be violated together in the same segment.

The following tableaux give a schematic overview of the local conjunction analysis of chain shifts. A phonological process, /B/ → [C], applies only if there is no double violation of faithfulness in the same segment, $[F(A \rightarrow B) \& F(B \rightarrow C)]_{Seg}$.[7]

(9) *Chain shifts in local conjunction*

No violation of local conjunction (/B/ → [C])

/BD/	$[F(A\rightarrow B) \& F(B\rightarrow C)]_{Seg}$	*BD	F(B→C)	F(A→B)
a. BD		*!		
☞ b. CD			*	

Violation of local conjunction (/A/ → [B], *[C])

/AD/	$[F(A\rightarrow B) \& F(B\rightarrow C)]_{Seg}$	*BD	F(B→C)	F(A→B)
☞ a. BD		*		*
b. CD	*!		*	*

Local conjunction blocks /A/ from mapping all the way onto [C] (see the second tableau in (9)). In other words, /B/ → [C] is blocked for underlying [A]s, but applies to underlying [B]s (compare the two tableaux in (9)).

Applied to Sea Dayak (shown in (7)), the local conjunction analysis would be as follows:[8]

(10) *Sea Dayak in a local conjunction analysis*

No violation of local conjunction

/naŋa/	$[\text{IDENT[nas]} \& \text{MAX}]_{AdjSeg}$	*NV	IDENT[nas]	MAX
a. nàŋa		*!	*	
☞ b. nāŋā			**	

Violation of local conjunction

/naŋga/	$[\text{IDENT[nas]} \& \text{MAX}]_{AdjSeg}$	*NV	IDENT[nas]	MAX
☞ a. nàŋa		*	*	*
b. nāŋā	*!		**	*

[7] Faithfulness constraints in the conjunction $[F(A\rightarrow B) \& F(B\rightarrow C)]_{Seg}$ refer to the dimensions of similarity between segments and do not require that [B] is visible in the output. As suggested by an editor of the *Companion to Phonology*, it can be rewritten as $[F(A\rightarrow B) \& F(A\rightarrow C)]_{Seg}$. The difference lies in the formulation of the latter faithfulness constraint, F(B→C) *vs.* F(A→C). For notational convenience, I have chosen to refer to the latter conjunct as F(B→C).

[8] The constraints are as follows: *NV (no nasal followed by a non-nasal vowel), MAX (no deletion), IDENT[nas] (no change in nasality compared with the input), and $[\text{IDENT[nas]} \& \text{MAX}]_{AdjSeg}$ (no change in nasality and deletion in adjacent segments).

Due to local conjunction, there is no nasalization in the form that undergoes cluster simplification (see the second tableau in (10)). However, nasalization applies in the form with no cluster simplification (see the first tableau).

3.3 *OT with contrast*

An alternative explanation for chain shifts in OT starts from the observation that chain shifts always preserve one underlying contrast at the expense of neutralizing another underlying contrast (Łubowicz 2003). In Finnish (shown in (1) and (2)), the contrast between underlying /aai/ and /ai/, originally one of length, is preserved, albeit in a different form – as a rounding contrast (underlying /aai/ *vs.* /ai/, surface [ai] *vs.* [oi]). The contrast between underlying /ai/ and /oi/, the original rounding contrast, is lost (both become [oi]). Thus:

(11)	*Input*		*Output*
	length contrast	→	rounding contrast
	rounding contrast	→	neutralized

This is referred to as *contrast transformation.*[9]

This observation has given rise to Preserve Contrast (PC) theory (Łubowicz 2003, 2004, 2007, forthcoming).[10] The key idea is that contrast exists as an imperative in a phonological system formalized as a family of rankable and violable constraints which demand that contrast be preserved, "PC constraints" (cf. Flemming 1995, 2004; Padgett 1997, 2003; Padgett and Zygis 2007; amongst others). PC constraints demand that pairs of words that contrast underlyingly with respect to a given phonological property P contrast on the surface (not necessarily with respect to P). Such constraints are defined in (12).

(12) PC(P)

For each pair of inputs contrasting in P that map onto the same output in a scenario, assign a violation mark. Formally, assign one mark for every pair of inputs, in_a and in_b, if in_a has P and in_b lacks P, $in_a \rightarrow out_k$, and $in_b \rightarrow out_k$. (Informally, if inputs are distinct in P, they need to remain distinct in the output (not necessarily in P).)[11]

P is a potentially contrastive phonological property, such as a distinctive feature, length, stress, or presence *vs.* absence of a segment. The properties P, then, are essentially the same as the properties governed by faithfulness constraints in standard OT. Indeed, PC(P) constraints are like faithfulness constraints in that they look at two levels of representation. But they are novel in that they evaluate contrasts for pairs of underlying words and corresponding output words instead

[9] Earlier works on contrast include Kaye (1974, 1975), Gussmann (1976), and Kisseberth (1976).

[10] Other works on PC theory include Tessier (2004), Barrie (2006), Flack (2007), and Riggs (2008). Similar ideas are expressed in the Dispersion Theory of Contrast (Flemming 1995, 2004; Padgett 1997, 2003; Itô and Mester 2004; Bradley 2001; Padgett and Zygis 2007).

[11] What it means to contrast in P is defined as follows: a pair of inputs, in_a and in_b, contrast in P when corresponding segments in those inputs, seg_a and seg_b, are such that seg_a has P and seg_b lacks P.

of evaluating individual input–output mappings. To evaluate constraints on contrast, candidates must be sets of mappings, called *scenarios*.

There are also generalized faithfulness constraints in PC theory that will become important in the discussion of circular shifts in §4.3.

The following tableau shows how PC theory accounts for chain shifts. The tableau compares three scenarios: a chain shift scenario, an identity scenario where all inputs map onto identical outputs, and a transparent scenario with no /B/ → [C]. The constraint ranking proposed below captures the observation that the initial mapping in the shift, /A/ → [B], is due to markedness, but the subsequent mapping, /B/ → [C], is facilitated by contrast preservation. The relevant ranking is: *A, PC(A/B) >> PC(B/C).[12]

(13) *Chain shifts under contrast*

	Scenarios	*A	PC(A/B)	PC(B/C)
☞ a. Chain shift	/A/ → [B] /B/ → [C] /C/ → [C]			*
b. Identity	/A/ → [A] /B/ → [B] /C/ → [C]	*!		
c. Transparent	/A/ → [B] /B/ → [B] /C/ → [C]		*!	

The identity scenario, scenario (13b), loses because it violates markedness *A. The transparent scenario, scenario (13c), neutralizes contrast, thus violating PC(A/B). The chain shift scenario, scenario (13a), wins under the proposed constraint ranking. In the chain shift scenario, the /B/ → [C] mapping applies to preserve contrast between /A/ and /B/.[13]

The predictions of PC theory will be further discussed in the following section.[14]

4 Implications for the typology of chain shifts

The goal of this section is to examine what chain shifts are predicted to occur under various approaches. Chain shifts will be subdivided on the basis of the relationship between various stages in the shift. This typology can be referred to as *teleological*, because it differentiates reasons as to why a chain shift takes place, rather than merely describing the chain shift. This view is not novel, and is also

[12] It is important to point out that not all chain shifts are push shift mappings in this approach to phonology. PC theory can also describe the "regular" shifts (see §4.4).

[13] A competing derived environment scenario where A → C and B → B incurs the same violations of PC and markedness constraints as the winning chain shift scenario. It has been proposed that the two scenarios would be differentiated in the second stage of Eval by generalized faithfulness constraints (Łubowicz 2003).

[14] For a comprehensive discussion of PC theory and its predictions, see Łubowicz (2003, forthcoming).

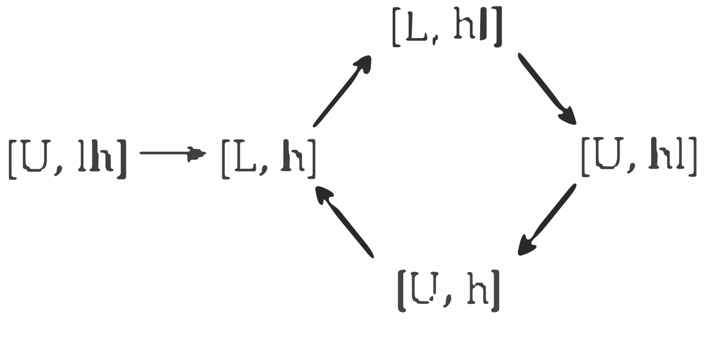
taken by Optimality Theory. Rule-based approaches are non-teleological, since rules do not look for the reason why a phonological process takes place. Predictions of both rule-based approaches and OT approaches to chain shifts are described below.[15]

4.1 Push shifts

Recall that in push shift mappings (see §2.2.1), the latter mapping in the shift is a consequence of the prior mapping and not an independently motivated phonological process. To argue that A → B → C is a push shift mapping, one must show that there is no independent motivation for the latter mapping in the shift, B → C (in OT, no markedness constraint that favors [C] over [B]). In addition, one must establish that the mapping /B/ → [C] is always linked with /A/ → [B].

The tonal system of Xiamen, a dialect of the Min language of the Sino-Tibetan family (see Chen 1987; Barrie 2006; CHAPTER 107: CHINESE TONE SANDHI) provides an example of a push shift mapping. The diagram below (14) uses the following notation: [U] is upper register, [L] lower register, [lh] rising pitch, and [hl] falling pitch. Thus [U, lh], for example, is a high-rising tone, where the pitch moves from the low end of the upper register to the high end.

(14) *Xiamen tone sandhi*



Barrie (2006) observes that the tonal shift above is an example of a push shift, because the latter mapping in the shift, [L, h] → [L, hl], is a result of the prior mapping, [U, lh] → [L, h]. Specifically, the initial mapping in the shift, [U, lh] → [L, h], is due to markedness improvement, such as avoidance of rising tones (*[U, lh]). However, the subsequent mapping, [L, h] → [L, hl], does not improve on markedness, since it creates a contour tone from an input level tone, and research on tonal markedness has shown that contour tones are more marked than level tones (Alderete 2001a; Yip 2004; CHAPTER 45: THE REPRESENTATION OF TONE). Barrie proposes that the latter mapping in the shift, [L, h] → [L, hl], is a result of the initial mapping, [U, lh] → [L, h], and the need to maintain contrast between tones of various registers, [U] *vs.* [L]. In other words, due to the chain shift effect, the input tones [U, lh] and [L, h] map onto [L, h] and [L, hl], respectively. Otherwise, if there was no shift both input tones would map onto the same output, [L, h]. The remaining mappings in the shift are analyzed in a similar manner (for a complete analysis see Hsieh 2005 and Barrie 2006).

Push shifts are problematic for rule-based approaches to phonology. In rule-based approaches, chain shifts are accounted for by rule ordering (see §3.1). But in a push shift mapping, there is no separate rule that accounts for the latter

[15] Thanks to a reviewer for comments on this point.

mapping in the shift. If there was a separate rule for the latter mapping in the shift, it would be predicted that the process would occur outside of the chain shifting context. But in a push shift, the latter process is always linked with the prior process, and does not apply independently. In consequence, synchronic push shifts cannot be described under a rule-based analysis.

Similarly, push shifts are problematic to OT without contrast (see §3.2). This approach does not admit push shifts, because in OT, a phonological process can only apply if there is a high-ranked markedness constraint against it (Moreton 1996). But in a push shift, /B/ → [C] is not due to markedness. In Xiamen (see (14)), both the initial mapping in the shift, [U, lh] → [L, h], and the subsequent mapping, [L, h] → [L, hl], would need to be forced by markedness constraints. The problem here is that [L, h] → [L, hl] cannot be accounted for by markedness improvement. Thus, OT without contrast would not account for push shifts, because in a push shift mapping there is no high-ranked markedness constraint that triggers the latter mapping in the shift.[16]

OT with contrast (see §3.3), on the other hand, proposes a solution to push shift mappings which makes use of contrast as an independent principle in a phonological system. The key observation is that the initial mapping in the shift, /A/ → [B], takes place to improve on markedness, but the latter mapping in the shift, /B/ → [C], is due to contrast preservation. In tone sandhi (shown in (14)) OT with contrast proposes that the latter mapping in the shift, [L, h] onto [L, hl], takes place to preserve contrast between tones of various registers, [U] *vs.* [L] (Barrie 2006). Thus, OT with contrast, unlike other approaches, admits push shifts. Depending on further evidence on push shift mapping, OT with contrast is promising in this respect.

4.2 *Pull shifts*

Another type of chain shift found in the literature is a pull shift mapping. In a pull shift mapping, also known as a drag shift, the prior mapping in the shift, /A/ → [B], takes place because the latter mapping, /B/ → [C], occurs (King 1969). In OT, the prior mapping in the pull shift does not occur due to markedness. Unlike push shifts, pull shift mappings are ruled out in all approaches to chain shifts discussed so far.

Rule-based approaches do not allow pull shifts, because they require that there is a separate rule that accounts for each mapping in the shift (see §3.1). In a pull shift mapping, however, there is no separate rule that accounts for the initial mapping in the shift. Rather, the initial mapping in the shift is always linked with the latter mapping, and does not occur independently. Thus, pull shifts are predicted not to occur by rule ordering.[17]

[16] In response to a reviewer's comments, it is important to point out that additional mechanisms in OT such as strata, candidate chains, sympathy, etc. do not admit push shifts. All these models *require* markedness to force a phonological process, and in a push shift markedness is not enough. As discussed here, by the introduction of contrast constraints into Con, a possible account of push shifts is provided.

[17] A reviewer points out that a rule-based approach to phonology is inherently non-teleological, unlike OT, and thus would not differentiate chain shifts based on what is the cause of a particular process, such as pull shifts *vs.* push shifts.

In OT without contrast (see §3.2), similarly, pull shifts cannot be described. The only way to obtain a phonological mapping in standard OT is by markedness improvement (Moreton 1996). But in a pull shift mapping, it has been proposed that the initial mapping in the shift cannot be explained by markedness improvement and thus is not predicted to occur.

OT with contrast (see §3.3) also rules out pull shifts. In OT with contrast, a phonological mapping can take place to improve on markedness or to preserve contrast. However, as will be shown below, in a pull shift mapping there is no improvement on either markedness or contrast.

Consider A → B → C as an example of a pull shift. The mapping /B/ → [C] is forced by markedness, but the initial mapping /A/ → [B] is not. This is illustrated in the following tableau, which compares a no shift scenario, scenario (15a), to a pull shift scenario, scenario (15b), under the constraint ranking PC(A/B), *B >> PC(B/C). The crucial point is that there is no markedness constraint that compels the /A/ → [B] mapping.

(15) *Pull shifts cannot be described in OT with contrast*

	Scenarios	PC(A/B)	*B	PC(B/C)
☞ a. No shift A B → C	/A/ → [A] /B/ → [C] /C/ → [C]			*
b. Pull shift A → B → C	/A/ → [B] /B/ → [C] /C/ → [C]		*!	*

Both scenarios preserve the contrast between /A/ and /B/, because underlying /A/ and /B/ map onto different outputs in both scenarios. Both scenarios map /B/ onto [C] in accordance with the ranking of the markedness constraint *B above PC(B/C). But the pull shift scenario, scenario (15b), incurs a fatal violation of *B, and is thus harmonically bounded by the no shift scenario, (15a).[18] Thus, in OT with contrast, pull shifts are predicted not to occur.[19]

In summary, synchronic pull shifts cannot be described in any of the approaches described above. Diachronic pull shifts are different, as they can be seen as different processes that apply at different stages in the development of the language (Holt 2003).

4.3 *Circular shifts*

Another type of chain shift found in the literature is a circular shift, also known as an exchange process (Anderson and Browne 1973). Examples of circular shifts

[18] A harmonically bounded scenario incurs a superset of violation marks in comparison to a competing scenario. This is under the assumption that there is no other constraint in Con that favors scenario (15b) over (15a).

[19] Though not allowing for pull shifts, OT with contrast admits a sequence of changes that *resemble* a pull shift effect. Take a situation where /A/ → [B] "wants to happen" (due to *A) but is blocked by /B/ → [B] (to avoid neutralization). Then /B/ → [C] comes along. Now the /A/ → [B] map can emerge. This scenario is what I refer to as a regular shift. It is not a pull shift, because there is a markedness constraint *A that drives /A/ → [B].

were given in §2.2.3. There is a debate in the literature on whether circular shifts exist and to what extent they are conditioned by morphology. Many of the circular shifts have been reanalyzed (see Moreton 1996) or, as we saw in Xiamen (see (14)), do not constitute a complete circle, as they contain a termination (neutralization) point.

Rule-based approaches can describe circular shifts with no termination point (see McCarthy 1999).[20] Unlike rule-based approaches, both standard OT (see §3.2) and the contrast approach (see §3.3) rule out circular shifts with no termination point, for example /A/ → [B] and /B/ → [A]. As shown by Moreton (1996), a circular shift is not admitted to non-contrast OT, since it does not improve on markedness.

As will be shown below, another feature that makes circular shifts with no termination point incompatible with OT with contrast is the fact that they do not improve on markedness or contrast, and involve an unmotivated violation of generalized faithfulness. The following tableau compares a circular shift scenario, scenario (16a), to an identity scenario, scenario (16b).

(16) *No circular shifts: Unmotivated violations of faithfulness*

	Scenarios	PC	Markedness	Faith
a. Circular shift A B ↑↓ ↑↓ D C	/A/ → [D] /B/ → [C] /C/ → [B] /D/ → [A]		[A] [B] [C] [D]	****! /A/ → [D] /B/ → [C] /C/ → [B] /D/ → [A]
☞ b. Identity ↻ ↻ A B ↻ ↻ D C	/A/ → [A] /B/ → [B] /C/ → [C] /D/ → [D]		[A] [B] [C] [D]	

The circular shift is harmonically bounded by the competing identity scenario. Both scenarios satisfy PC constraints and incur the same violations of markedness (they have the same outputs). But the circular shift scenario, scenario (16a), is ruled out by generalized faithfulness. There are unmotivated violations of faithfulness in this scenario.[21]

Circular shifts *with a termination point*, for example /A/ → [B], /B/ → [C], /C/ → [B], are predicted to occur in OT with contrast, since the initial step in the shift improves on markedness (see §4.1). A circular shift *with a termination point* refers to a shift where one of the original inputs is never used as the output. In the example cited here, the input /A/ is never used as the output in the shift. The tone sandhi analyzed in Barrie (2006) and Hsieh (2005) is of that form. It is important to note that such a scenario would not be harmonically bounded by a no shift scenario, /A/ → [B], /B/ → [B], /C/ → [C], because the two scenarios differ as to whether they preserve contrast between underlying /A/ and /B/.

[20] McCarthy (1999) characterizes these shifts as involving the "multi-process Duke of York gambit."

[21] Some instances of circular shifts are also ruled out by relational PC (see Łubowicz 2003).

To sum up, circular shifts are not admitted to OT, with or without contrast. But they are predicted to occur in rule-based approaches.

4.4 *Regular shifts*

Finally, there are regular shifts where each mapping occurs independently, but one of the mappings is blocked when they co-occur (see §2.2.4). All approaches to chain shifts predict regular shifts. Rule-based approaches predict regular shifts by rule ordering (see §3.1). Standard OT describes regular shifts by blocking an otherwise regular phonological process, for example by special faithfulness constraints (see §3.2). Similarly, OT with contrast (see §3.3) accounts for regular shifts by blocking an otherwise regular phonological process by PC constraints, as shown below.

Consider a regular chain shift A → B → C, where each step in the shift is forced by markedness constraints. But the process /B/ → [C] is blocked for underlying /A/ to maintain the contrast between underlying /A/ and /B/. The ranking and the relevant tableau are given below. A no shift scenario, scenario (18a), where both /A/ and /B/ map onto [C], is compared with a regular shift, scenario (18b).

(17) *The role of contrast*

PC(A/B) >> *B >> PC(B/C) (also *A >> *B)

(18) *Regular shifts exist*

	Scenarios	PC(A/B)	*B	PC(B/C)
a. No shift A B → C	/A/ → [C] /B/ → [C] /C/ → [C]	*!		*
☞ b. Regular shift A → B → C	/A/ → [B] /B/ → [C] /C/ → [C]		*	*

Scenario (18a) loses since it merges the contrast between underlying /A/ and /B/, thus violating PC(A/B). Scenario (18b) wins because in this scenario /A/ and /B/ map onto distinct outputs. This is at the expense of violating the markedness constraint *B. Thus, OT with contrast predicts regular shifts to occur.

To summarize, this section has contributed to the debate on what chain shift mappings are possible. It has provided an analysis on what types of chain shifts are predicted to occur under different theoretical proposals. OT with contrast has been shown to predict push shift mappings which are attested, but cannot be described under any of the other analyses.

5 Summary

This chapter has provided a typology of chain shift mappings in the context of a number of theoretical approaches to chain shifts. Chain shifts have puzzled researchers for a long time. Many types of chain shifts have been described in the literature (see §2) and numerous theoretical approaches have been put forward to account for them (see §3). The goal of this chapter has been to evaluate the various approaches to chain shifts and compare their implications for the typology of chain shift mappings (see §4).

Four types of chain shifts have been examined: push shifts, pull shifts, circular shifts, and regular shifts. Out of those, regular shifts have been the easiest to account for (see §4.4). Circular shifts are only allowed in rule-based phonology (see §4.3). Pull shifts are left with no straightforward phonological explanation (see §4.2). Push shifts are only admitted by OT with contrast (see §4.1). Depending on further evidence on push shifts, the contrast approach seems to be promising in this respect and differs crucially from other existing accounts of chain shifts in the predictions which it makes.

ACKNOWLEDGMENTS

I would like to thank Marc van Oostendorp, Beth Hume, and two anonymous reviewers for their comments and valuable suggestions.

REFERENCES

Ahn, Sang-Cheol. 2004. Towards the optimal account of diachronic chain shifts. *Studies in Phonetics, Phonology and Morphology* 10. 43–67.

Alderete, John. 2001a. *Morphologically governed accent in Optimality Theory*. New York & London: Routledge.

Alderete, John. 2001b. Dominance effects as transderivational anti-faithfulness. *Phonology* 18. 201–253.

Anderson, Stephen R. & Wayles Browne. 1973. On keeping exchange rules in Czech. *Papers in Linguistics* 6. 445–482.

Anttila, Arto. 1995. Deriving variation from grammar: A study of Finnish genitives. Unpublished ms., Stanford University (ROA-63).

Anttila, Arto. 2002a. Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20. 1–42.

Anttila, Arto. 2002b. Variation and phonological theory. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.) *The handbook of language variation and change*, 206–243. Oxford: Blackwell.

Barrie, Michael. 2006. Tone circles and contrast preservation. *Linguistic Inquiry* 37. 131–141.

Bauer, Laurie. 1979. The second Great Vowel Shift? *Journal of the International Phonetic Association* 9. 57–66.

Bauer, Laurie. 1992. The second Great Vowel Shift revisited. *English World-Wide* 18. 253–268.

Benediktsson, Hreinn (ed.) 1970. *The Nordic languages and modern linguistics*. Reykjavik: Societas Scientarium Islandica.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Ph.D. dissertation, University of Massachusetts, Amherst. Published 2000 as *Phonological relations between words*. New York: Garland.

Bermúdez-Otero, Ricardo. 2007. Morphological structure and phonological domains in Spanish denominal derivation. In Fernando Martinez-Gil & Sonia Colina (eds.) *Optimality-theoretic studies in Spanish phonology*, 278–311. Amsterdam & Philadelphia: John Benjamins.

Bradley, Travis G. 2001. The phonetics and phonology of rhotic duration contrast and neutralization. Ph.D. dissertation, Pennsylvania State University (ROA-473).

Bradshaw, Mary M. 1996. One-step raising in Gbanu. *Ohio State Working Papers in Linguistics* 48. 1–11.

Braine, Martin D. S. 1976. Review of Smith (1973). *Language* 52. 489–498.

Burzio, Luigi. 1998. Multiple correspondence. *Lingua* 104. 79–109.

Chacha, Chacha Nyaigotti & David Odden. 1998. The phonology of vocalic height in Kikuria. *Studies in African Linguistics* 27. 129–158.

Chen, Matthew Y. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4. 109–149.

Cheng, Richard L. 1968. Tone sandhi in Taiwanese. *Linguistics* 41. 19–42.

Cheng, Richard L. 1973. Some notes on tone sandhi in Taiwanese. *Linguistics* 100. 5–25.

Clements, G. N. 1991. Vowel height assimilation in Bantu languages. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 17. 25–63.

Crowhurst, Megan J. 2000. A flip-flop in Sirionó (Tupian): The mutual exchange of /i i/. *International Journal of American Linguistics* 66. 57–75.

Dinnsen, Daniel A. & Jessica Barlow. 1998. On the characterization of a chain shift in normal and delayed phonological acquisition. *Journal of Child Language* 25. 61–94.

Dinnsen, Daniel A., Kathleen O'Connor & Judith Gierut. 2001. The puzzle-puddle-pickle problem and the Duke-of-York gambit in acquisition. *Journal of Linguistics* 37. 503–525.

Endzelin, J. 1922. *Lettische Grammatik*. Heidelberg: Carl Winter.

Flack, Kathryn. 2007. Ambiguity avoidance as contrast preservation: Case and word order freezing in Japanese. *University of Massachusetts Occasional Papers in Linguistics* 32. 57–88.

Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles.

Flemming, Edward. 2004. Contrast and perceptual distinctness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 232–276. Cambridge: Cambridge University Press.

Gnanadesikan, Amalia E. 1997. Phonology with ternary scales. Ph.D. dissertation, University of Massachusetts (ROA-195).

Goldrick, Matthew & Paul Smolensky. 1999. Opacity and turbid representations in Optimality Theory. Paper presented at the 35th Annual Meeting of the Chicago Linguistic Society.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury & Peter Trudgill. 2004. *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press.

Gregersen, Edgar A. 1972. Consonant polarity in Nilotic. In Erhard Voeltz (ed.) *3rd Annual Conference on African Linguistics*, 105–109. Bloomington: Indiana University.

Gussmann, Edmund. 1976. Recoverable derivations and phonological change. *Lingua* 40. 281–303.

Harrikari, Heli. 2000. Segmental length in Finnish: Studies within a constraint-based approach. Ph.D. dissertation, University of Helsinki.

Hayes, Bruce. 1986. Assimilation as spreading in Toba Batak. *Linguistic Inquiry* 17. 467–499.

Hendon, Rufus S. 1966. *The phonology and morphology of Ulu Muar Malay*. Harvard: Yale University.

Holt, D. Eric (ed.) 2003. *Optimality Theory and language change*. Dordrecht: Kluwer.

Horwood, Graham. 2001. Anti-faithfulness and subtractive morphology. Unpublished ms., Rutgers University (ROA-466).

Hsieh, Feng-fan. 2005. Tonal chain-shifts as anti-neutralization-induced tone sandhi. In Sudha Arunachalam, Tatjana Scheffler, Sandhya Sundaresan & Joshua Tauberer (eds.) *Proceedings of the 28th Annual Penn Linguistics Colloquium*, 99–112. Philadelphia: Department of Linguistics, University of Pennsylvania.

Hualde, José Ignacio. 1989. Autosegmental and metrical spreading in the vowel-harmony systems of northwestern Spain. *Linguistics* 27. 773–805.

Itô, Junko & Armin Mester. 2004. Morphological contrast and merger: Ranuki in Japanese. *Journal of Japanese Linguistics* 20. 1–18.

Jesney, Karen. 2007. Child chain shifts as faithfulness to input prominence. In Alyona Belikova, Luisa Meroni & Mari Umeda (eds.) *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition, North America*, 188–199. Somerville, MA: Cascadilla Press.

Jespersen, Otto. 1949. *A Modern English grammar on historical principles*. Part I: *Sounds and spellings*. London: George Allen & Unwin.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Karlsson, Fred. 1999. *Finnish: An essential grammar*. 2nd edn. London & New York: Routledge.

Kaye, Jonathan. 1974. Opacity and recoverability in phonology. *Canadian Journal of Linguistics* 19. 134–149.

Kaye, Jonathan. 1975. A functional explanation for rule ordering in phonology. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 11. 244–252.

Kaze, Jeffrey W. 1989. Metaphony in Italian and Spanish dialects revisited. Ph.D. thesis, University of Illinois.

Kenstowicz, Michael & Charles W. Kisseberth. 1979. *Generative phonology: Description and theory*. New York: Academic Press.

Keyser, Samuel J. & Paul Kiparsky. 1984. Syllable structure in Finnish phonology. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*, 7–31. Cambridge, MA: MIT Press.

King, Robert D. 1969. Push chains and drag chains. *Glossa* 3. 3–21.

Kiparsky, Paul. 1973. Phonological representations: Abstractness, opacity, and global rules. In Osamu Fujimura (ed.) *Three dimensions of linguistic theory*, 57–86. Tokyo: Taikusha.

Kiparsky, Paul. 1993. Blocking in nonderived environments. In Sharon Hargus & Ellen M. Kaisse (eds.) *Studies in Lexical Phonology*, 277–313. San Diego: Academic Press.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–365.

Kirchner, Robert. 1996. Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry* 27. 341–350.

Kisseberth, Charles W. 1976. The interaction of phonological rules and the polarity of language. In Andreas Koutsoudas (ed.) *The application and ordering of phonological rules*, 41–54. The Hague: Mouton.

Labov, William. 1994. *Principles of linguistic change*, vol. 1: *Internal factors*. Oxford: Blackwell.

Labov, William, Sharon Ash & Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology and sound change*. Berlin & New York: Mouton de Gruyter.

Labov, William, Malcah Yaeger & Richard Steiner. 1972. *A quantitative study of sound change in progress*. Philadelphia: U.S. Regional Survey.

Langdon, Margaret. 1970. *A grammar of Diegueño: The Mesa Grande dialect*. Berkeley: University of California Press.

Lass, Roger. 1999. Phonology and morphology. In Richard M. Hogg (ed.) *The Cambridge history of the English language*, vol. 3: *1476–1776*, 56–186. Cambridge: Cambridge University Press.

Lehtinen, Meri. 1967. *Basic course in Finnish*. Bloomington: Indiana University Publications.

Lin, Hua. Forthcoming. Apparent deletion in a grammar of gestural coordination. Ph.D. dissertation, University of Southern California.

Łubowicz, Anna. 2003. Contrast preservation in phonological mappings. Ph.D. dissertation, University of Massachusetts, Amherst (ROA-554).

Łubowicz, Anna. 2004. Counter-feeding opacity as a chain shift effect. *Proceedings of the West Coast Conference on Formal Linguistics* 22. 315–327.

Łubowicz, Anna. 2007. Paradigmatic contrast in Polish. *Journal of Slavic Linguistics* 15. 229–262.
Łubowicz, Anna. Forthcoming. *The phonology of contrast*. London: Equinox.
Luick, K. 1914. *Historische Grammatik der englischen Sprache*. 2 vols. Stuttgart: Tauchnitz.
Macken, Marlys A. 1980. The child's lexical representation: The *"puzzle-puddle-pickle"* evidence. *Journal of Linguistics* 16. 1–17.
Maclagan, Margaret & Jennifer Hay. 2004. The rise and rise of New Zealand English DRESS. In S. Cassidy, F. Cox, R. Mannell & S. Palethorpe (eds.) *Proceedings of the 10th Australian International Conference on Speech Science and Technology*, 183–188. Sydney: Australian Speech Science and Technology Association.
Martinet, André. 1952. Function, structure, and sound change. *Word* 8. 1–32.
Martinet, André. 1955. *Économie des changements phonétiques*. Berne: Francke.
McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16. 331–399.
McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.
McCarthy, John J. 2003a. Sympathy, cumulativity, and the Duke-of-York gambit. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*, 23–76. Cambridge: Cambridge University Press.
McCarthy, John J. 2003b. Comparative markedness. *Theoretical Linguistics* 29. 1–51.
McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.
McCawley, James D. 1964. The morphophonemics of the Finnish noun. Unpublished ms., MIT.
McLaughlin, John E. 1984. A revised approach to Southern Paiute phonology. *Kansas Working Papers in Linguistics* 9. 47–79.
Miglio, Viola & Bruce Morén. 2003. Merger avoidance and lexical reconstruction. In Holt (2003), 191–228.
Minkova, Donka & Robert Stockwell. 2003. English vowelshifts and "optimal" diphthongs: Is there a logical link? In Holt (2003), 169–190.
Moreton, Elliott. 1996. Non-computable functions in Optimality Theory. Unpublished ms., University of Massachusetts, Amherst (ROA-364).
Moreton, Elliott. 2004. A compendium of synchronic chain shifts. Unpublished ms., University of North Carolina, Chapel Hill. Available (June 2010) at www.unc.edu/~moreton/Materials/Chainshifts.pdf.
Moreton, Elliott & Paul Smolensky. 2002. Typological consequences of local constraint conjunction. *Proceedings of the West Coast Conference on Formal Linguistics* 21. 306–319.
Ní Chiosáin, Máire. 1991. Topics in the phonology of Irish. Ph.D. dissertation, University of Massachusetts, Amherst.
Okoth-Okombo, Duncan. 1982. *Dholuo morphophonemics in a generative framework*. Berlin: Dietrich Reimer.
Oostendorp, Marc van. 2004. Phonological recoverability in dialects of Dutch. Unpublished ms., Meertens Instituut. Available (June 2010) at www.vanoostendorp.nl/pdf/recoverable.pdf.
Padgett, Jaye. 1997. Candidates as systems: Saussure lives! Paper presented at the Hopkins Optimality Theory Workshop/University of Maryland Mayfest 1997.
Padgett, Jaye. 2003. Contrast and post-velar fronting in Russian. *Natural Language and Linguistic Theory* 21. 39–87.
Padgett, Jaye & Marzena Zygis. 2007. The evolution of sibilants in Polish and Russian. *Journal of Slavic Linguistics* 15. 291–324.
Parkinson, Frederick B. 1996. The representation of vowel height in phonology. Ph.D. dissertation, Ohio State University.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Riggs, Daylen. 2008. Contrast preservation in the Yupik languages. *Proceedings of the Western Conference on Linguistics* 18. 217–234.

Rubach, Jerzy. 1984. *Cyclic and Lexical Phonology: The structure of Polish.* Dordrecht: Foris.

Rubach, Jerzy. 2003. Polish palatalization in Derivational Optimality Theory. *Lingua* 113. 197–237.

Sapir, Edward. 1930. Southern Paiute, a Shoshonean language. *Proceedings of the American Academy of Arts and Sciences* 65. 1–296.

Schendl, Herbert & Nikolaus Ritt. 2002. Of vowel shifts great, small, long and short. *Language Sciences* 24. 409–421.

Schmidt, Deborah. 1996. Vowel raising in Basaa: A synchronic chain shift. *Phonology* 13. 239–267.

Smith, Neil V. 1973. *The acquisition of phonology: A case study.* Cambridge: Cambridge University Press.

Smolensky, Paul. 1993. Harmony, markedness, and phonological activity. Paper presented at the Rutgers Optimality Workshop 1, Rutgers University (ROA-87).

Smolensky, Paul. 1997. Constraint interaction in generative grammar II: Local conjunction or random rules in Universal Grammar. Paper presented at the Hopkins Optimality Theory Workshop/University of Maryland Mayfest 1997.

Tessier, Anne-Michelle. 2004. Input "clusters" and contrast preservation in OT. *Proceedings of the West Coast Conference on Formal Linguistics* 23. 759–772.

Trudgill, Peter, Elizabeth Gordon & Gillian Lewis. 1998. New dialect formation and Southern Hemisphere English: The New Zealand short front vowels. *Journal of Sociolinguistics* 2. 35–51.

Ultan, Russell. 1970. Some sources of consonant gradation. *Working Papers on Language Universals* 2. 1–30.

Walker, Douglas C. 1970. Diegueño plural formation. *Linguistic Notes from La Jolla* 4.

Wilson, Colin. 2001. Consonant cluster neutralisation and targeted constraints. *Phonology* 18. 147–197.

Yip, Moira. 1980. The tonal phonology of Chinese. Ph.D. dissertation, MIT.

Yip, Moira. 2004. Phonological markedness and allomorph selection in Zahao. *Language and Linguistics* 5. 969–1001.

Zonneveld, Wim. 1976. An exchange rule in Flemish Brussels Dutch. *Linguistic Analysis* 2. 109–114.

74 Rule Ordering

Joan Mascaró

1 The bases of rule ordering

The distributional properties of sound in natural languages are explained by appeal to a level of underlying structure in addition to the level of observed phonetic or surface representation (CHAPTER 1: UNDERLYING REPRESENTATIONS), and to a function that maps underlying representations into surface representations. This function has been conceived since the beginning of generative grammar as an ordered set of rules. In this chapter I will first introduce the main properties of rule ordering and the arguments for ordering rules (§1), and I will review various proposals to modify rule ordering in early generative phonology (§2), including cyclic ordering (§3). In §4 I discuss feeding, bleeding, and similar interactions in more detail, §5 discusses serial ordering and parallel approaches, and §6 draws some conclusions.

A rule expresses a significant generalization about the sound structure of a given natural language. The rules of generative phonology, as formalized in Chomsky and Halle (1968; *SPE*) and subsequent work, were formalized adaptations of descriptive statements about phonology of earlier frameworks, even though their function was not the same. Both the relationship of generative rules to statements of descriptive grammars and the reasons for imposing ordering on them can be gathered from the following example, taken from Halle (1962: 57–58). (1a)–(1d) are taken from the description of Sanskrit vowel sandhi in Whitney (1889). The rules in (1e)–(1h) are a formalization of the corresponding generative rules. For simplification, in (1e)–(1h) I have included only the rules that apply to front vowels.

(1) a. Two similar simple vowels, short or long, coalesce and form the corresponding long vowel. (§126)

b. An *a*-vowel combines with a following *i*-vowel to *e*; with a *u*-vowel, to *o*. (§127)

c. The *i*-vowels, the *u*-vowels, and the *r* before a dissimilar vowel or a diphthong are each converted into its own corresponding semi-vowel, *j* or *v* or *r*. (§129)

d. Of a diphthong, the final *i*- or *u*-element is changed into its corresponding semi-vowel *j* or *v* before any vowel or diphthong: thus *e* (really *ai* . . .) becomes *aj*, and *o* (that is *au* . . .) becomes *av*. (§131)

e. $V_iV_j \rightarrow V_i$: $V_i = V_j$
f. ai → e
g. i → j / __ V_i $V_i \neq$ i
h. i → j / V_i __ V_j $V_j \neq$ i

The similarity of the rules to the descriptive statements is obvious. But, as Halle notices, if the ordering (e)–(g)–(f) is imposed on the rules in (1), "significant simplifications can be achieved." A similar comment is made by Chomsky and Halle (1968: 18) with respect to ordering: "it [is] possible to formulate grammatical processes that would otherwise not be expressible with comparable generality." Indeed, the condition on dissimilarity of (1g) can be eliminated, since when (1g) applies, all similar VV sequences will have coalesced by the application of (1e). Moreover, (1h) can be dispensed with, because V*i*V sequences will not be turned into *e*V by (If), since (1g) will have changed the vowel into a glide. We find here one of the main reasons for imposing ordered rules: ordering allows for simplification of grammars and for a better expression of linguistically significant generalizations.

Another typical argument in favor of rule ordering is language variation. Since *SPE* relates underlying and surface representations via a set of ordered rules, it follows that language variation must be due to differences in underlying representations, in the set of rules and in their ordering. A famous example of difference in grammars stemming from different orderings of the same rules is Canadian Raising, an example introduced in Halle (1962: 63–64), based on data from Joos (1942), which is also discussed in Chomsky and Halle (1968: 342).[1]

In certain Canadian and US dialects the first elements in the diphthongs /aɪ aʊ/ are raised to [ʌɪ ʌʊ] before voiceless consonants.[2] At the same time there is regular change of /t/ to [ɾ] in the American English flapping environment. The interaction of these phenomena gives different results in two dialects, A and B. This causes, according to Joos, a dilemma: in a word like *writer*, which is pronounced [rʌɪɾɚ] in dialect A, Joos's generalization that "/a/ is a lower-mid vowel ... [only] in diphthongs followed by fortis [≈ voiceless] consonants" is not true – and in Joos's view, descriptive statements are about surface representations, hence true of surface representations. Halle's solution to the dilemma stems from the recognition that statements of regularities ("rules") should be true of steps in the derivation, but need not be true of surface representations. This is the case if rules are ordered, and hence the application of a later rule can change the context that conditioned an earlier rule, as in this case, or the result of the rule itself. In other words, rule ordering solves Joos's dilemma. (2) shows the derivation of *typewriter* with the diphthong /aɪ/ both before non-flapping (/p/) and flapping (/t/) environments in both dialects. The statement in (2c) (also in (2f)) is true of surface representations (2d) and (2h), but the rule in (2b) (also in (2g)) is true of (2h), but not of (2d), which contains the sequence [ʌɪd], if we

[1] The opaque case (dialect A) had already been discussed by Harris (1951: 70–71) and Chomsky (1962: 156–157).

[2] I transcribe the first vowel of the diphthong as [ʌ], and the voiced *t* as [ɾ], following Chambers (1973, 2006); Joos's phonetic description is slightly different (basically [ɐ] and [d], respectively). Canadian Raising has generated a great deal of discussion. Kaye (1990) casts some doubts on the existence of dialect B, which are not clearly formulated. Mielke *et al.* (2003) claim that the difference has been phonemicized, e.g. as /nʌjf/ *vs.* /najv/, but Idsardi (2006) argues convincingly that there are actual alternations.

interpret the rule in the sense that "/ᴀɪ ᴀʊ/ appear [phonetically] only before voiceless consonants." In (2) I simplify the flapping context to V__V.

(2) *Dialect A*

a.		/taɪpraɪtə/
b.	a → ᴀ / __ [C, −voice]	tᴀɪprᴀɪtə
c.	t → ɾ / V __ V	tᴀɪprᴀɪɾə
d.	*output*	[tᴀɪprᴀɪɾə]

Dialect B

e.		/taɪpraɪtə/
f.	t → ɾ / V __ V	taɪpraɪɾə
g.	a → ᴀ / __ [C, −voice]	tᴀɪpraɪɾə
h.	*output*	[tᴀɪpraɪɾə]

Another example of the same argument for imposing ordering on rules, grammars differing only in rule ordering, is examined in Kiparsky (1982b: 65–66). German devoices obstruents in coda position (3a) (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION) and simplifies /ŋg/ clusters to [ŋ] (3b). Two of the inflectional forms of the adjective meaning 'long', *lang* and *lange*, contrast in two dialect groups, one showing [laŋk], [laŋə], the other [laŋ], [laŋə], respectively. Application of any of the two rules renders the other rule inapplicable (a case of mutual bleeding, see §2), therefore only the first rule applies in each ordering in every instance:

(3) a. Devoicing $[+\text{obstr}] \rightarrow [-\text{voice}] \;/\; \_\_ \begin{Bmatrix} +\text{C} \\ \# \end{Bmatrix}$

b. *g*-deletion g → Ø / [+nasal] __

c. *Dialect group I*

	/laŋg/	/laŋg+ə/
Devoicing	laŋk	—
g-deletion	—	laŋə

d. *Dialect group II*

	/laŋg/	/laŋg+ə/
g-deletion	laŋ	laŋə
Devoicing	—	—

Rule ordering is closely connected to rule application. As shown by Whitney's example, descriptive grammars and many versions of structuralist phonology implicitly assume simultaneous rule application (see Postal 1968: 140–152). This follows from the assumption that rules (or descriptive statements) are true of surface representation, i.e. they are generalizations about surface representation. In simultaneous rule application, the string is scanned for the structural description of each rule and all the rules whose structural description is met apply simultaneously. Chomsky and Halle (1968: 19) provide an interesting abstract example of simultaneous application, which is compared to rule ordering.[3] I adapt

[3] Chomsky and Halle's (1968) example consists of the rules B → X / __ Y and A → Y / __ X and the input representations /ABY/ and /BAX/.

it with a hypothetical example. Consider rules (4a), (4b), the underlying representations (4c) and (4d), and the results of simultaneous application (4e) and of ordered rules (4f), (4g):

(4) a. t → tʃ / __ i
b. e → i / __ tʃ

	Underlying	*Surface* e. Simultaneous application	f. Rules ordered (a)–(b)	g. Rules ordered (b)–(a)
c.	/eti/	etʃi	itʃi	etʃi
d.	/tetʃ/	titʃ	titʃ	tʃitʃ

The problem is now empirical, i.e. the question to ask is whether natural languages have input–output relations like (4c), (4d) to (4e), or rather like (4c), (4d) to (4f) or (4c), (4d) to (4g). In the case of ordered rules, the first rule creates a representation that allows the application of the second rule (feeding). With the ordering (4a) < (4b) ("<" denotes "is ordered before") feeding takes place in /eti/ → etʃi → [itʃi]; with the ordering (4b) < (4a) feeding takes place in /tetʃ/ → titʃ → [tʃitʃ]. Simultaneous application makes these feeding relationships impossible. Thus, since feeding relations are clearly observable in natural languages, the rule-ordering hypothesis receives more support than the hypothesis of simultaneous application.

A similar example can be constructed with two rules, each of which prevents the application of the other if applied first (mutual bleeding; §2). Consider a language with palatalization of velars before /i/ (CHAPTER 71: PALATALIZATION) and backing of /i/ to [u] after velars, and the underlying representation /ki/:

(5) a. k → tʃ / __ i
b. e → i / __ tʃ

	Underlying	*Surface* d. Simultaneous application	e. Rules ordered (a)–(b)	f. Rules ordered (b)–(a)
c.	/ki/	tʃu	tʃi	ku

Under rule ordering, for order (5a) < (5b) we can only apply palatalization (/ki/ → tʃi → (*n/a*)). For order (5b) < (5a), we can only apply backing (/ki/ → ku → (*n/a*)). Here simultaneous application makes bleeding impossible: both rules must apply. Simultaneous application faces another problem. A set of ordered rules assigns one and only one surface representation to any underlying representation. But consider simultaneous application of two rules, one lowering mid nasalized vowels (e.g. [ẽ] → [ã]), another raising mid unstressed vowels (e.g. [e] → [i]). They will force /'ẽ/ → ['ã], /e/ → [i], under any application mode. But unstressed /ẽ/ will satisfy both rules. Under ordering, the first rule applied always wins (we have again mutual bleeding): with the ordering lowering–raising the vowel is lowered; with the reverse ordering it is raised. Under simultaneous application, since /ẽ/ meets the structural description of both rules, two simultaneous contradictory changes must apply to /ẽ/: it has to be lowered *and* raised.

2 Total strict order and other ordering relations

Rule ordering has been a formal property of generative phonology since its beginnings (Chomsky 1951; Chomsky *et al.* 1956; Halle 1959). As defined in *SPE*, the set of rules form a total (or linear) strict order, i.e. the relation "precede" or "is ordered before" relating two rules has the properties (6a)–(6d) (see CHAPTER 34: PRECEDENCE RELATIONS IN PHONOLOGY for formal discussion of a different interpretation of precedence):

(6) Precedence (A < B, or "rule A precedes rule B") is:

- a. asymmetric: given any two rules A, B, it is not the case that A precedes B and B precedes A;
- b. (hence) irreflexive: given any rule A, it is not the case that A precedes A,
- c. transitive: given any three rules A, B, C, if A precedes B and B precedes C, then A precedes C;
- d. total (or connected): for any pair of rules A, B, either A precedes B or B precedes A.

Notice that in many cases some rules are not *crucially* ordered: the same surface form will result with the ordering A < B and B < A. Still, we assume that in the generation of the derivation that begins with the underlying form and ends with the surface form they apply in a given order. In *SPE* there were small departures from total strict order, most notably in the case of disjunctive ordering and simultaneous application in the case of infinite rule schemata, both discussed below.

Rule ordering generated a lot of discussion in the 1960s and the beginning of the 1970s. Some of the questions that were asked were whether ordering of processes was justified, whether it could be partially or totally derived from general principles, and what were the possible types of ordering. I examine now some of the proposals that were advanced in this period.

2.1 *Eliminating extrinsic ordering*

Since in many cases rules are not crucially ordered, a first question that was asked is whether extrinsic ordering could be dispensed with. The ordering between two rules A and B is *extrinsic* if it is imposed language-specifically, as in the Canadian Raising examples. On the other hand, *intrinsic* ordering refers to an ordering imposed by universal properties of grammars and by formal properties of rules. Consider the interaction of /ɣ/-deletion (7a) and Diphthongization (7b) in Finnish, analyzed by Kiparsky (1968: 177) as the interaction of two extrinsically ordered rules. But Koutsoudas *et al.* (1974) argue that if we assume that rules apply sequentially, but are unordered extrinsically in the sense that they are "simply applied to every representation that satisfies its structural description," we get the right results, no matter which rule we try to apply first ((7a) or (7b)). Consider the derivation of /vee/ and /teɣe/ in (7). /ɣ/-deletion will not be able to apply in the derivation of /vee/ either before or after Diphthongization, and we will get /vee/ → [vie]. In the second example, Diphthongization cannot apply to /teɣe/, but /ɣ/-deletion can, with the result /teɣe/ → tee. Now Diphthongization can apply, deriving the final output, tee → [tie]. Here the ordering /ɣ/-deletion < Diphthongization is not established extrinsically, but it is determined intrinsically by the form of the rules.

(7)					/vee/	/teɣe/
a.	ɣ	→	Ø / V __ V		—	tee
	ee	→	ie		vie	tie
b.	ee	→	ie		vie	—
	ɣ	→	Ø / V __ V		—	tee
	ee	→	ie		—	tie

Using this and other examples, Koutsoudas *et al.* (1974) conclude that rules are unrestricted in their application, every rule applying to every representation that satisfies its structural description, and that no extrinsic ordering needs to be imposed on them. They base their argumentation on four possible situations that derive from the possible orderings in terms of feeding, counterfeeding and counterbleeding relations. I will analyze such relations in more detail in §3; here I only give the basic definitions:

(8) Given rules A, B, where A < B:

a. A feeds B (or A and B are in feeding order) iff A can increase the number of representations to which B can apply;
b. A bleeds B (or A and B are in bleeding order) iff A can decrease the number of representations to which B can apply;
c. B counterfeeds A (or A and B are in counterfeeding order) iff B would feed A if the order were B < A;
d. B counterbleeds A (or A and B are in counterbleeding order) iff B would bleed A if the order were B < A.

Cases of pure feeding order (when it does not involve counterfeeding or counterbleeding at the same time), can be dealt with under these assumptions, as we have seen in (7). The second possible situation is a pure counterbleeding relation. Consider Koutsoudas *et al.*'s example, taken from Kiparsky (1968: 199). In certain Low German dialects, post-vocalic voiced stops spirantize. A word like *Tag* 'day' spirantizes the underlying /g/ in the plural [ta:ɣə], but in the singular it spirantizes and devoices by Final Devoicing, yielding [ta:x]. Kiparsky proposes a pair of ordered rules (9a), (9b) that determine the derivation in (9c). But Koutsoudas *et al.* argue that if the rules apply simultaneously the same result is achieved, as shown in (9d): since /g/ is a post-vocalic voiced stop and also a word-final obstruent, it meets the structural description of both rules.

(9) a. Spirantization $\begin{bmatrix}\text{+stop}\\ \text{+voice}\end{bmatrix} \rightarrow [\text{–stop}]$ / V __

b. Devoicing [+obstruent] → [–voice] / __ #

c.		/ta:g/	d.		/ta:g/
	Spirantization	ta:ɣ		Spirantization, Devoicing	ta:x
	Devoicing	ta:x		(applied simultaneously)	

A third set of cases involves two rules that are in both bleeding and counter-bleeding order. One such case is the Latin American Spanish [l ~ j] alternation in pairs like [a'kel] ~ [a'kejos] 'that (MASC SG ~ MASC PL)', analyzed by Saporta (1965) with two extrinsically ordered rules. Given underlying forms /akeʎ/, /akeʎos/ the first rule depalatalizes underlying /ʎ/ word-finally; the second turns any remaining /ʎ/ into [j].

(10)			/akeʎ/	/akeʎos/
	Final Depalatalization	ʎ → l / __#	akel	—
	Delateralization	ʎ → j	—	akejos

Cases like these are dealt with by Koutsoudas *et al.* (1974) by invoking a universal principle, very similar to Kiparsky's Elsewhere Condition (see §2.3), the Proper Inclusion Precedence Principle, which determines intrinsic ordering for a certain class of rules:

(11) *Proper Inclusion Precedence*

For any representation R, which meets the structural descriptions of each of two rules A and B, A takes applicational precedence over B with respect to R if and only if the structural description of A properly includes the structural description of B.

Given Proper Inclusion Precedence, the rules in (10) can be left (extrinsically) unordered; since the structural description of Final Depalatalization (ʎ# = "any word-final ʎ") is included in the structural description of Delateralization (ʎ = "any ʎ"), Final Depalatalization takes precedence. Notice that this is similar to disjunctive ordering in *SPE*, where a rule containing abbreviatory parentheses like A → B / __ C(D) stands for two rules, A → B / __ CD and A → B / __ C, the first rule applying first and application of one rule excluding application of the other rule. Disjunctive ordering, though, is limited to rules that are adjacent in the ordering.

The fourth set of cases is formed by pairs of rules which are in some other ordering relation. For such cases Koutsoudas *et al.* argue that there are alternative analyses in which no extrinsic ordering is necessary.

The proposal that extrinsic ordering should be dispensed with was also made by Natural Generative Phonology, a theory that began with work by Theo Vennemann and was presented in detail in Hooper (1976). In Natural Generative Phonology, many ordering relations just disappeared, because typical *SPE* phonological rules became either morphophonemic rules or "via rules," rules relating surface representations and hence not taking any part in derivations. In addition, the "No-Ordering Condition" explicitly prohibited extrinsic ordering.

2.2 *Multiple application, local ordering, and global rules*

Although based on total strict ordering (6), *SPE* allowed for multiple application of a rule in the case of infinite schemata (Chomsky and Halle 1968: 343–348), an abbreviatory convention by which X_n stands for an infinite set of sequences of *n* or more instances of X. Thus given a rule like C → Ø / __ C_1# (where C_1

abbreviates the infinite set {C, CC, CCC, . . .}) and a string . . . VCCC, both the first and the second C meet the structural description of the rule, which applies simultaneously to both and derives . . . VC#.

Another case of multiple application of a rule that was introduced after *SPE* is iterative application. In iterative rule application, which is usually directional, after the first structural change takes place, the string is scanned again and if the structural description is met, the rule reapplies, the string is scanned again, etc., until the rule is no longer applicable. Processes like harmony (CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS; CHAPTER 118: TURKISH VOWEL HARMONY; CHAPTER 123: HUNGARIAN VOWEL HARMONY; CHAPTER 110: METAPHONY IN ROMANCE), stress (CHAPTER 40: THE FOOT; CHAPTER 41: THE REPRESENTATION OF WORD STRESS; CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE), tone spreading (CHAPTER 45: THE REPRESENTATION OF TONE; CHAPTER 114: BANTU TONE), and metrical structure construction are cases of iterative application. Notice that iterative rules do not follow strict ordering because a rule can precede itself, hence the property of irreflexivity in (6b) is abandoned. Kenstowicz and Kisseberth (1977: 180–183) present the following case of iterative application. In Gidabal, a long vowel shortens after another long vowel. In a word like /djalum-baː-daːŋ-beː/ 'is certainly right on the fish', simultaneous application would derive the wrong form, *[djalum-baː-daŋ-be], with two vowels shortened, but iterative left-to-right application yields the correct derivation. We see /djalum-baː-daːŋ-beː/ → djalum-baː-daŋ-beː in the first iteration, and no change, because the structural description is not met, in the second iteration. Different versions of iterative application are discussed in Anderson (1974: 124–133) and Kenstowicz and Kisseberth (1977: 177–195).

Another case in which precedence can be reflexive, against strict order (6), is the case of "anywhere" or "persistent" rules.[4] These are rules that apply in the derivation whenever their structural description is met. A typical case is (re)syllabification. If a rule A establishes some syllabic properties to which rule B is sensitive, and rule B introduces changes, like insertion, deletion, or changes in sonority properties, to which the syllabification rule A is sensitive, A has to apply before and after B. Myers (1991) argues for several other cases of iterative ordering.

The asymmetric property of strict order (6a) is also abandoned in the case of *local ordering*, a specific type of ordering proposed in Anderson (1969) and also discussed in Anderson (1974). He proposes that two rules might have to apply in different orderings to different representations. The ordering in which they apply is the *unmarked order*. Relative to a given representation, feeding is unmarked with respect to a neutral ordering, and bleeding is marked with respect to a neutral ordering, hence the unmarked orderings are feeding and counterbleeding. This proposal stems from the observation in Kiparsky (1968) that when a linguistic change consists of a reversal in the ordering of two rules, feeding order tends to be maximized and bleeding order tend to be minimized. To see how this works, consider Icelandic Umlaut and Syncope (Anderson 1969, 20–27, 1974: 141–147) applying to [kœtlum] 'kettle-DAT PL' and [bœggli] 'parcel-DAT SG':

[4] The term "persistent rule" is due to Chafe (1968).

(12) *Local ordering: (a) applies before (b) in (c); (b) applies before (a) in (d)*

a. Umlaut a → œ / __C_0u

b. Syncope $\begin{bmatrix}+\text{syll}\\ -\text{stress}\end{bmatrix} \rightarrow \varnothing$ / C __ C+V

	c.		d.	
	/katil+um/			/bagg+ul+e/
Syncope	Ø		Umlaut	œ
Umlaut	œ		Syncope	Ø
output	kœtlum			bœggli

Notice that in (12c) for the representation /katil+um/, Syncope feeds Umlaut, and if the order were the opposite, Umlaut would either feed or bleed Syncope, i.e. order would be neutral. Hence the unmarked order Syncope < Umlaut is chosen. In (12d), for the representation /bagg+ul+e/, the first rule, Umlaut, doesn't either feed or bleed the second rule, Syncope, i.e. the order is neutral. But if the order were the opposite, Syncope would bleed Umlaut. The neutral ordering Umlaut < Syncope is therefore preferred.

There is yet another type of rule that relates to rule application and to ordering, the global rule. A global rule must be applied at the level at which it appears in the ordering, but crucial information for its structural description is ordered elsewhere in the derivation. Kisseberth and Abasheikh (1975) present the following case. In Mwini the perfect is formed by adding the suffix /-iːɫ ~ -eːɫ/ to the verb base, as shown in (13). After a base-final sibilant or /ɲ/, the /ɫ/ in the suffix spirantizes to [z] (13b). The final consonant of the verb base undergoes Mutation, which turns coronal and labial stops into [s], /k/ into [ʃ], and a nasal voiced stop sequence into [nz] (13c) (the data are from Hyman's 1993 analysis of the processes).

(13)	*Root*	*Perfect*	
a.	/jiːb/	jib-iːɫ-e	'he answered'
	/soːm/	som-eːl-e	'he read'
b.	/aːnz/	anz-iːz-e	'he began'
	/t̪oʃ/	t̪oʃ-eːz-e	'he thought'
c.	/fuːɲg/	-fuːnz-il-e	'he closed'
	/pik/	-piʃ-il-e	'he swore'

As shown by the examples in (13c), the vowel in the suffix undergoes a further change: it shortens. But shortening cannot be determined by the base-final sibilant because the derived sibilants in (13b) do not cause shortening. Hence the correct generalization seems to be that shortening applies just in case Mutation has applied, and therefore we need to make reference to another stage in the derivation instead of being restricted to the local stage of application.[5]

2.3 *The Elsewhere Condition*

The *SPE* formalism distinguishes between conjunctive ordering of two rules (the normal mode of application) and disjunctive ordering, where only one of the rules

5 Hyman (1993) discusses the Mwini case and the alternatives to a global analysis.

can apply. In *SPE*, disjunctive ordering derives from abbreviatory devices: . . . A(B) . . . corresponds to two disjunctively ordered rules, . . . AB . . . and . . . B . . . , the first immediately preceding the second, and the application of one rule excluding the application of the other rule. Disjunctive ordering was later derived as an effect of a more general principle, the Elsewhere Condition, originally formulated by Kiparsky (1973a: 94). (14) gives the revised formulation in Kiparsky (1982a: 136):

(14) Rules A, B apply disjunctively to a form Φ if and only if:

a. The structural description of A (the special rule) properly includes the structural description of B (the general rule).

b. The result of applying A to Φ is distinct from the result of applying B to Φ.

In that case, A is applied first, and if it takes effect, then B is not applied.

Consider, as an illustration, the interaction of lengthening and shortening in English (Myers 1987; Halle 1995). Lengthening and shortening affect stressed non-high vowels in branching feet. They are lengthened if, in addition, they are followed by an open syllable ending in an /i/, which in turn must be followed by a vowel in hiatus (15c), (15d); lengthened vowel underlined). Otherwise, they are shortened (15e) (see also CHAPTER 88: DERIVED ENVIRONMENT EFFECTS):

(15)	a.	$\begin{bmatrix} \text{V} \\ -\text{high} \end{bmatrix} \rightarrow \text{V:} \ / \ [\ \_\_ \ \text{C}_0 \ \text{i}]_{\text{StressFt}} \ \text{V}$	c.	*re[medi]al* *co[loni]al*	*[radi]al* *De[voni]an*
	b.	$\begin{bmatrix} \text{V} \\ -\text{high} \end{bmatrix} \rightarrow \text{V:} \ / \ [\ \_\_ \ \text{C}_0 \ \text{V}]_{\text{StressFt}}$	d.	*[trivi]al* *[natu]ral*	*di[vini]ty* *ex[plana]tory*

(15a), (15b) are two rules applying in the general context "disyllabic stress foot with stressed non-high vowel." (15a) being more restrictive, it applies only when, in addition, the second vowel is *i* and it is in hiatus with a following vowel. Since $[[\text{V}, -\text{high}]\ \text{C}_0\ \text{i}]_{\text{StressFt}}\ \text{V}$ is a subset of $[[\text{V}, -\text{high}]\ \text{C}_0\ \text{V}]_{\text{StressFt}}$, (15a) and (15b) are in an Elsewhere relationship: (15a) is the specific rule applying in $[[\text{V}, -\text{high}]\ \text{C}_0\ \text{i}]_{\text{StressFt}}$ V and (15b), the general rule, applies elsewhere in the general common context.

Kiparsky (1982a) makes ingenious use of the Elsewhere Condition to predict blocking of general morphological processes by more specific ones, and strict cyclic effects like non-derived environment blocking (NDEB). An example of the first case is the plural rule in (16c), which applies to the singular form *person*, but shouldn't apply to the lexical plural *people*. Kiparsky assumes that every lexical item L is an identity rule L $\leftrightarrow$ L. Since the structural description $[\textit{people}]_{\text{Noun, +Pl}}$ in (16a) is a proper subset of the structural description $]_{\text{Noun, +Pl}}$ in (16c), the identity rule applies (vacuously) and blocks the application of the plural rule. The structural description $[\textit{person}]_{\text{Noun, +Pl}}$, on the other hand, is not a lexical item, but the pluralization of the lexical item, so that the plural rule applies regularly to derive *persons*.

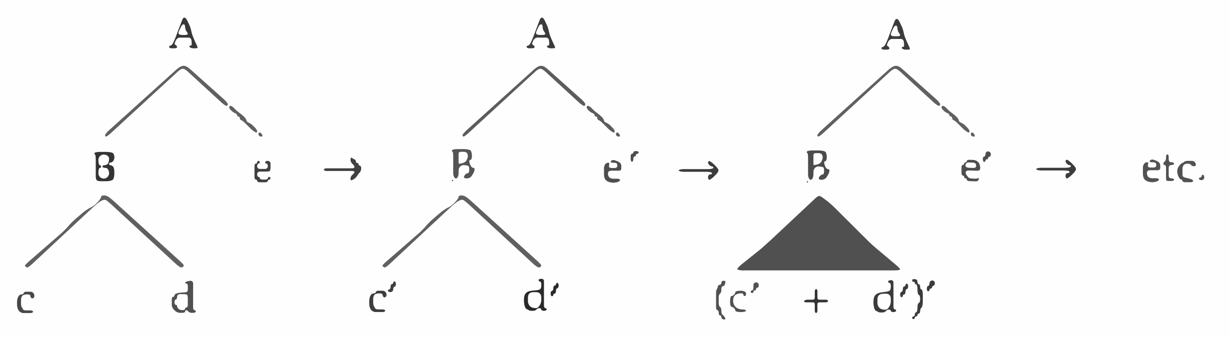
(16)	*Rules*	*Derivations*	
		$[\textit{people}]_{\text{Noun, +Pl}}$	$[\textit{person}]_{\text{Noun, +Pl}}$
a.	$[\textit{people}]_{\text{Noun, +Pl}} \leftrightarrow [\textit{people}]_{\text{Noun, +Pl}}$	$[\textit{people}]_{\text{Noun, +Pl}}$	
b.	$[\textit{person}] \leftrightarrow [\textit{person}]$		—
c.	$\emptyset \rightarrow \text{z} \ / \ \_\_ \ ]_{\text{Noun, +Pl}}$	—	$[\textit{person}]_{\text{Noun, +Pl}}$ /-z/

The Elsewhere Condition can also prevent rule application in non-derived environments (non-derived environment blocking, or NDEB; CHAPTER 88: DERIVED ENVIRONMENT EFFECTS), an effect which is also predicted by the Strict Cycle Condition (see next section). Thus the shortening rule (cf. (15b)) shortens [[*san*]*ity*] to [[*săn*]*ity*], because it is derived, but fails to apply to [*nightingale*] (*[*nĭghtingale*]), because it is underived. Again the identity rule [*san*] ↔ [*san*] cannot block shortening, because the structural description of the identity rule, [*san*], and the structural description of shortening, $[V\ C_0\ V]_{SUF}$, are not in a proper inclusion relation, but in intersection. But [*nightingale*] properly includes $[V\ C_0\ V]_{StressTr}$. Cyclic effects in later cycles/levels derive from the assumption that the output of every cycle/level is a lexical item, hence an identity rule.[6]

3 Cyclic ordering

In cyclic application the input to the set of rules is a phonological representation organized by constituent structure (CHAPTER 85: CYCLICITY). Consider a structure $[[c\ d]_B\ e]_A$, which is represented in (17) by a tree diagram; A, B are categories, and d, e, c are phonological strings, matrices, or structures.

(17)



(17) shows a cyclic derivation of $[[c\ d]_B\ e]_A$. The set of ordered rules $\mathfrak{R}$ applies first separately to the innermost constituents, i.e. to c, to d and to e, giving as a result c′, d′, e′. This is the first cycle. We now proceed to the second cycle, the next degree of embedding, namely B, and apply the set of rules to its domain, the concatenation c′ + d′, whose output is (c′ + d′)′. The domain of the following (and in this case the final) cycle is whatever is dominated by A, namely (c′ + d′)′ + e.

The cycle was first proposed in phonology (Chomsky *et al.* 1956) to deal with stress in compounds like *blackboard eraser*, showing primary–tertiary–secondary stress distinctions (stress levels are indicated above the vowel in (18), and as subscripts in the text below):

(18) 1 3 2
$[[[black]\ [board]]_B\ [eraser]]_A$

[6] Given Optimality Theory, some properties of Elsewhere-type interactions follow as a theorem, Pāṇini's Theorem on Contraint-ranking (PTC) (Prince and Smolensky 1993: §5.3, §7.2.1). PTC relates the activity of two constraints, S and G, in a constraint hierarchy relative to an input *i*. Assume that S applies non-vacuously to *i* (i.e. it distinguishes the set of candidates Gen (*i*)). If G >> S and G is active on *i* (i.e. it distinguishes the set of candidates Gen(*i*) when it applies) then S is not active on *i*. For a clear and interesting discussion of the relation between the Elsewhere Condition and PTC, see Baković (2006); also Prince (1997).

After stress has applied to individual words, the compound stress rule locates stressed vowels and maintains primary stress on the leftmost stressed vowel and weakens other stresses by one degree. After assigning vacuously primary stress in the first cycle to [*bla$_1$ck*], to [*boa$_1$rd*] and to [*era$_1$ser*], it applies in the second cycle in the domain B to [*black board*], yielding [*bla$_1$ck boa$_2$rd*], and in the last cycle in the domain A to [*bla$_1$ck boa$_2$rd era$_1$ser*] to give the final [*bla$_1$ck boa$_3$rd era$_2$ser*].

Cyclicity was later applied in syntax as a result of the elimination of generalized transformations and the generation of embedded sentences by base rules (Chomsky 1965). Later on, Chomsky (1973) proposed a limitation on cyclic application in syntax, the Strict Cycle Condition (SCC, or "strict cyclicity"), by which no rule can apply to a constituent I in such a way as to affect solely a subconstituent of I. Kean (1974) presented two cases that argued for the application of this version of the SCC also in phonology. In (17), for instance, in the second cycle, cycle B, a rule cannot apply to the domain of B if it affects just c′. An actual example is the interaction of Glide Formation and Destressing in Catalan (Mascaró 1976: 24–36). Glide Formation applies to post-vocalic unstressed high vowels. In *produirà* 'it will produce' it cannot apply to [[prudu'i]$_1$'ra]$_2$ at cycle 1, because post-vocalic /i/ is stressed. At cycle 2, a following stress causes destressing of /i/. Therefore at cycle 3, and at later cycles, the sequence /ui/ meets the structural description of the rule; but /ui/ is entirely within cycle 1 and the SCC blocks application, resulting in [pruðui'ra] *[pruðuj'ra].

The SCC was further refined in Mascaró (1976: 1–40), as in (19). Case (19b.i) corresponds roughly to the SCC as formulated in Chomsky (1973) and used by Kean (1974).

(19) Given a bracketed expression [. . . [. . . , [. . .]$_1$, . . .]$_{n-1}$ . . .]$_n$, and a (partially ordered) set of cyclic rules C:

a. C applies to the domain [. . .]$_j$ after having applied to the domain [. . .]$_{j-1}$, each rule in C applying in the given order whenever it applies properly in j.

b. *Proper application of rules*. For a cyclic rule R to apply properly in any given cycle j, it must make specific use of information proper to (i.e. introduced by virtue of) cycle j. This situation obtains if (i), (ii), or (iii) is met:

i. R makes specific use of information uniquely in cycle j. That is, it refers specifically to some A in [XAY [. . .]$_{j-1}$ Z]$_j$ or [Z [. . .]$_{j-1}$ XAY]$_j$.

ii. R makes specific use of information within different constituents of the previous cycle which cannot be referred to simultaneously until cycle j. R refers thus to some A, B in [X [. . . A . . .]$_{j-1}$ Y [. . . B . . .]$_{j-1}$ Z]$_j$.

iii. R makes use of information assigned on cycle j by a rule applying before R.

A states the general procedure for cyclic application; B gives the conditions for proper application: morphologically derived environments in inflection (19b.i), derived environments by compounding or syntax (19b.ii), and rule-derived environments (19b.iii). Effects of derived environments on application of processes, irrespective of the theoretical mechanism they derive from, are usually referred to as *derived environment effects* (DEE). We just saw a case, the interaction of glide formation and destressing in Catalan, which falls under (19b.i). The rule of /t/ → [s] assibilation in Finnish illustrates both (19b.i) and (19b.ii). Assibilation (20a) applies in morphologically derived environments like (20c): the structural description [ti] is met by material in the root cycle and in the inflected word cycle. It also applies in rule-derived environments (20d): here the structural description [ti] is met because at its cycle of application, the rule of raising (20b) has created it. But it fails to apply in the non-derived environments (20e), because none of the conditions for proper application in (19b) is met:

(20)	a.	t → s / __ i				
	b.	e → i / __##				
	c.	/halut-i/ → halus-i		'wanted'	cf. halut-a	'to want'
	d.	/vete/ → veti → vesi		'water (NOM SG)'	cf. vete-næ	'water (ESS SG)'
	e.	tila	'place, room'			
		æiti	'mother'			
		itikka	'mosquito'			

An instance of (19b.ii) is the application of glide formation in Central Catalan to vowels of different words. As we have just seen, glide formation applying to post-vocalic high vowels is blocked in *produirà* [[pɾuðu'i]$_1$'ɾa]$_2$. Consider now *produirà oxidació* 'it will produce oxidation':

(21)		[[[pɾuðu'i]$_1$ 'ɾa]$_2$	[[['ɔksið]$_1$ 'a]$_2$ 'sjo]$_3$]$_4$
	Cycle 2	[[pɾuðu'i]$_1$'ɾa]$_2$	[['ɔksið]$_1$ 'a]$_2$
		[[pɾuðui]$_1$ 'ɾa]$_2$	[[uksið]$_1$ 'a]$_2$
	Cycle 3		[[[uksið]$_1$ ə]$_2$ 'sjo]$_3$
			[[[uksið]$_1$ ə]$_2$ 'sjo]$_3$
	Cycle 4	[[[pɾuðui]$_1$ 'ɾa]$_2$	[[[uksið]$_1$ ə]$_2$ 'sjo]$_3$]$_4$
		[[[pɾuðui]$_1$ 'ɾa]$_2$	[[[wksið]$_1$ ə]$_2$ 'sjo]$_3$]$_4$

In the second word, at cycle 2, the initial /'ɔ/ is destressed by a following stress and becomes [u] by a rule of vowel reduction. At cycle 4, the sequence /au/ meets the structural description of glide formation and the SCC does not block Glide Formation, the application being proper by (19b.ii), because /au/ is not within the domain of a single previous cycle: /a/ is in cycle 2; /u/ is in cycle 3. Hence the rule applies, yielding ['aw].

It was assumed that the SCC applied to cyclic, obligatory neutralization rules, and dealt with DEEs. These were previously accounted for by the Alternation Condition proposed by Kiparsky (1973b: 65), according to which "neutralization processes apply only to derived forms . . . [i.e.] if the input involves crucially a sequence that arises in morpheme combinations or through the application of phonological processes." Cyclic application and derived environment effects were reformulated within Lexical Phonology through lexical strata and post-lexical phonology, which correspond to cycles and to the effect of the Elsewhere Condition from which DEEs are derived. In Stratal Optimality Theory (see CHAPTER 85: CYCLICITY), cycles correspond to strata to which Gen and Eval apply successively. Within OT, output-output faithfulness constraints (Benua 1997) ensure similarity of larger constituents to its inner components. Strict cycle effects (DEEs) are also obtained by local conjunction of markedness and faithfulness constraints (Łubowicz 2002). To see how DEEs are derived from local conjunction, consider the interaction of Velar Palatalization and Spirantization in Polish (Łubowicz 2002: §3). We find the following descriptive generalizations. Spirantization applies to rule-derived [ʤ] (22a), but not to underlying /ʤ/ (22b); similarly, in (22c) Velar Palatalization applies only to morphologically derived velar + [e i] sequences.

(22)	a. /rog-ek/ 'horn'	b. /banʤ-o/ 'banjo'	c. /xemik-ek/ 'chemist-DIM'
Velar Palatalization	roʤ-ek	—	xemiʧ-ek *blocked in* xe
Spirantization of /ʤ/	roʒ-ek	*blocked*	—
output	roʒ-ek	banʤ-o *banʒ-o	xemiʧ-ek *ʃemiʧ-ek

Let us examine rule-derived environments first. Given the ranking *ʤ >> IDENT[cont], we will normally have the mapping /ʤ/ → [ʒ]. The difference from derived and non-derived environments stems from the fact that in the first case the mapping is /g/ → ʤ → [ʒ], whereas in the second case it would be /ʤ/ → [ʒ]. The candidate with [ʤ] deriving from /g/ will violate both *ʤ and IDENT[cor], hence also the constraint conjunction *ʤ & IDENT(cor). But if /ʤ/ is underlying, *ʤ will be violated, but not IDENT[cor], therefore the conjunction *ʤ & IDENT[cor] will be satisfied.

For morphologically derived environments, as in the example in (22c), Łubowicz uses conjunction of markedness and ANCHOR. Velar palatalization applies to the morphologically derived sequence [k–e], but not to the non-derived sequence /xe/ in /xemik-ek/. Since the velar /k/ is stem-final, but not syllable-final, in $[xe.mi.k]_{Stem}$-ek, the sequence $k]_{Stem}$-e will violate R-ANCHOR(Stem, s), and it will also violate PAL, the constraint against velar + {e i} sequences that forces palatalization. It will therefore also violate $[$PAL & R-ANCHOR(Stem, s)$]_D$. But since morphologically underived /xe/ satisfies R-ANCHOR(Stem, s), the conjunction will be satisfied in this case and palatalization will not take place.

4 Rule interaction, ordering, and applicability: Feeding and bleeding

In a system in which rules are ordered, rules can interact: both the applicability and the result of application of a rule can depend on the application of previous rules. The notions of *feeding* and *bleeding* that I made reference to in §2 were introduced by Kiparsky (1968) in order to explain the direction of linguistic change. These concepts have been widely used since. In this section I examine them in some detail.

Since it is not uncommon to detect terminological inadequacies in the literature, in order to avoid confusion I will start with some terminological observations. In Kiparsky's original terminology, feeding and bleeding relations between rules are distinguished from feeding order and bleeding order. Feeding and bleeding relations (or the terms "X feeds/bleeds Y") are defined as functional relations between two rules, with no actual ordering between them presupposed. A feeds B if A "creates representations to which B is applicable"; A bleeds B if A "removes representations to which B would otherwise be applicable," where "representations" means *possible* representations (Kiparsky 1968: 37, 39). The terms *feeding order* and *bleeding order* are relations between rules that are in a specific order. Since feeding and bleeding relations are functional relations between rules, whether two rules are in a *feeding* or *bleeding relation* can be determined by mere inspection of the rules.[7] I will keep this distinction (feeding/bleeding relation *vs.* feeding/bleeding order), but I will reserve the use of the predicates *feed* and *bleed* applied to arguments A and B for feeding/bleeding *order*, and I will make use of the predicates *p-feed* and *p-bleed* ("p" for "potentially") in the case of feeding/bleeding *relations*. (23) provides an illustration using our previous German example (3):

(23) *German, group II* (*g*-deletion < Devoicing)

a.

Feeding/bleeding relation	A p-feeds/ p-bleeds B	devoicing p-bleeds *g*-deletion *g*-deletion p-bleeds devoicing
Feeding/bleeding order	A feeds/ bleeds B	*g*-deletion bleeds devoicing

b.

Devoicing	$[\text{obstr}] \rightarrow [-\text{voice}] \ / \ \_\_ \begin{Bmatrix} +\text{C} \\ \# \end{Bmatrix}$
g-deletion	$g \rightarrow \emptyset \ / \ [+\text{nasal}] \ \_\_$

c. *Dialect group II*

	/laŋg/
g-deletion	laŋ
devoicing	—

Devoicing p-bleeds *g*-deletion by devoicing *g* in [+nasal] + *g*, and *g*-deletion p-bleeds devoicing by deleting *g* in the same context. Given the ordering *g*-deletion < devoicing, *g*-deletion bleeds devoicing, as shown in the derivation in (23).

Feeding and bleeding relations can be formally defined as follows:

(24) *Feeding and bleeding relations*

a. Rule A is in feeding relation with respect to B (or A p-feeds B) iff there is a possible input I such that B cannot apply to I, A can apply to I, and B can apply to the result of applying A to I.

b. Rule A is in bleeding relation with respect to B (or A p-bleeds B) iff there is a possible input I such that B can apply to I, A can apply to I, and B cannot apply to the result of applying A to I.

[7] Of course one might want to to relativize these notions to a given set of representations, e.g. the lexicon. For instance, a rule A that centralizes the place of articulation of all consonants in word-final position feeds a rule B that vocalizes /l/ to [w] in coda position, because it can create the representation . . . Vl]_{Coda}## from /. . . $\text{V}ʎ\text{]}_{\text{Coda}}$##/, to which B is applicable. But in a language with a single lateral *l*, the feeding interaction will never take place. In such cases, in order to avoid terminological ambiguities we can say that A feeds B, but A doesn't feed B for lexicon L, or that A doesn't l-feed B. Similarly, if we relativize feeding and bleeding to specific derivations, we can say that a rule A does/does not d-feed or d-bleed a rule B, meaning that the feeding or bleeding relation is/is not actually instantiated in that particular derivation.

It is important to notice that in the definitions in (24) "apply" is usually interpreted as "apply non-vacuously." In the German example in (3), in dialect group I, Devoicing bleeds *g*-deletion, (/laŋg/ → /laŋk/ → (*n/a*)). But for the word *Bank* 'bank', whose derivation is /baŋk/ → (vacuous devoicing) baŋk → (*n/a*), we don't want to say that Devoicing bleeds *g*-deletion, because the input to Devoicing didn't meet its structural description. Kiparsky's (1968) terms "creates" and "removes," cited above, already indicate that vacuous application doesn't count.

On the other hand, *feeding order* and *bleeding order* (or the terms A *feeds* B and A *bleeds* B) refer to relations between two rules A and B which presuppose both feeding/bleeding relations and the specific ordering A < B (i.e. A precedes B) in the grammar. Most definitions are formulated for cases in which A immediately precedes B, or cases in which intervening rules don't interact with A and B. In such a situation the definitions become simpler: A is in feeding/bleeding order with respect to B iff A < B and A p-feeds/bleeds B. For the general case the definitions have to be refined as follows:

(25) *Feeding order and bleeding order*

Let G be a grammar, A, B rules, and D a derivation of G.

a. A is in feeding order with respect to B (or A *feeds* B) in grammar G iff
   i. A < B
   ii. There is a derivation D by G such that B would not apply to the input to A, and B applies to the output of A and would apply to all intermediate stages up to its own input.

b. A is in bleeding order with respect to B (or A *bleeds* B) in grammar G iff
   i. A < B
   ii. There is a derivation D by G such that B would apply to the input to A, and B does not apply to the output of A and would not apply to all intermediate stages up to its own input.

When A immediately precedes B or in cases where intermediate rules don't interact we get derivations like those in (26): (26a.i) is in feeding order with respect to (26a.ii) because the second rule (26a.ii) wouldn't apply to AQ, but applies to BQ, the output of the first rule (26a.i); (26b.i) is in bleeding order with respect to (26b.ii) because the second rule (26b.ii) would apply to AQ, but doesn't apply to BQ, the output of the first rule (26b.ii).

(26) a. *Feeding order*
(No intervening interacting rules)

		AQ
i.	A → B / __ Q	BQ
ii.	Q → R / B __	BR

b. *Bleeding order*

		AQ
i.	A → B / __ Q	BQ
ii.	Q → R / A __	—

The case of feeding order for two adjacent rules can be illustrated with the interaction of /æi̯/ → [aː] and Umlaut in a group of Swiss German dialects (Kiparsky 1982b: 190). Bleeding order, also for adjacent rules, can be illustrated with our earlier example (2e)–(2h), Canadian Raising, in the word *writer* in dialect B:

(27) a. *Feeding: Swiss German (dialect group I)*

		/æi̯-li/	'egg-DIM'
i.	æi̯ → a: / __ {C, ##}	a:li	
ii.	Umlaut (fronting)	æli	(does not apply non-vacuously to /æi̯-li/)

b. *Bleeding: Canadian Raising (dialect B)*

		/raɪtə˞/	
i.	t → ɾ / V __ V	raɪɾə˞	
ii.	a → ʌ / __ [C, −voice]	—	(applies non-vacuously to /raɪtə˞/)

Consider now the cases with interacting rules intervening between (i) and (ii) that motivate the definitions in (25). (28a) exemplifies feeding cases and (28b) bleeding cases. The rules (i.) and (ii.) are the rules in feeding/bleeding relation; (iii) is the intervening rule.

(28) a. *Feeding order* (Intervening interacting rules)

		QA
i.	Q → R / __ A	RA
iii.	A → B / R __	RB
ii.	A → C / R __	—

b. *Bleeding order*

		QA
i.	Q → R / __ A	RA
iii.	R → Q / __ A	QA
ii.	A → B / Q __	QB

In the feeding example, rule (i) p-feeds rule (ii) and precedes (ii), but given conditions (25a.ii) and (25b.iii), it does not feed rule (ii), because some rule ordered between them, namely (iii), undoes the change that caused the feeding (it bleeds rule (ii)). In terms of the definitions in (25), there are representations between the two rules, in particular the input to rule (ii), to which the second rule cannot apply. Similarly, in the bleeding example, rule (i) p-bleeds rule ii. and would indeed bleed rule (ii), if it were not for (iii), which feeds rule (ii).

Of course a pair of rules can show non-feeding or non-bleeding interactions like those in (28) in some derivations, but feeding or bleeding interactions in other derivations. As already indicated in note 7, I will use the terms d-feed and d-bleed when feeding and bleeding is relativized to a specific derivation. English stress provides an actual example for bleeding. Stress is assigned twice in words like *context* ['kan,tekst] or *Ahab* ['eɪ,hæb]. But after a light syllable the second stress is removed (the "Arab rule"; Ross 1972), as in *Arab* ['ærəb], and the destressed vowel reduces to [ə]. Stress bleeds vowel reduction, but in the derivation of *Arab* destressing undoes the bleeding (/ærəb/ → 'æ,ræb → 'æræb → ['ærəb]). Here we must say that stress bleeds reduction, because there are derivations that show actual bleeding, as in ['eɪ,hæb], but if we relativize bleeding to specific derivations, some of them do not show a bleeding interaction: in these derivations stress doesn't d-bleed reduction.[8]

There is yet another interesting case of p-feeding/bleeding with no actual feeding/bleeding. If a rule A p-feeds a rule B and precedes it, and there is a

[8] Notice that the example is adequate only if we assume that the underlying /ærəb/ has no stress structure and the unstressed character of the second vowel is introduced by the stress rule.

representation to which A applies and B would not apply, it is possible, according to (25), to have no feeding order even if, contrary to what happens in the previous examples, B actually applies. The same is true, *mutatis mutandis*, of bleeding order. This happens in Duke of York derivations (Pullum 1976; McCarthy 2003), which are derivations in which a rule reverses the action of a previous rule, e.g. . . . A . . . → . . . B . . . → . . . A . . . Consider the derivations in (29), which contain a Duke of York subderivation, highlighted in bold (notice that (28b) above is also an instance of a Duke of York derivation):

(29) a. *Feeding order* b. *Bleeding order*
(Intervening interacting rules; Duke of York derivations)

		QA			**QA**
i.	Q → R / __ A	**RA**	i.	Q → R / __ A	**RA**
iii.	A → B / R __	**RB**	iii.	R → Q / __ A	**QA**
iv.	B → A / R __	**RA**	iv.	A → B / Q __	QB
ii.	A → C / R __	BC	ii.	A → C / Q __	—

In (29a) (i) p-feeds and precedes (ii), and (ii) does apply, but (i) does not feed (ii), because there is an intermediate representation to which the second rule would not apply, namely RB, created by (iii). In fact it is the other intervening rule, (iv), that now feeds (i). Similarly, in (29b) (i) p-bleeds the last rule (ii), but it does not bleed it, even if the rule does not apply, because of the intermediate representation QA created by (iii), to which the rule would apply. Here it is rule (iv) which actually bleeds (ii).

Feeding and bleeding interactions have been used in different contexts and for different purposes, so it is conceivable to have slightly different changes in the definitions. One such change is desirable in cases in which usual definitions do not yield a feeding/bleeding relation, and yet this relation is intuitively correct. Consider a case like (30), in which glide formation, vowel reduction, and destressing interact in Central Catalan. (30a)–(30c) show that glide formation affects postvocalic high unstressed vowels (30a), but not non-high vowels (30b), where *Osiris* is a lexical exception to vowel reduction, or stressed vowels (30c). It also affects high unstressed vowels that are not underlying, as in (30d).

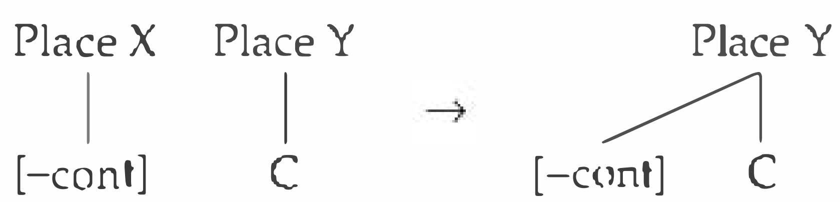
(30)

	a.	*ser*['a] [u]*mid* 'it will be wet'	b.	*ser*['a] [o]*siris* 'it will be Osiris'
Destressing				
Vowel Reduction				
Glide Formation		'a w		
output		*ser*['a] [w]*mid*		*ser*['a] [o]*siris*
	c.	*ser*['a] ['u]*til* 'it will be useful'	d.	*ser*['a] ['ɔ]*ciós* 'it will be idle'
Destressing				'a ɔ
Vowel Reduction				'a u
Glide Formation				'a w
output		*ser*['a] ['u]*til*		*ser*['a] [w]*ciós*

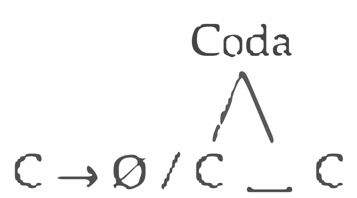
Notice now that the structural description V V[+high, −stress] is met in (30d), because vowel reduction has turned [ɔ] into a high vowel, but also because destressing has created the other condition for gliding. In such a case we want to say that these two rules jointly feed glide formation. The definitions in (25) can be changed accordingly, to meet such situations.

Notice also that a rule can stand in both feeding and bleeding order with respect to another rule. In Majorcan Catalan stops assimilate in place to a following consonant (place assimilation), and the second consonant in a two-consonant coda cluster deletes before another consonant (cluster simplification). As shown in (31), deletion of the medial C causes bleeding when the medial C is the target of assimilation, and feeding when it intervenes between the trigger and the target of assimilation.

(31) a. *Place assimilation*



b. *Cluster simplification*



c.

	Bleeding	*Feeding*
input	/'bujd 'trens/ 'I empty trains'	/'trens 'bujds/ 'empty trains'
Cluster simplification	'buj 'trens	'tren 'bujds
Place assimilation	—	'trem 'bujds
output	'buj 'trens	'trem 'bujts

It should be observed that the fact that two rules A, B do not have a feeding or bleeding interaction does not mean that they don't interact. In (32) rule (a) deletes final consonants, while (b) stresses the final syllable if it is heavy, otherwise the penult. Different orderings give different results, but the interaction isn't either a feeding or a bleeding relation.

(32) a. C → Ø / __ ## /satopek/ $\xrightarrow{a}$ satope $\xrightarrow{b}$ sa'tope
b. V → [+stress] / __ C(V)## /satopek/ $\xrightarrow{b}$ sato'pek $\xrightarrow{a}$ sato'pe

I will now examine *counterfeeding* and *counterbleeding*. These notions refer only to rules that are in a specific order (potential situations don't make sense in this context). Basically, a counterfeeding/bleeding order is an order that would be feeding/bleeding if the order of the rules were reversed. Since there is some confusion in the use of the predicates, I will follow the practice in Koutsoudas *et al.* (1974) and use as subject of "counterfeed/counterbleed" the second rule in the ordering, e.g. B counterfeeds A means that A < B, and B would feed A if B < A.

(33) a. A and B *are in counterfeeding order* (B *counterfeeds* A) in grammar G iff
i. A < B
ii. B p-feeds A

b. A and B *are in counterbleeding order* (B *counterbleeds* A) in grammar G iff
i. A < B
ii. B p-bleeds A

Counterfeeding order can be illustrated with the same processes of Swiss German presented in (27) in another dialect group, group II, which shows the opposite ordering. Counterbleeding is illustrated with Canadian Raising in dialect A (2a):

(34) a. *Counterfeeding: Swiss German (dialect group II)*

	/æi̯-li/	'egg-DIM'
i. Umlaut (fronting)	—	
ii. $\text{æi̯} \rightarrow \text{aː} \;/\; \_\_ \begin{Bmatrix} C \\ \#\# \end{Bmatrix}$	aːli	

The opposite, feeding ordering would yield /æi̯-li/ → aː-li → æː-li

b. *Counterbleeding: Canadian Raising (dialect A)*

	/raɪtə/
i. a → ʌ / __ [C, −voice]	rʌɪtə
ii. t → ɾ / V __ V	rʌɪɾə

The opposite, bleeding ordering would yield /raɪtə/ → raɪɾə → (*n/a*)

"Counter" orderings have important properties. Assume the simple case where rules A and B are adjacent, and A < B. Since in feeding order (B < A) there must be at least one input I such that B is applicable to I, A is not applicable to I, and A is applicable to the output of B (35a), it follows that in the corresponding counterfeeding order where A < B there must be an input (namely I) to which the first rule, now A, does not apply and to which the second rule, now B, applies (35b). Hence the generalization expressed by A does not appear in the output: we can say, using McCarthy's (1999) terms, that it is *not surface-true*, it is not true of the output of B, usually the surface representation. In the bleeding order B < A there must be by definition at least one input I such that both A and B are applicable to I (giving different results, I′ and I″, respectively), and A is not applicable to the output of B (35c). It follows that in the corresponding counterbleeding order A < B there can be an input to which the first rule, now A, applies and to which the second rule, now B, might apply and change the context of application of the first rule (35d). Hence the generalization expressed by A about the input I does not appear in the output: following McCarthy we can say that it is *not surface-apparent*, because the generalization A about I is not apparent in the output of A, usually the surface representation.

(35)

a. *Feeding*		b. *Counterfeeding*		c. *Bleeding*		d. *Counterbleeding*	
	I		I		I		I
B.	I′	A.	—	B.	I′	A.	I″
A.	I″	B.	I‴	A.	—	B.	— /I‴

In our previous example in (34) for counterfeeding in Swiss German, the fronting dictated by Umlaut is not true in the surface form [a:li]; for the counterbleeding in Canadian Raising, dialect A, the fact that [ʌ] derived from /a/ appears before voiceless consonants is not apparent in the surface form [rʌɪɾə].

It is also important to notice that the existential quantification in the definitions in (25) of feeding and bleeding orders (hence also of counterfeeding and counterbleeding orders) allows for the existence of multiple feeding and bleeding relations between two rules. For feeding, and given two ordered rules A < B, the requirement (25a.ii) that there be an input I whose derivation D meets the conditions required in (25a.ii) does not prevent the existence of another input I′ that meets the condition (25b.ii) for bleeding. Hence A can both feed and bleed B (and B can both counterfeed and counterbleed A).

5 Serial and parallel approaches

Rule interactions of the sort just discussed have become important in the theoretical comparative analysis of serial and parallel approaches, in particular in relation to *opaque* rule interactions. If we compare a standard serial theory like *SPE* with a parallel theory based on constraints like Optimality Theory (OT), pure feeding and pure bleeding order effects (i.e. those that are not also counterfeeding or counterbleeding) are *transparent* interactions and can be derived from both. Consider the well-known case of *e*-raising and /t/ → [s] interaction in Finnish (Kiparsky 1973b: 166–172), partially repeated from (20):

(36)		vete 'water-NOM SG'	halut-i 'wanted'
	a. e → i / __##	veti	—
	b. t → s / __ i	vesi	halus-i

Because both (36a) and (36b) are statements that are true of surface forms, constraints of the form *e##, *ti, dominating conflicting faithfulness constraints, together with other constraints determining the choice of [i] and [s], will derive the output of /vesi/, /halus-i/.

But counterfeeding and counterbleeding are opaque interactions and cause problems for a parallel approach. A process (37a) is opaque (Kiparsky 1973b: 79) to the extent that there are phonetic forms in (37b) or (37c); otherwise it is transparent. The derivations (37d) and (37e) illustrate (37b) and (37c), respectively.

(37) a. *Rule:* A → B / C __ D

Opaque surface forms

b. A in the environment C __ D

c. B derived by (a) in an environment different from C __ D

d.		/EAD/
	A → B / C __ D	—
	E → C / __ A	CAD
e.		/CAD/
	A → B / C __ D	CBD
	C → E / __ A	EBD

In (37d) the generalization "A does not appear in C __ D; B appears instead" expressed by (37a) is not surface-true; the rule *underapplies* with respect to surface representations. In (37e) the generalization "underlying (or intermediate) A is represented by B in C __ D" is not true of the derivation, it is not surface-apparent; the rule *overapplies* with respect to surface representations, since it applies outside its environment. To illustrate with a real example, consider counterfeeding in Madurese (Austronesian, Indonesia) (McCarthy 2002: 174–175). Nasality spreads rightwards onto following vowels, but is blocked by oral consonants, and voiced stops delete after a nasal (CHAPTER 78: NASAL HARMONY):

(38)	a.		/naŋgaʔ/	/naŋaʔ/
	b.	V → [+nas] / N __	nãŋgaʔ	nãŋãʔ
	c.	/b d g/ → Ø / N __	nãŋaʔ	—
		Surface	[nãŋaʔ]	[nãŋãʔ]

In the first derivation, rule (38c) has deleted an oral consonant and has thus partially changed the context of application of rule (38b); rule (38b) underapplies, because if it did apply to the surface representation it would nasalize the second vowel, *[nãŋãʔ]. The generalization that a nasal vowel nasalizes following vowels across non-oral consonants is not surface-true.

Such an opaque interaction is derivable in an ordered rule system, but not in a system in which markedness generalizations are about surface forms. Consider now a model like OT. For an input /nãŋaʔ/ (cf. the second derivation in (38)), the constraint hierarchy must favor candidate [nãŋãʔ] over candidate *[nãŋaʔ] (nasalization spreads across non-oral consonants). Therefore it will also favor the non-opaque candidate *[nãŋãʔ] over candidate [nãŋaʔ] if the input is /nãŋgaʔ/. Similar considerations apply to counterbleeding opacity. Consider our earlier example, Canadian Raising in dialect A. The change /aɪt/ → [ʌɪt] does not appear as such in the phonetic representation of *writer*, because the second rule has modified the result of the change, turning the triggering voiceless /t/ into [ɾ]:

(39)	a.		/raɪtə/		/taɪp/
	b.	aɪ → ʌɪ / __ [C, −voice]	rʌɪtə		tʌɪp
	c.	t → ɾ / V __ V	rʌɪɾə	(ʌɪɾ, *not* ʌɪt)	tʌɪp

Here in order to obtain the transparent [tʌɪp] in *type*, both *aɪ[C, −voice] and *VtV must be active. But for *writer* the input /raɪtə/, where both constraints are relevant, cannot have as output [rʌɪɾə], because the candidate [rʌɪtə] also satisfies both markedness constraints and is, in addition, more faithful to the input:

(40)

/raɪtə/	*aɪ[C, −voice]	*VtV	Faith[aɪ]	Faith[t]
a. raɪtə	*!	*		
b. rʌɪtə		*!	*	
☞ c. raɪɾə				*
d. rʌɪɾə			*!	*

6 Conclusion

Rules are generalizations about the distribution of sound in natural languages. Rule ordering is a specific theory about how these generalizations interact to derive a surface representation. The intensive study of many phonological systems using rule ordering has not only produced a rich body of descriptive work, but has also unveiled many deep properties of phonological systems and many theoretical problems that go beyond the model that generated them. When the problem of the theoretical status of phonology was first addressed seriously, it was immediately realized that phonological generalizations could not have two properties at the same time: they could not be absolute generalizations and generalizations about the surface representation. In other words, they could not map a lexical representation to a surface representation in one step (simultaneous rule application), as illustrated by Joos's paradox discussed at the beginning of §1. The response to this fact was that the requirement that generalizations be true of surface representations should be abandoned, and hence that phonological processes had to be ordered. The conviction of many present-day phonologists that the right response is to abandon the other requirement, i.e. that generalizations be absolute, and keep the idea that they apply to surface representations, has been made possible by many decades of work in a framework based on rule ordering. Even if many things have changed since the days in which a phonological description could be based on a system with a depth of ordering of 20 or 30 (i.e. 20 or 30 rules that had to be linearly ordered),[9] serial approaches haven't achieved a total elimination of ordering through mechanisms like the ones described in §2.3. At the same time, many of the properties of phonological systems that have been discovered as the result of work on rule ordering – the existence of opacity, disjunctivity as predicted by the Elsewhere Condition, derived environment effects, and many morphology–phonology interactions – are still important problems that will stimulate further research, for both serial and parallel approaches.

REFERENCES

Anderson, Stephen R. 1969. West Scandinavian vowel systems and the ordering of phonological rules. Ph.D. dissertation, MIT.

Anderson, Stephen R. 1974. *The organization of phonology*. New York: Academic Press.

Anderson, Stephen R. & Paul Kiparsky (eds.) 1973. *A Festschrift for Morris Halle*. New York: Holt, Rinehart & Winston.

Baković, Eric. 2006. Elsewhere effects in Optimality Theory. In Eric Baković, Junko Itô & John J. McCarthy (eds.) *Wondering at the natural fecundity of things: Essays in honor of Alan Prince*, 23–70. Santa Cruz: Linguistics Research Center, University of California, Santa Cruz. Available at http://repositories.cdlib.org/lrc/prince/4.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Ph.D. dissertation, University of Massachusetts, Amherst. Published 2000, New York: Garland.

9 "In the segment of the phonological component for Modern Hebrew presented in Chomsky (1951), a depth of ordering that reaches the range of twenty to thirty is demonstrated and this is surely an underestimate" (Chomsky 1964: 71).

Chafe, Wallace L. 1968. The ordering of phonological rules. *International Journal of American Linguistics* 34. 115–136.

Chambers, J. K. 1973. Canadian Raising. *Canadian Journal of Linguistics* 18. 131–135.

Chambers, J. K. (ed.) 1975. *Canadian English: Origins and structures*. Toronto: Methuen.

Chambers, J. K. 2006. Canadian Raising retrospect and prospect. *Canadian Journal of Linguistics* 51. 105–118.

Chomsky, Noam. 1951. Morphophonemics of modern Hebrew. M.A. thesis, University of Pennsylvania. Published 1979, New York: Garland.

Chomsky, Noam. 1962. A transformational approach to syntax. In Archibald A. Hill (ed.) *Proceedings of the 3rd Texas Conference on Problems of Linguistic Analysis in English*, 124–158. Austin: University of Texas Press.

Chomsky, Noam. 1964. *Current issues in linguistic theory*. The Hague & Paris: Mouton.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1973. Conditions on transformations. In Anderson & Kiparsky (1973), 232–286.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Chomsky, Noam, Morris Halle & Fred Lukoff. 1956. On accent and juncture in English. In Morris Halle, Horace Lunt, Hugh MacLean & Cornelis van Schooneveld (eds.) *For Roman Jakobson: Essays on the occasion of his sixtieth birthday*, 65–80. The Hague: Mouton.

Halle, Morris. 1959. *The sound pattern of Russian: A linguistic and acoustical investigation*. The Hague: Mouton.

Halle, Morris. 1962. Phonology in generative grammar. *Word* 18. 54–72.

Halle, Morris. 1995. Comments on Luigi Burzio's "The rise of Optimality Theory." *Glot International* 1. 9–10.

Harris, Zellig S. 1951. *Methods in structural linguistics*. Chicago: University of Chicago Press.

Hooper, Joan B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.

Hyman, Larry M. 1993. Problems for rule ordering in Bantu: Two Bantu test cases. In John A. Goldsmith (ed.) *The last phonological rule: Reflections on constraints and derivations*, 195–222. Chicago & London: University of Chicago Press.

Idsardi, William J. 2006. Canadian Raising, opacity, and rephonemicization. *Canadian Journal of Linguistics* 51. 119–126.

Joos, Martin. 1942. A phonological dilemma in Canadian English. *Language* 18. 141–144. Reprinted in Chambers (1975), 79–82.

Kaye, Jonathan. 1990. What ever happened to dialect B? In Joan Mascaró & Marina Nespor (eds.) *Grammar in progress: Glow essays for Henk van Riemsdijk*, 259–263. Dordrecht: Foris.

Kean, Marie-Louise. 1974. The strict cycle in phonology. *Linguistic Inquiry* 5. 179–203.

Kenstowicz, Michael & Charles W. Kisseberth. 1977. *Topics in phonological theory*. New York: Academic Press.

Kiparsky, Paul. 1968. Linguistic universals and linguistic change. In Emmon Bach & Robert T. Harms (eds.) *Universals in linguistic theory*, 171–202. New York: Holt, Rinehart & Winston. Reprinted in Kiparsky (1982b), 13–43.

Kiparsky, Paul. 1973a. "Elsewhere" in phonology. In Anderson & Kiparsky (1973), 93–106.

Kiparsky, Paul. 1973b. Phonological representations: Abstractness, opacity, and global rules. In Osamu Fujimura (ed.) *Three dimensions of linguistic theory*, 57–86. Tokyo: Taikusha.

Kiparsky, Paul. 1982a. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.

Kiparsky, Paul. 1982b. *Explanation in phonology*. Dordrecht: Foris.

Kisseberth, Charles W. & Mohammad Imam Abasheikh. 1975. The perfective stem in Chimwi:ni and global rules. *Studies in African Linguistics* 6. 249–266.

Koutsoudas, Andreas, Gerald Sanders & Craig Noll. 1974. The application of phonological rules. *Language* 50. 1–28.

Łubowicz, Anna. 2002. Derived environment effects in Optimality Theory. *Lingua* 112. 243–280.

Mascaró, Joan. 1976. Catalan phonology and the phonological cycle. Ph.D. dissertation, MIT. Published 1978, Indiana University Linguistics Club.

McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16. 331–399.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John J. 2003. Sympathy, cumulativity, and the Duke-of-York gambit. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*, 23–76. Cambridge: Cambridge University Press.

Mielke, Jeff, Mike Armstrong & Elizabeth Hume. 2003. Looking through opacity. *Theoretical Linguistics* 29. 123–139.

Myers, Scott. 1987. Vowel shortening in English. *Natural Language and Linguistic Theory* 5. 485–518.

Myers, Scott. 1991. Persistent rules. *Linguistic Inquiry* 22. 315–344.

Postal, Paul. 1968. *Aspects of phonological theory*. New York: Harper & Row.

Prince, Alan. 1997. Elsewhere and otherwise. *Glot International* 2. 23–24 (ROA-217).

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Pullum, Geoffrey K. 1976. The Duke of York gambit. *Journal of Linguistics* 12. 83–102.

Ross, John R. 1972. A reanalysis of English word stress. In Michael K. Brame (ed.) *Contributions to generative phonology*, 229–323. Austin: University of Texas Press.

Saporta, Sol. 1965. Ordered rules, dialect differences, and historical processes. *Language* 41. 218–224.

Whitney, William Dwight. 1889. *Sanskrit grammar, including both the classical language, and the older dialects of Veda and Brahmana*. 2nd edn. Cambridge, MA: Harvard University Press. Reprinted 1975, Cambridge, MA: Harvard University Press.

75 Consonant–Vowel Place Feature Interactions

Jaye Padgett

1 Introduction

Both consonants and vowels are formed with constrictions in the oral cavity, made by the lips, the tongue blade, the tongue body, and/or the tongue root. Since they make demands on the same organs, it should not be surprising that the place features of consonants can influence those of vowels and vice versa. Indeed, such interactions are common: consonants and vowels frequently assimilate in place to one another, or dissimilate. But the empirical territory is not simple, and attempts to understand consonant–vowel place interactions (henceforth "C–V interactions") have led to much unresolved debate in phonological theory.

The questions most debated have had to do with the nature of the phonological features we assume, with questions of feature structure and with claims about the locality of phonological processes. However, as the field of phonology gravitated toward questions of constraint interaction under the influence of Optimality Theory (Prince and Smolensky 1993), attention toward these representational questions faded without having been resolved. Whatever the theoretical framework though, the empirical puzzles underlying the debate about C–V interactions remain, and remain interesting.

The discussion in this chapter will necessarily reflect the open-endedness of the historical discussion, as well as the framework in which that discussion was held – autosegmental phonology and feature geometry. §2 begins by presenting a typology of C–V interactions. §3 puts forward an influential model of feature geometry as a point of departure and reviews the challenges raised for that model by C–V interactions. §4 discusses a prominent approach to these challenges, a "unified feature" approach to consonants and vowels advocated by Herzallah (1990), Clements (1991), Hume (1994, 1996), Clements and Hume (1995), and others. In §5 we pause to consider issues of locality and transparency in C–V interactions. §6 covers an alternative to the unified feature approach, due to Ní Chiosáin and Padgett (1993) and Flemming (1995, 2003), called the "inherent vowel place" approach here. §7 concludes.

2 A typology of C–V interactions

The typology given here is not meant as an exhaustive survey of the kinds of C–V interaction known. Instead the goal is to classify processes according to the challenges they have presented for phonological theory. In particular, a key distinction will be made between "within-category" C–V interactions and "cross-category" C–V interactions.[1] Also, for space reasons the main focus will be on assimilations (CHAPTER 81: LOCAL ASSIMILATION), with only occasional reference made to dissimilatory cases (see CHAPTER 60: DISSIMILATION).

2.1 *Within-category interactions*

It may seem incoherent to posit "within-category" interactions between the distinct categories of consonant and vowel. However, it is well known that consonants can have secondary articulations that are essentially vocalic in nature: vowel- or glide-like gestures, produced along with a consonant's primary place of articulation. Some representative examples are illustrated in (1) (see also CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION; CHAPTER 71: PALATALIZATION; CHAPTER 121: SLAVIC PALATALIZATION).[2]

(1) *(Semi-)vocalic secondary articulations*

labialization	palatalization	velarization	pharyngealization
t^w	t^j	t^{γ}	$t^{ʕ}$

Indeed, glides themselves are consonants with vocalic properties (CHAPTER 15: GLIDES). "Within-category" interactions are those between a vowel and another (semi)-vocalic element, whether the latter is a secondary articulation or a primary one (a glide).

Let us begin with interactions between vowels and glides. The examples in (2) are from Kabardian (Colarusso 1992: 32–33). Kabardian has a "vertical" vowel system arguably consisting of only the two phonemes /ə a/. These vowels assimilate in backness and roundness to a following coda glide. According to Colarusso, the triggering glide is elided in all but careful speech, with some compensatory lengthening (not shown). Effects like this of glides on vowels, affecting either vowel color (backness and/or roundness) or height (CHAPTER 19: VOWEL PLACE; CHAPTER 21: VOWEL HEIGHT), seem common in languages.

(2)	/q'əw/	[q'uw]	'swan'	/bəj/	[bij]	'enemy'
	/psaw/	[psow]	'alive'	/tsaj/	[tsej]	'one of wool (kind of coat)'

Turning to vocalic secondary articulations, non-low short vowels in Irish are front before palatalized consonants and back before non-palatalized consonants; the

[1] These terms are borrowed from Clements (1991). Compare the "Type I" *vs.* "Type II" distinction of Ní Chiosáin and Padgett (1993).

[2] This presentation simplifies reality in some ways. For example, sounds transcribed C^w or C^{γ} might be labial-velarized and not just labialized or velarized. In addition, "pharyngealized" sounds are more accurately described as "uvularized" in at least some cases (McCarthy 1994).

latter are velarized. The symbols "I/E" denote underlying high and mid vowels (respectively) of indeterminate backness.

(3)	/mˠIdʲ/	[mˠɪdʲ]	'we/us'	/pˠIntˠ/	[pˠuntˠ]	'pound'
	/sʲIvʲ/	[ʃɪvʲ]	'you (PL)'	/skʲIbˠ/	[skʲubˠ]	'snatch'
	/tˠEtʲ/	[tˠɛtʲ]	'smoke'	/bˠEsˠ/	[bˠʌsˠ]	'palm (of hand)'
	/tʲEpʲ/	[tʲɛpʲ]	'fail'	/lʲEmˠ/	[lʲʌmˠ]	'with me'

Similarly, labialized consonants can cause a neighboring vowel to be round, as in Kabardian /dəʁʷ/ → [doʁʷ] 'thief' (Colarusso 1992: 30). In a case involving pharyngealization (or uvularization; see note 2), emphatic consonants in Palestinian Arabic cause /a/ to ablaut to [u] instead of [i] in first measure imperfect verbs (Herzallah 1990): the imperfect form of [naðˤam] 'compose' is [ji-nðˤum] rather than the expected *[ji-nðˤim] (cf. [katab], [ji-ktib] 'write'). Herzallah argues that secondary pharyngealization involves a component of backness that spreads to the vowel in these cases. In a more typical case of emphasis spread, vowels in Ayt Seghrouchen Tamazight Berber are backed and lowered next to emphatic consonants (Rose 1996), as shown in (4). Rose argues that emphasis spread is the spreading of the feature [RTR] ([Retracted Tongue Root]) (see also CHAPTER 25: PHARYNGEALS; CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS).

(4)	a.	[izi]	'fly'	b.	/izˤi/	[ezˤe]	'bladder'
		[llef]	'to divorce'		/tˤtˤef/	[tˤtˤɛf]	'to hold'
		[nðu]	'to be shaken'		/nðˤu/	[nðˤo]	'to cross'

Consonants commonly acquire vocalic secondary articulations by assimilating to adjacent vowels. For example, Russian consonants are palatalized before certain suffixes beginning in [i] or [e] (Padgett, forthcoming; see also CHAPTER 121: SLAVIC PALATALIZATION):

(5)	*Nom sg*	*Nom sg (dim)*	*Loc sg*	
	stol	stolʲik	stolʲe	'table'
	dom	domʲik	domʲe	'house'
	ʂar	ʂarʲik	ʂarʲe	'ball'
	zont	zontʲik	zontʲe	'umbrella'

A similar palatalization occurs in Nupe (Hyman 1970). Also in Nupe, consonants are rounded (or labial-velarized) before rounded vowels, e.g. [egʷũ] and [egʷo] for /egũ/ 'mud' and /ego/ 'grass' (tones not shown).

It is worth noting that the examples of within-category assimilation presented above never involve a vowel changing the features of a glide or of a consonant's secondary articulation. Such cases seem at best rare, but it is not clear why that should be.

Again, what all within-category interactions have in common is interaction among overtly (semi-)vocalic elements. Glides are [−consonantal] in the feature theory of Chomsky and Halle (1968; *SPE*). Vocalic secondary articulations are likewise basically vocalic in constriction degree, even if they accompany primary constrictions that are [+consonantal]. For reasons that will become clear below, C–V interactions of this sort have created little controversy in phonological theory. This is in contrast to C–V interactions in which the *primary* articulation of a [+consonantal] segment (CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION) appears to interact with the place of a vowel, cases called "cross-category" here.

2.2 *Cross-category interactions*

Numerous cases are known in which plain (not rounded) labial consonants cause vowels to be round. This happens, for example, in a dialect of Mapila Malayalam described by Bright (1972).[3] In this dialect, a vowel is inserted for apparently phonotactic reasons. The vowel is generally something like [ɨ] (not a phoneme of the language), as shown in (6a); but it surfaces as [u] after [o] or [u] (6b) or after a labial consonant (6c). The rule is productive, applying even in borrowings like [trippu].

(6)	a.	pa:lɨ	'milk'	b.	onnu	'one'	c.	ʧa:vu	'death'
		pandɨ	'shake'		nu:ru	'hundred'		ʤappu	'pound'
		kurva:nɨ	'Koran'		uɳɳu	'dine!'		isla:mu	'Islam'
		dressɨ	'dress'		o:ɖu	'run!'		trippu	'trip'

Another well-known case occurs in Turkish (Lees 1961; Lightner 1972). Within historically native Turkish roots, a high vowel following [ɑ] and any intervening consonants is normally [ɯ] (sometimes transcribed [ɨ]). But it is [u] when a labial consonant intervenes, e.g. [jɑvru] 'cub, chick', [ɑrmud] 'pear'. Cross-category dissimilations also occur. For example, in Cantonese a syllable rhyme cannot have both a rounded vowel and a labial coda, e.g. *[up] (Cheng 1991). Other languages showing C–V interactions involving vowel rounding and plain labial consonants are discussed by Hyman (1973), Campbell (1974), Sagey (1986), Clements (1991), Selkirk (1993), Flemming (1995), and Anttila (2002).

There seems to be a similar connection between coronal place of articulation (CHAPTER 12: CORONALS) and front vowels. A frequently cited example comes from Maltese Arabic (Brame 1972; Hume 1994, 1996). In imperfective Measure I verbs, the prefix vowel is normally identical to the vowel of the stem, as shown in (7a). However, when the stem begins with a coronal obstruent, the prefix vowel is [i], (7b). Note that some of the verbs in (7b) undergo an independently existing ablaut by which the imperfective stem vowel becomes [o]; this occurs in verbs without initial coronal obstruents too, e.g. [barad] *vs.* [jo-brod] 'to file'. In these verbs the prefix vowel is normally [o].

(7)		*perfective*	*imperfective*		
	a.	kotor	jo-ktor	'to increase'	
		ʔasam	ja-ʔsam	'to break'	
		ħeles	je-ħles	'to set free'	
		nizel	ji-nzel	'to descend'	UR = /nizil/

[3] Bright relies on Upadhyaya (1968) for data. The Mapila Malayalam data resemble the more often-cited Tulu facts also discussed by Bright; in fact, Bright suggests that the Mapila Malayalam facts are due to contact with Tulu.

b.	daħal	ji-dħol	'to enter'
	talab	ji-tlob	'to pray'
	seħet	ji-sħet	'to curse'
	ʤabar	ji-ʤbor	'to collect'

Hume (1994, 1996) treats the general vowel copy as a case of feature stability: when the first vowel of the imperfective stem is deleted (by a normal syncope rule), its vowel place features surface on the underlyingly featureless prefix vowel. In (7b), however, a rule applies by which the prefix vowel acquires its frontness from a coronal obstruent; this rule takes precedence over the feature stability rule.

In another widely cited case, non-low vowels in Cantonese must be front when between coronal consonants (Cheng 1991); next to [tit] 'iron' and [tøn] 'a shield' there are no forms like *[tut] or *[ton]. Similarly, in Kabardian the vowels /a/ and /ə/ are allophonically fronted before coronal consonants, e.g. /ʃəd/ 'donkey' and /zaz/ 'bile' become [ʃɛd] and [zæz], respectively (Colarusso 1992: 30). Discussion of other cases can be found in Clements (1976, 1991), Hume (1994), and Flemming (1995, 2003).

Dorsal consonants can trigger backing of vowels. This is clearly true when the consonants in question are uvular, analyzed by many as having a [pharyngeal] (Herzallah 1990; McCarthy 1994) or [RTR] (Rose 1996) component in their place of articulation, as well as a [dorsal] one. In fact, uvulars, which are [+back, –high] in the *SPE* framework, can trigger backing and/or lowering. The data in (8) from Inuktitut are taken from Buckley (2000), who cites Schultz-Lorentzen (1945) and Fortescue (1984). The high vowels seen in (8b) are lowered to mid before either of the language's uvular segments, [q] and [ʀ], (8a). This is an allophonic change, since the vowel phonemes of Inuktitut are /i u a/. According to Rischel (1974), this rule involves retraction as well as (or even more than) lowering, though this is not obvious from the transcriptions. As Elorrieta (1991) notes, this is consistent with the notion that uvulars are pharyngealized dorsal consonants.

(8)	a.	seʀmɛ-q	'glacier'	b.	seʀmi-t	'glaciers'
		ikɛ-ʀput	'our wound'		iki-t	'your wound'
		uvdlɔ-q	'day'		uvdlu-t	'days'

In Kabardian, the phonemes /ə a/ are backed before uvulars, e.g. /baq/ → [bɑq^{h}] 'cow shed' (Colarusso 1992: 30).[4]

Velar consonants, which are also [dorsal], can also cause backing, and/or raising. One example is from Maxakalí (Gudschinsky *et al.* 1970; Clements 1991). Tautosyllabic VC sequences tend to display an excrescent vowel which either replaces the consonant or forms a transition from vowel to consonant, depending on aspects of the environment. The place of this excrescent vowel depends on the place of the consonant. As shown in (9a), that vowel is /ə/ before alveolars. But it is a high back vowel before velars, (9b). (The vowel is also [i] before "alveo-palatals," and something like [ʌ] before labials.)[5]

[4] Colarusso states that backing affects /ə/ also, and his rule predicts [ɣ], but he transcribes [ə].

[5] The relevant excrescent vowel is underlined. Along with the excrescent vowel, a preceding glide can appear. A breve indicates the vowel is non-syllabic. Gudschinsky *et al.* actually indicate a good deal of variation in these excrescent vowel qualities.

(9)	a.	/mit/	[mbijə̲t̚]	'sound of a jaguar's footsteps'
		/kot nak/	[kowə̲ daɯ̆ɨx]	'dry manioc'
	b.	/noʔok/	[ndoʔoɯ̲̆x]	'to wave (something)'
		/kɯcakkɯk/	[kɯʃaɯ̲kɯx]	'capybara (species of rodent)'

If Clements (1991) is right that [ə] represents the basic quality of the excrescent vowel, then [k] (and also [ŋ]) seem to cause it to raise and back. Similarly, in Yoruba, certain *i*-initial nouns show the /i/ backing to [u] when a velar precedes. This occurs in a reduplicative context, e.g. /ki + isɔ/ → [isɔku̲sɔ] 'saying, foolish/loose talk' (Pulleyblank 1988: 245–246; tones omitted). Other cross-category cases involving velars are discussed in Ní Chiosáin and Padgett (1993), Clements and Hume (1995), and references therein.[6]

Finally, pharyngeal consonants often cause vowels to lower and back, particularly to [a]. In fact, this can be a property of all "guttural" consonants – uvulars, pharyngeals, and (for some languages) laryngeals – which can all be analyzed as having a [pharyngeal] component to their place of articulation (McCarthy 1994) (see also CHAPTER 25: PHARYNGEALS). The examples in (10), taken from Rose (1996), who cites Cowell (1964), are from Syrian Arabic. The feminine suffix /-e/, seen in (10a), is realized as [a] after gutturals, (10b).

(10)	a.	daraʒ-e	'step'	b.	waːʒh-a	'display'
		ʃerk-e	'society'		mniːħ-a	'good'
		madras-e	'school'		daggaːʀ-a	'tanning'

The examples of cross-category assimilation discussed so far involve consonants affecting vowels. A striking fact is that consonant-to-vowel cross-category assimilations are notably missing (Ní Chiosáin and Padgett 1993). The one clear exception to this claim is the case of palatalizing mutations. As many have noted (e.g. Clements 1976; Mester and Itô 1989), front vowels, especially higher ones, often trigger mutations of velars or dentals/alveolars to palato-alveolar (or a similar) place of articulation. Hume (1996) cites a case of velar mutation in Slovak (Rubach 1993) by which /k g x ɣ/ become [ʧ ʤ ʃ ʒ] respectively before any of [j i e æ] (see CHAPTER 65: CONSONANT MUTATION for more on mutations):

(11)	a.	vnuk	'grandson'	vnuʧik	(DIM)
	b.	tsveng	'sound'	tsvenʤat[j]	'to sound'[7]
	c.	strax	'fright'	straʃit[j]	'frighten'
	d.	boɣ	'god'	boʒe	(VOC)

Such cases are common, and clearly involve assimilation of a velar consonant to a front vowel. The existence of these cases might lead us to expect equally frequent assimilations to round vowels, such as /ku/ → [pu], or assimilations to [a], such as /fa/ → [ħa]. But assimilations like these, or the many others that

6 Some researchers have suggested that cross-category assimilations of vowels to velars like these are unexpectedly rare, compared to cases involving labial or coronal consonants (Ihiunu and Kenstowicz 1994; Flemming 1995). This seems possible, but no comprehensive comparative survey has been done. Flemming's claim that they do not exist at all seems too strong.

7 The triggering vowel is assumed to be /æ/, which backs later in the derivation.

can be imagined if vowels can cause place assimilation of a consonant, are glaringly absent.[8]

Apart from this asymmetry, another interesting fact about cross-category assimilations should be noted. Compared to within-category assimilations, they seem "weak" in several respects. First, they appear to be much less frequent. This seems especially clear if we compare within-category effects in which a consonant's secondary articulation affects a vowel to cross-category effects in which a consonant's *primary* articulation affects a vowel, e.g. /k^{w}ɨ/ → [k^{w}u] *vs.* /pɨ/ → [pu]. If we keep in mind that consonants with secondary articulations occur in a minority of languages while *all* languages have plain consonants (see e.g. Maddieson 1984), the difference is very striking. The seeming exception to this generalization involves gutturals, which, when present in a language, seem likely to trigger assimilation of a vowel (McCarthy 1994; Rose 1996).

Second, cases in which a consonant's primary place takes precedence over its secondary place in determining a vowel's place features, e.g. /p^{j}ɨ/ → [p^{j}u], seem non-existent (Ní Chiosáin and Padgett 1993; Flemming 1995, 2003).

Third, cross-category effects often seem to involve vowels that are "underspecified" (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION) in the sense of being either epenthetic (CHAPTER 67: VOWEL EPENTHESIS), reduplicative (CHAPTER 100: REDUPLICATION), or central (see also CHAPTER 58: THE EMERGENCE OF THE UNMARKED). The case of Mapila Malayalam is typical: the vowel [ɨ] is not one of the language's phonemes /i e a o u/, it is predictably inserted, and it is central. Central vowels (whether predictable or not) are often hypothesized to lack specification to some degree, whether phonetically (Browman and Goldstein 1992) or phonologically (Kaye *et al.* 1985; Clements 1991; Lombardi 2003) (see also CHAPTER 26: SCHWA). That cross-category effects often seem to be linked to central and epenthetic vowels suggests that they often can be feature-filling, but not feature changing.

Fourth, cross-category effects generally seem to be highly local: the consonant and vowel must be immediately adjacent or nearly so (CHAPTER 81: LOCAL ASSIMILATION). Though some within-category effects seem to have this property too (e.g. rounding or palatalization of consonants by vowels), some clearly do not. For example, consonants can dissimilate across vowels, and yet cross-category dissimilations are local.

Finally, some cross-category effects seem to need to "gang up" in order to apply. In the case of Cantonese vowel fronting mentioned above, the vowel must be *surrounded* by coronals in order to undergo the rule. In Feʔ-feʔ-Bamileke, a labial consonant causes an adjacent reduplicating vowel to be round, but only when a round vowel is also present; likewise, a coronal consonant causes it to be front only when a front vowel is also present (Hyman 1972).[9]

These facts about cross-category effects, suggesting that they are in some sense "weak," should arguably follow from any account of them.

[8] Some apparent counterexamples are discussed below and in Ní Chiosáin and Padgett (1993). (See also CHAPTER 72: CONSONANT HARMONY IN CHILD LANGUAGE for a discussion of this possibility in child language.)

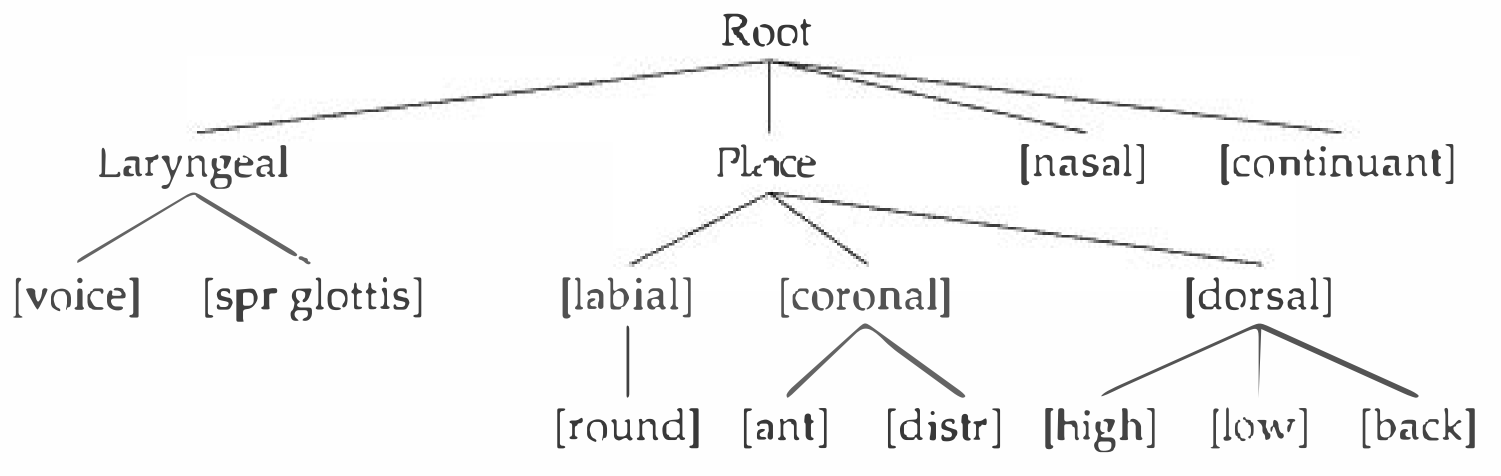
[9] If the neighboring vowel is high, it is sufficient to cause the change. Otherwise the vowel and consonant together cause it.

3 C–V interactions and feature theory

Research on C–V interactions became particularly active within the context of the development of feature geometry theory. This chapter assumes a basic familiarity with the workings of autosegmental phonology and feature geometry (see CHAPTER 14: AUTOSEGMENTS and CHAPTER 27: THE ORGANIZATION OF FEATURES.)

A good starting point for our discussion is a feature geometry representation based on the influential work of Sagey (1986), with some modifications suggested by McCarthy (1988), shown in (12) (some details omitted).

(12) *Feature geometry*



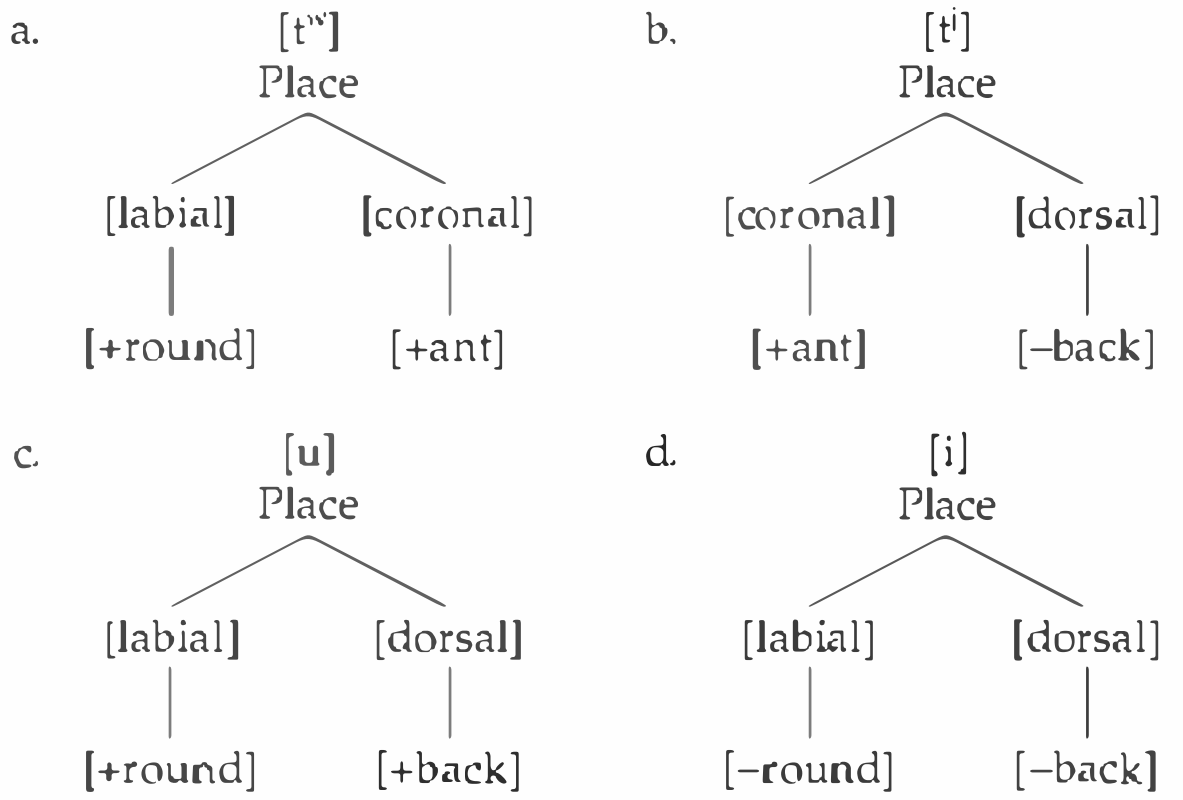
Focusing on the place of articulation features, one notable property of Sagey's model is its basis in active articulators of the vocal tract: [labial], involving the lips, [coronal], involving the tongue tip and/or blade, and [dorsal], involving the tongue body. In its grounding in articulation the model is in the tradition of *SPE*. However, Sagey's model departs from *SPE* in various respects, including in holding that the articulator nodes [labial], [coronal], and [dorsal] are privative, and that they are organizational nodes in feature geometry, as shown in (12).[10]

As Sagey argues, an advantage of an articulator-based model like this is that it easily represents complex segments – segments that have more than one place of articulation, such as [g͡b] (CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION). In *SPE*, by comparison, velars are [−anterior, −coronal] and labials are [+anterior, −coronal]. In such a system it is unclear how to specify a segment that is both labial and velar. This point is relevant to us, since consonants bearing vocalic secondary articulations are complex segments. For example, in Sagey's terms the segments [t^w] and [t^j] are represented as in (13a) and (13b), focusing only on place features.[11] Vowels are also specified in terms of the features in (12) and are often complex segments themselves. The vowels [u] and [i], for instance, are specified as in (13c) and (13d). The representation in (13d) assumes that [i] has an active lip-spreading gesture, which requires involvement of the lips. Since all vowels are specified for tongue body features, all vowels have a [dorsal] specification.

[10] For Sagey they are privative because they are *class nodes* rather than features. Like Clements (1991) and others, I interpret them as features and assume features can be dependent on other features.

[11] Some features are omitted for simplicity, including [distributed], [high], and [low]. [t^w] is understood as labialized, not labial-velarized.

(13)



To give a more complete understanding of this feature system, Table 75.1 shows place specifications for plain, palatalized, and rounded consonants of all three major places of articulation and for five representative vowels. The symbol "✓" indicates specification of a privative (major) place feature. In this theory, specification of a feature such as [round] or [back] is possible only if the relevant major place – here [labial] or [dorsal] respectively – is specified. Otherwise, full feature specification is assumed for the sake of discussion. For our purposes, what is particularly worth noticing is the disjointedness of the consonantal *vs.* vocalic place specifications. Unless they are labialized or palatalized, coronal consonants have nothing in common with vowels. Plain labials are like rounded vowels in having a [labial] specification – but not in being rounded. The exception to this disjointedness is with [dorsal] consonants, which, following *SPE*, are specified for vocalic tongue body features. This is in fact how velars are distinguished from uvulars ([–high, –low]) and pharyngeals ([–high, +low]).

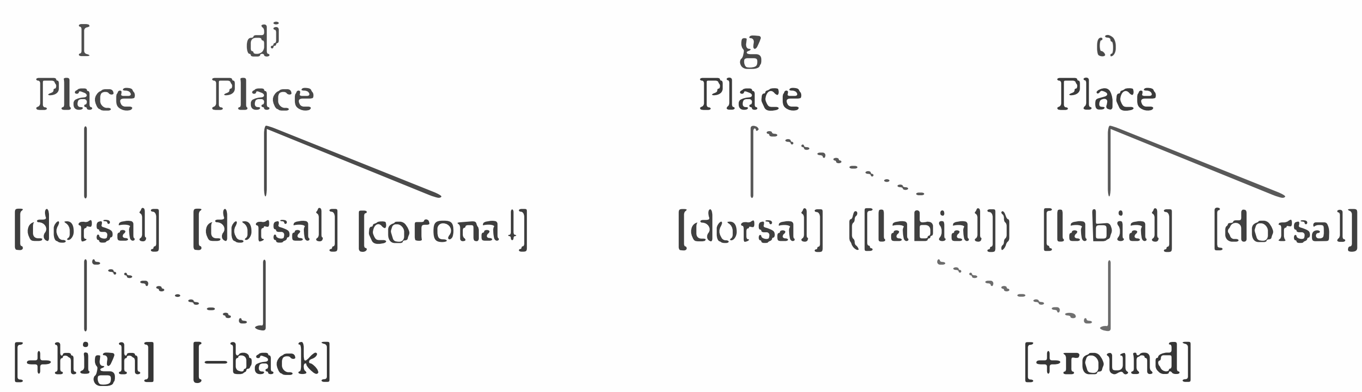
Within such a model, many within-tier interactions are straightforward to represent. For example, the dependence in Irish of a short vowel's backness

Table 75.1 Feature specifications (Sagey 1986)

		p	pʲ	pʷ	t̪	t̪ʲ	t̪ʷ	k	kʲ	kʷ	i	y	a	o	u
[labial]		✓	✓	✓			✓			✓	✓	✓		✓	✓
	[round]	–	–	+			+			+	–	+		+	+
[coronal]					✓	✓	✓								
	[anterior]				+	+	+								
	[distributed]				+	+	+								
[dorsal]			✓			✓		✓	✓	✓	✓	✓	✓	✓	✓
	[high]		+			+		+	+	+	+	+	–	–	+
	[low]		–			–		–	–	–	–	–	+	–	–
	[back]		–			–		+	–	+	–	–	+	+	+

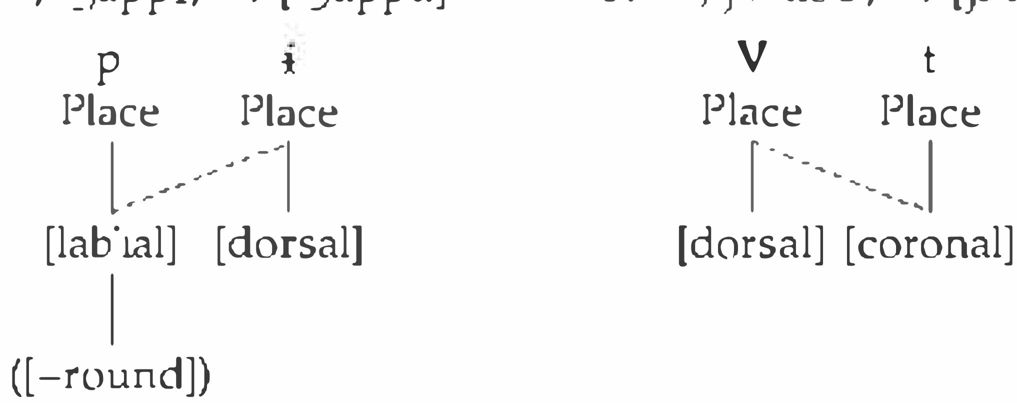
on the following consonant (see (3)) is an assimilation as shown in (14a).[12] The assimilation of a plain consonant to a round vowel, as in Nupe [egʷo] 'grass', is as in (14b). (The rule spreads [+round], and [labial] is inserted on the consonant by Node Interpolation (Sagey 1986). Alternatively, [labial] spreads to the consonant.) Apart from the advantages of an articulator-based representation, it is not the particular geometry assumed here that makes within-category effects easy to represent. The point is that vowels, glides, and (semi-)vocalic secondary articulations on consonants are assumed to employ the very same set of place features. If that is true, then whatever the geometry, there is no (at least general) difficulty in representing within-category processes.

(14) a. /mˠɪdʲ/ → [mˠɪdʲ] b. /ego/ → [egʷo]



Compare the situation with cross-category assimilations. Recall that in Mapila Malayalam the vowel /ɨ/ is rounded after a plain labial consonant (see (6)). Since it is a plain (not labialized) labial consonant in question, it cannot be specified [+round]. Whether the consonant is specified [–round] or unspecified for [round], spreading [labial] as in (15a) will not cause the vowel to round. In the case of Maltese Arabic (see (7)), the problem seems even worse. The feature [–back] needed to achieve [i] is not part of the representation of a coronal consonant.[13] Spreading [coronal], the only seeming option, is of no help. The issue once again is not about the feature geometry pursued. The problem is that the place features that are assumed to define consonantal place are largely disjoint from those assumed to define vocalic place. There is therefore no straightforward way to explain why plain consonants can affect vowels in this way.

(15) a. ??/ʤappɨ/ → [ʤappu] b. ??/jV-tlob/ → [ji-tlob]

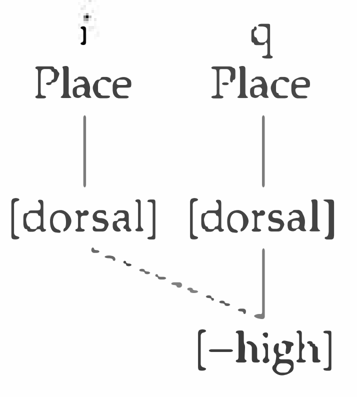


This problem of disjoint place features for consonants and vowels has been noted for some time. For example, before the advent of autosegmental phonology or feature geometry, Campbell (1974) and Clements (1976) had pointed out the

[12] Irrelevant detail will often be omitted in representations shown.
[13] Hume (1994) argues that the vowel's height is a default value and so doesn't need to be spread.

problem for the *SPE* feature theory of processes like those of Mapila Malayalam and Maltese, respectively. Campbell noted the lack of connection between the features [labial] and [round], and Clements did the same for [coronal] and [back]. By comparison, cross-category effects involving [dorsal] consonants make sense in the feature theory of *SPE*/Sagey (1986), because such consonants are specified for [high], [low], and [back]. For example, the vowel lowering of Inuktitut (see (8)) can be represented as in (16). The point holds equally if Inuktitut is better interpreted as [RTR] spread from a consonant. To put it differently, according to this theory assimilations by vowels to [dorsal] consonants are, in a sense, within-category effects: the influence of [dorsal] consonants is accomplished through vowel place features.

(16) /seʀmi-q/ → [seʀmɛ-q]



4 Unified feature theory

The idea that consonants and vowels should be specified by the same set of place features has been motivated by researchers in diverse frameworks, including Schane (1984, 1987), Kaye *et al.* (1985), Anderson and Ewen (1987), Selkirk (1988, 1993), van der Hulst (1989), and Clements (1991). Selkirk and Clements cast the idea roughly in terms of the features of Sagey (1986), as shown in Table 75.2.[14] Following McCarthy (1994), a feature [pharyngeal] is now included. McCarthy argues that uvular, pharyngeal and, for at least some languages, laryngeal consonants have in common a [pharyngeal] specification.

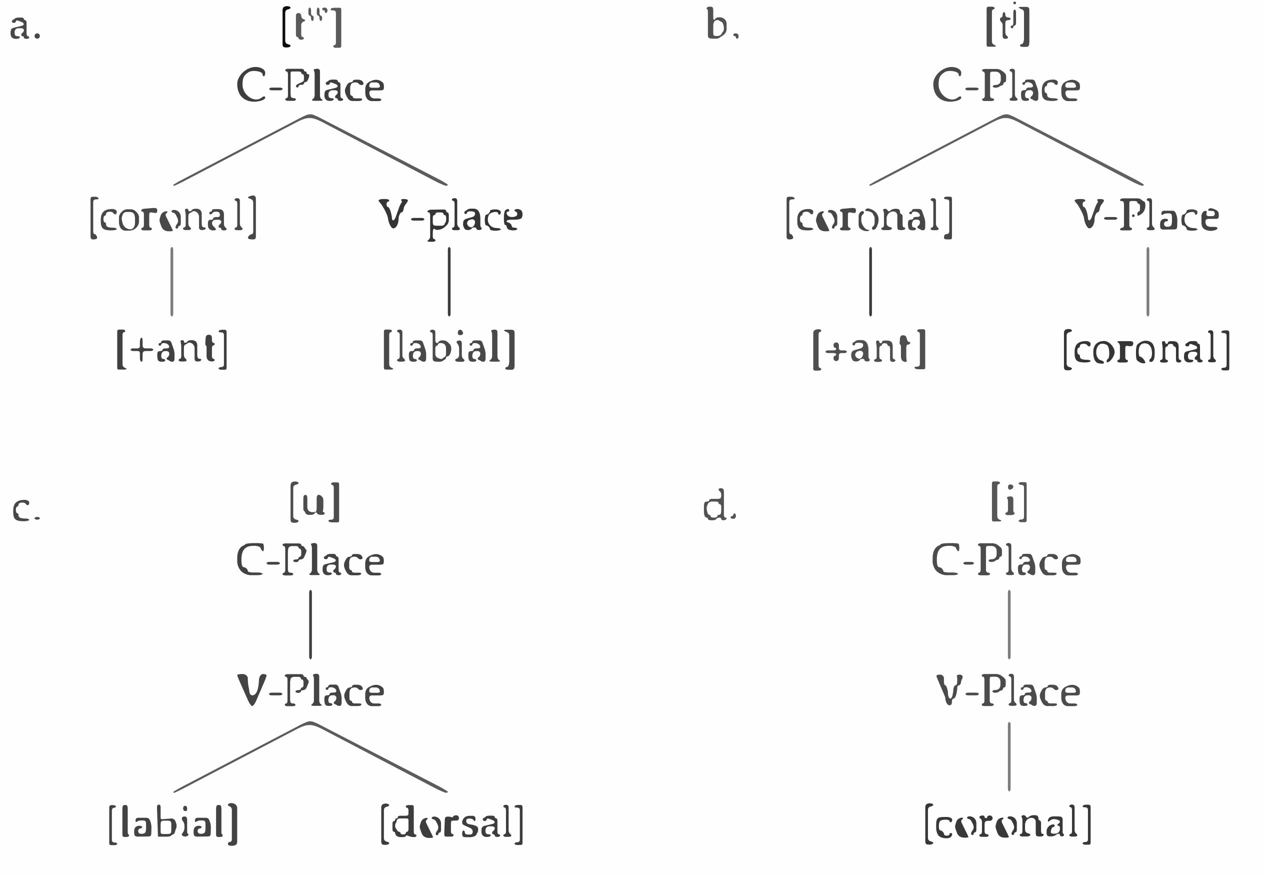
Table 75.2 Unified place features for consonants and vowels (Clements 1991)

		p	pʲ	pʷ	t̪	t̪ʲ	t̪ʷ	k	kʲ	kʷ	ħ	i	y	a	o	u
C-Place	[labial]	✓	✓	✓												
	[coronal]				✓	✓	✓									
	[dorsal]							✓	✓	✓						
	[pharyngeal]										✓					
V-Place	[labial]			✓			✓			✓			✓		✓	✓
	[coronal]		✓			✓			✓			✓	✓			
	[dorsal]													✓	✓	✓
	[pharyngeal]													✓		

[14] However, Clements treats these features as binary valued. The features [anterior] and [distributed] are not shown here.

The unified feature approach capitalizes on the apparent articulatory parallelism between consonants and vowels: both labial consonants and round vowels involve a constriction at the lips; both coronal consonants and front vowels involve a constriction at the tip/blade/front of the tongue; both dorsal consonants and back vowels involve a constriction at the tongue dorsum; and both pharyngeal consonants and low vowels involve a constriction between the tongue root and the pharynx wall. (The parallelism in the case of [coronal] is the most questionable, as we will see later.) For Clements (1991) and others working within this framework (including Herzallah 1990; Hume 1990, 1994), the consonant–vowel parallelism does not extend to vowel height (or stricture features in consonants). Distinct features are still needed for these properties of segments. This means that the unified feature approach obviates the vowel color features [back] and [round], but not [high] and [low].

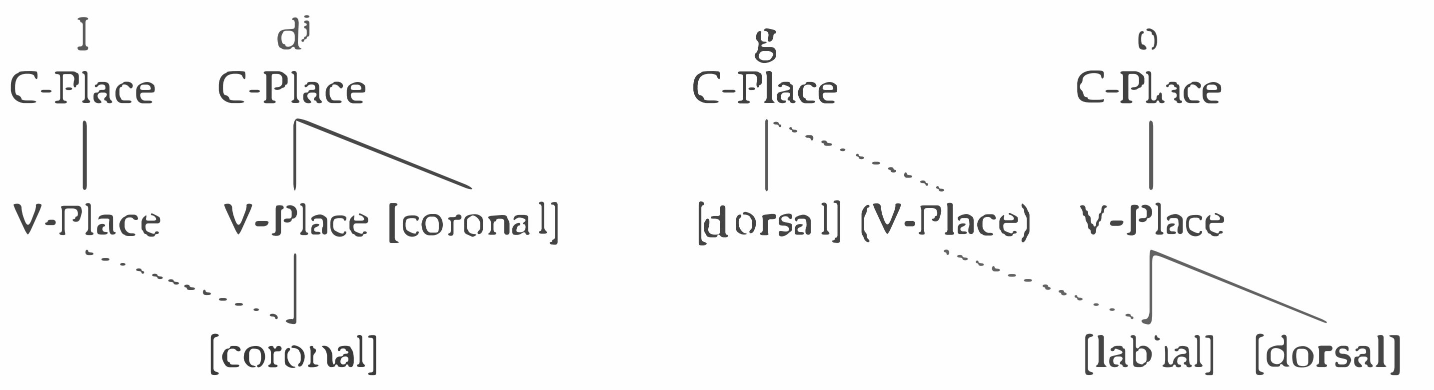
Of course, vocalic rounding is not articulatorily identical to the labial constriction of a consonant; likewise for the other parallel features. For these unified features to be phonetically interpreted we require reference to a segment's manner features. For example, if a sound is specified as [−consonantal] then [labial] is interpreted as lip rounding. Alternatively, the relevant information is read off feature-geometric structure. Thus Herzallah (1990), Clements (1991), Hume (1994, 1996), and Clements and Hume (1995) locate [labial], [coronal], [dorsal], and [pharyngeal] under separate C-Place and V-Place nodes, depending on whether a consonantal or vocalic constriction is intended. In these terms, the segments seen above in (13) are now rendered as in (17). Segment (17a) is interpreted as [t^w] because [labial] is a V-Place feature while [coronal] is a C-Place feature; and so on for the other representations.[15]

(17)


[15] These representations simplify the full geometry assumed by the references cited, to focus on what is crucial here. There are reasons for assuming that V-Place is a dependent of C-Place (or "Place" according to some), instead of a sister, for example, but a consideration of these would take us too far afield. See Clements (1991), Odden (1991), and Ní Chiosáin (1994) for discussion of this issue and for motivation of V-Place as a feature-geometric constituent.

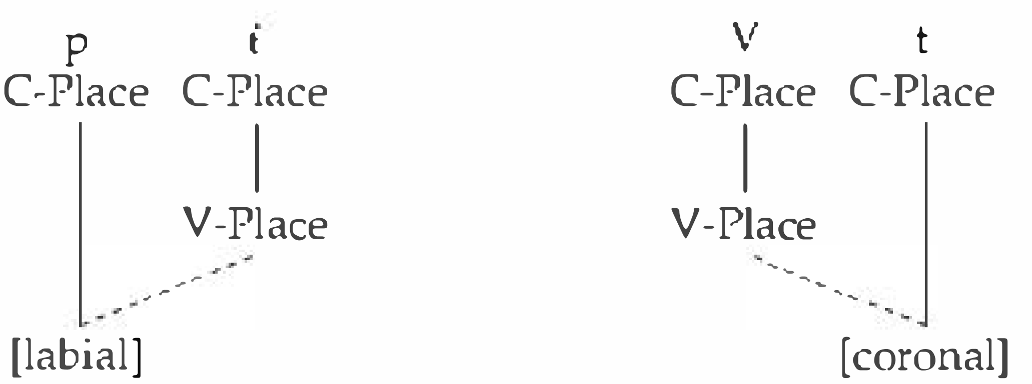
Naturally, it remains true in this theory that vowel place features and secondary vocalic articulation features on consonants are the same. Therefore it remains straightforward to characterize within-category assimilations as seen in (14) above, now understood as in (18). Cases such as Inuktitut (see (8)) are also arguably within-category, as noted above.

(18) a. /ɪnˠɪdʲ/ → [ɪnˠɨdʲ]   b. /ego/ → [egʷo]



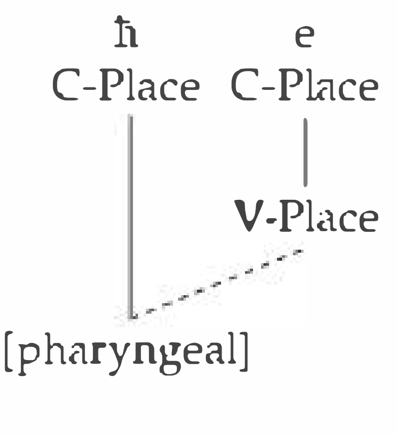
What is new with unified features is the possibility of directly capturing cross-category assimilations too. Compare the representations in (19) to the problematic (15) above. In (19a) (Mapila Malayalam), [labial] spreads from a consonant to a vowel. Notice that [labial] is linked to C-Place for the consonant and to V-Place for the vowel. It is therefore interpreted as consonantal lip constriction for the consonant and as rounding for the vowel. Similar reasoning holds for (19b) (Maltese; see Hume 1994, 1996). Cases of backing around dorsal consonants, as in Maxakalí (see (9)), similarly involve the spreading of [dorsal] from consonant to vowel.

(19) a. /ʤappɨ/ → [ʤappu]   b. /jV-tlob/ → [ji-tlob]



Finally, cases in which a vowel becomes [a] around a guttural consonant, as in Syrian Arabic (see (10)), are analyzed as the spreading of [pharyngeal] (Herzallah 1990; McCarthy 1994; Rose 1996), as shown below.

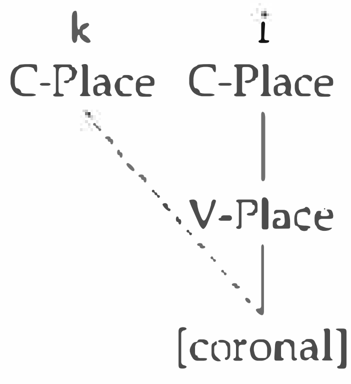
(20) /mniːħ-e/ → [mniːħ-a]



In short, unified feature theory solves the problem of cross-category interactions by eliminating the disjointedness of consonantal and vowel place features.

If features can spread from C-Place to V-Place, as in (19) and (20), then we might expect that the reverse can happen. This is just what is proposed for palatalizing mutations of the sort seen in Slovak, where /k g x ɣ/ become /ʧ ʤ ʃ ʒ/ respectively before any of /j i e æ/ (see (11)). Since front vowels are characterized as [coronal] instead of [−back] in the unified theory, this kind of mutation can be viewed as assimilation, specifically "coronalization" (Broselow and Niyondagara 1989; Mester and Itô 1989; Pulleyblank 1989; Lahiri and Evers 1991; Hume 1996):

(21) /vnuk/ → [vnuʧik]



The outputs of the Slovak rule are not simply [coronal]; they are fricated or affricated palato-alveolars. Hume (1994, 1996) reasonably assumes that front vowels are [−anterior] coronals. This entails that "coronalization" will output [−anterior] coronals too. The rest must follow from redundancy rules like [−anterior] → [+delayed release].

As noted above, this approach to unified features does not attempt to unify features for vowel height and consonantal stricture. This means that, when such features are affected by cross-category assimilation, it must be for independent reasons. For example, Hume (1994, 1996) argues that the front vowel derived in Maltese is [+high] [i] because this is the default height for vowels in the language.

On the other hand, since [+low] is not a likely default height in Arabic, Herzallah (1990: 185) assumes that [pharyngeal] spreading as in (20) leads to a low vowel because of a redundancy rule [pharyngeal] → [+low].[16] Since gutturals seem to cause assimilation to [a] typically, this redundancy rule will be needed for other cases too. This might be seen to somewhat undermine the argument of unified feature theory. The point of unified features is to capture the assimilatory nature of cross-category effects. The lowering that occurs around gutturals seems just as assimilatory as the spreading of the pharyngeal constriction, so why treat it differently? The case of Maxakali (see (9)) also supports the view that consonants can affect vowel height as well as vowel color. Recall that the inserted vowel in that language is [ə] before alveolars; Clements (1991) suggests that this is the default inserted vowel. But before velars the inserted vowel is both back and high, [ɯ]. A redundancy rule [dorsal] → [+high] might work, but, side-by-side with [pharyngeal] → [+low], it begs the question why we do not allow that consonants directly affect height as well as color.

This question aside, unified features have a clear appeal. They explain cross-category effects, because they assume that coronal consonants and front vowels form a natural class, as do labial consonants and round vowels, etc. Apart from

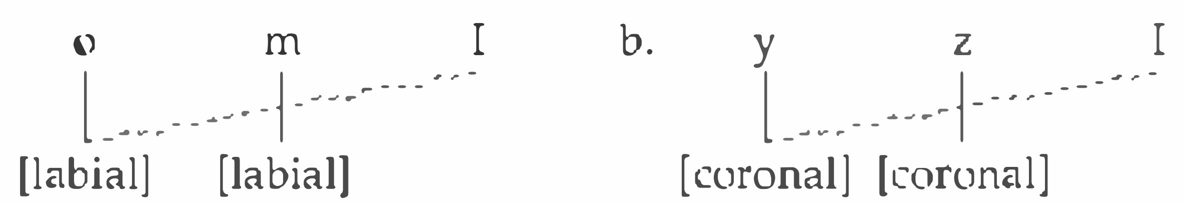
[16] Herzallah discusses Palestinian Arabic in this context. She also employs the vowel aperture features of Clements (1991), rather than [high] and [low].

assimilations and dissimilations like those already seen, more evidence for such natural classes comes from instances of vowel strengthening or consonantal weakening. For example, when the vowel [i] and glide [j] are strengthened to consonants, they are strengthened to coronal consonants, or at least consonants with a coronal component – palatals or palato-alveolars. This occurs in Porteño Spanish when the (semi-)vowel is in onset position, as in /jelo/ (or /ielo/) → [ʒelo] 'ice' (see Harris and Kaisse 1999 and references therein). A relevant example of weakening comes from Irish lenition (CHAPTER 66: LENITION), whereby /b m/ are reduced to [w] (see Ní Chiosáin 1991). If front vowels and palato-alveolars are both [coronal], and if labial consonants and [w] are both [labial], then these processes can be understood as the "promotion" or "demotion" of those features in the C-Place/V-Place representation. As with the assimilations, however, such accounts will often need the help of redundancy rules.

5 Non-interaction and locality

Our discussion of C–V interactions has so far ignored an important issue. There are ways in which consonants and vowels apparently *fail* to interact, and our theory needs to explain these too. Perhaps the most basic question arises from the simple observation that consonants typically seem to be transparent to vowel harmonies and other kinds of vowel-to-vowel place assimilations (CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS; CHAPTER 118: TURKISH VOWEL HARMONY; CHAPTER 123: HUNGARIAN VOWEL HARMONY). In Turkish, for example, vowels harmonize for roundness as well as backness (Lees 1961; Clements and Sezer 1982). Most consonants are transparent to the harmony.[17] Particularly relevant to the discussion here, labial consonants are transparent to round harmony, as in [somun] 'loaf', and coronal consonants are transparent to backness harmony, as in [økyz] 'ox'.[18] The issue raised by such cases is schematized in (22).

(22) a. b.



Unification of place features for consonants and vowels is motivated by the cross-category interactions we have seen so far. However, the ability to *block* spreading is also a kind of interaction (if a passive one), and the principles of autosegmental phonology imply that spreading as in (22) *should* be blocked. A similar implication arises for vocalic [dorsal] spreading through dorsal consonants (not shown). These representations cross lines, a maneuver ruled out within autosegmental phonology for features on the same tier.[19]

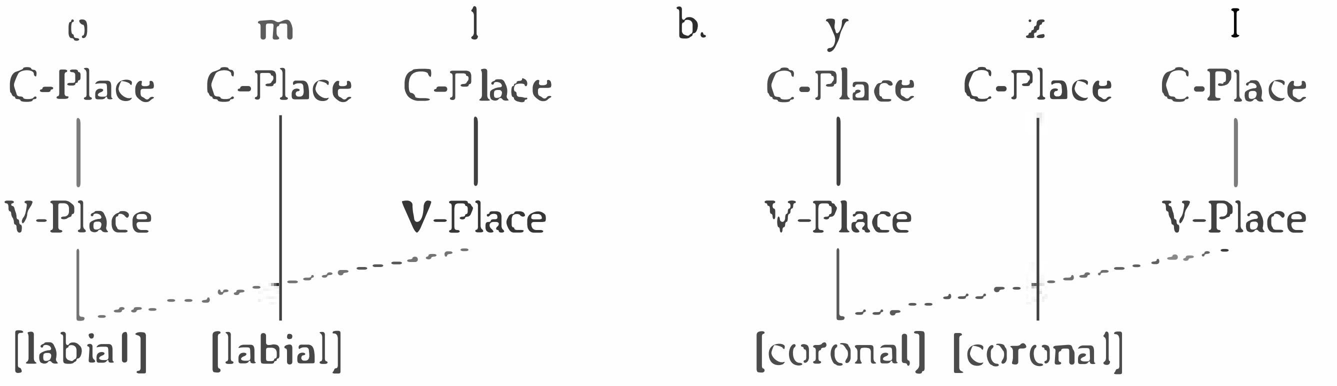
[17] The exceptions are palatalized consonants in certain limited circumstances. Since palatalization is a V-Place specification, this blocking is *within*-category. Blocking in such cases is the rule across languages, in contrast to the situation with plain consonants.

[18] These examples involve harmony in the root, but the observation about transparency holds equally for harmony between a stem and a suffix.

[19] See Hammond (1988), Sagey (1988), Bird and Klein (1990), Coleman (1991), Scobbie (1991), and Archangeli and Pulleyblank (1994) on deducing the ill-formedness of line crossing within the theory.

As we have seen, Clements (1991), Clements and Hume (1995), and others working within this unified features framework locate vocalic and consonantal place features under distinct nodes in feature geometry, V-Place and C-Place respectively. The full representations for the scenarios in (22) (modulo some irrelevant simplifications) are shown in (23) below. In feature geometry, a *plane* on which association lines spread is defined by adjacent tiers. Therefore, the plane defined by the [labial] tier and the C-Place tier in (23a) is different from that defined by the [labial] tier and the V-Place tier. Clements (1991) and Clements and Hume (1995) suggest that, even when the same feature such as [labial] is involved in the spreading, line crossing is prohibited only *within a plane*. Therefore spreading, as in (23a) and (23b), is allowed. (Put differently, apparent line-crossing is only a problem when the crossed lines link to the same mother node in the geometry.)

(23) a. b.



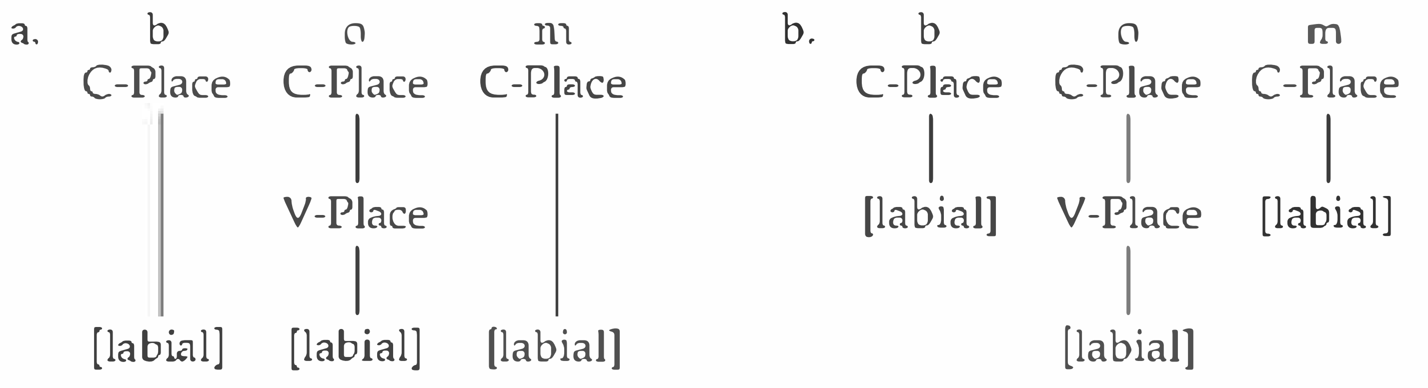
This suggestion raises questions about the formal understanding of tiers and planes that have not been fully explored. In any case, the worry about non-interaction of C-Place and V-Place features goes beyond this kind of spreading. For example, many languages place restrictions on homorganic consonants occurring within forms (CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS). In autosegmental phonology these have been explained by means of the Obligatory Contour Principle, which prohibits tier-adjacent identical feature specifications (see for example McCarthy 1986; Mester 1986; Yip 1989; Frisch *et al.* 2004 and references therein) (see also CHAPTER 14: AUTOSEGMENTS). Such restrictions can apply to consonants separated by vowels, and, crucially, do not seem to be blocked even by vowels of the "same" place of articulation; that is, forms such as [bom] are as ill-formed as [bam].[20] But given unified features, the consonants' [labial] features in (24a) are not tier-adjacent, since the vowel's [labial] intervenes. Why should this form be dispreferred? (The answer is *not* because sequences such as [bo] or [om] themselves are ruled out; they are not in most languages having this kind of dissimilation.) Analogous issues arise with front vowels and coronal consonants, etc., and the same general issue arises in reverse when vowels dissimilate across all consonants, as in Ainu (Itô 1984).[21] To deal with this question, Clements (1991) suggests that instances of a feature are not on the same tier when they are dominated by different mother nodes, and in fact to highlight this point he

[20] One exception is Akkadian (McCarthy 1979; Yip 1988; Hume 1994; Odden 1994). A prefix /m/ dissimilates to [n], given another labial consonant in the stem. The sounds [u w] do not trigger dissimilation, but they *do* block it.

[21] There are other theoretical approaches to dissimilation that do not appeal to tier-adjacency, including the idea of local self-conjunction advocated by Itô and Mester (1996) and Alderete (1997). Such approaches might avoid the question raised by (24).

draws representations as in (24b).[22] Hume (1994, 1996) and Clements and Hume (1995) frame a similar idea differently: there is one [labial] tier, as in (24a), but two instances of a feature can fail to interact when they are dominated by different mother nodes. Therefore the [labial] features dominated by C-Place may interact with each other and may each fail to interact with the intervening [labial].

(24)



Obviously there should be concern at this point about losing the gains made with unified features. If C-Place and V-Place features are on different tiers, or if they can fail to interact because C-Place and V-Place are different mother nodes, then why do C-Place and V-Place features *ever* interact? Selkirk (1988, 1993) considers many of the same issues, and makes the important observation that cross-category dissimilations and co-occurrence restrictions seem to hold only under segmental adjacency.[23] Both Selkirk and Clements suggest that this observation be elevated to a principle. A similar, though more general, observation was made at the end of §2 above: cross-category dissimilations *and* assimilations, unlike within-category cases, are always highly local.

To summarize: according to a unified features view of C–V interactions, consonant and vowel place features are unified, and so can interact. But cross-category interaction seems limited to (near-)segmental adjacency, and a unified feature theory must address this limitation by separate stipulation (e.g. interaction can happen across tiers/with different mother nodes only under (near-)segmental adjacency). If the empirical observations here are on the right track, one might still raise questions about the account. In particular, the very motivation for unified features seems weakened by the need to stipulate *non-interaction* except under close adjacency. In addition, the latter stipulation does not follow from anything else in the theory; cross-category effects are limited in a way not really explained.

6 Inherent vowel place specifications

An alternative approach to cross-category C–V interactions, also couched within feature geometry theory, is proposed by Ní Chiosáin and Padgett (1993) and Flemming (1995, 2003): perhaps "plain" consonants are not as plain as they are assumed to be.

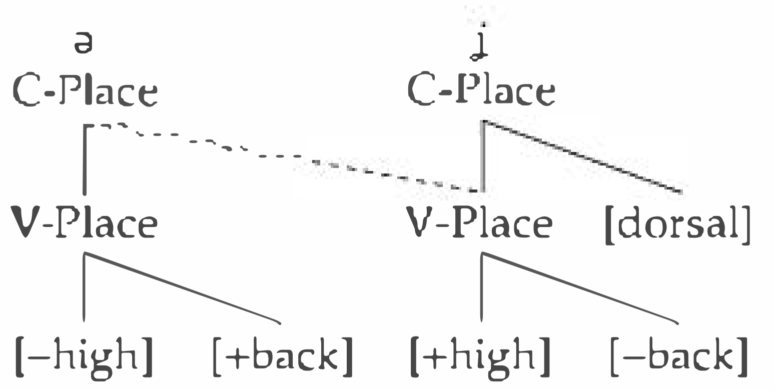
[22] This redefinition of the notion "tier" may render unnecessary the reference to planes in (23). If consonantal and vocalic [labial] are on different tiers, then there is by definition no line crossing (in the relevant sense) in such cases.

[23] Selkirk's notion of "cross-category" is actually more abstract than that employed here. Her notion makes some different empirical predictions.

To begin with, many consonants have been claimed to be complex segments, specified for both consonantal and vowel place features, even though they are "plain" in the sense of lacking secondary articulations that are transcribed. For example, there is a long-standing view that palatals, alveolo-palatals, and at least certain palato-alveolars are inherently specified for features indicating a high and front tongue body, and studies of their articulatory properties support this view (see Keating 1988, 1991; Keating and Lahiri 1993). This fact also explains why these segments are the most common outputs of palatalizing mutations of velars and coronals, assuming these mutations involve assimilation. Similarly, uvulars and pharyngeals may involve inherent specifications for vocalic tongue body and/or root position. In this view, they cause lowering or retraction of vowels because they are themselves specified for a feature like [−high] or [RTR] (see Chomsky and Halle 1968; and more recently Elorrieta 1991; Halle 1995; Rose 1996). If these consonants involve inherent vowel place specifications – in addition to independent consonantal specifiications – then effects such as the lowering or retraction of vowels before uvulars in Inuktitut, or the raising and fronting of vowels before palatals and palato-alveolars in Kabardian, e.g. /ʒəʝ/ → [ʒɪʝ] 'tree' and /ʒaʃ/ → [ʒɛʃ] '(to) be bored, tired' (Colarusso 1992: 30), are not cross-category assimilations at all.

For example, the raising and fronting seen in Kabardian might be understood as in (25). For the sake of discussion, we revert to the familiar vowel place features of *SPE*, but following the literature on V-Place constituency (see note 15) continue to assume this aspect of the geometry. The palatal fricative [ʝ] is assumed to have (at least) a primary [dorsal] specification. The point is that what spreads in this case are features of tongue body height and frontness that are uncontroversially relevant to vowels; in effect, palatals (and palato-alveolars in Kabardian) are understood as inherently palatalized segments.

(25) /ʒəʝ/ → [ʒɪʝ]



The question raised by work such as Ní Chiosáin and Padgett (1993) and Flemming (1995, 2003) is to what extent features relevant to vowel place might inhere within other "plain" consonants. If all cross-category C–V interactions were caused by such inherent features, then "cross-category," though a useful classificatory term, would lose any theoretical import: all C–V interactions would be within-category.

Ní Chiosáin and Padgett (1993) approach the phonetic claim from an articulatory point of view, following the general articulatory approach to features in the *SPE* tradition. They note that consonantal constrictions involve offsets and onsets, i.e. movement from a previous position into the consonant, and movement from the consonant into a following position. The idea is diagrammed as in (26), for a labial consonant in an intervocalic context such as [ɨbɨ]. During the onset transition there is a short period of time when the constriction is not yet consonantal, and yet this vocalic period is shaped by the impending consonant. Likewise, early in the release the constriction becomes vocalic while still shaped by the preceding consonant. As the diagram implies, this must have effects on the quality of the vowel near the closure.

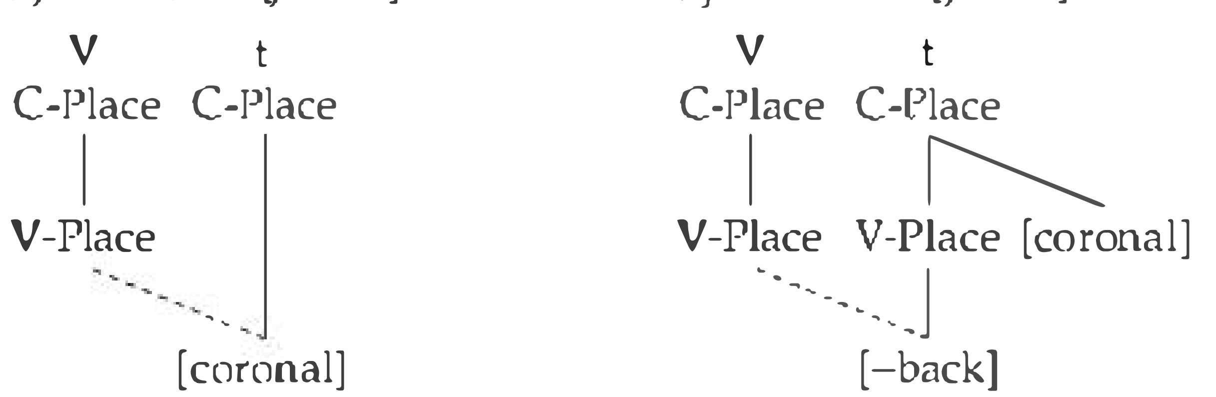
(26) *Stages of consonant production*

ɨ " b " ɨ
vowel onset closure offset vowel
→→→→→→→→→→→

Flemming (1995) emphasizes these acoustic or auditory effects of the transitions into and out of consonants, and in fact argues for the incorporation of auditory-based features into phonological theory. It is well known that consonantal offsets and onsets influence vowel formants; in fact, the resulting dynamic formant transitions are an important cue to both vocalic and consonantal place. Flemming notes that labial consonants, for example, lower vocalic F2 (second formant) values, while dental and alveolar coronals raise F2.[24] Since vowel place (particularly vowel color) is cued by the location of F2, consonants therefore have the inherent ability to affect perception of vowel place.

Flemming (2003) discusses the articulatory basis of the acoustic effects in more detail, focusing on coronal consonants. According to his survey of relevant studies, anterior coronals and palato-alveolars tend to cause tongue body fronting, due to coupling between the tongue blade/tip and the tongue body, and this is the reason for the rise in F2 around such consonants. With this in mind consider once again the fronting of vowels by coronal consonants, as in Maltese Arabic (see (7)). The idea is that fronting happens not because the consonant spreads its primary [coronal] articulation as in (27a), but because the consonant has some inherent tongue fronting, with a concomitant effect on F2, and this is what spreads to the vowel, (27b). For the sake of discussion this is formalized in terms of the traditional feature [back].

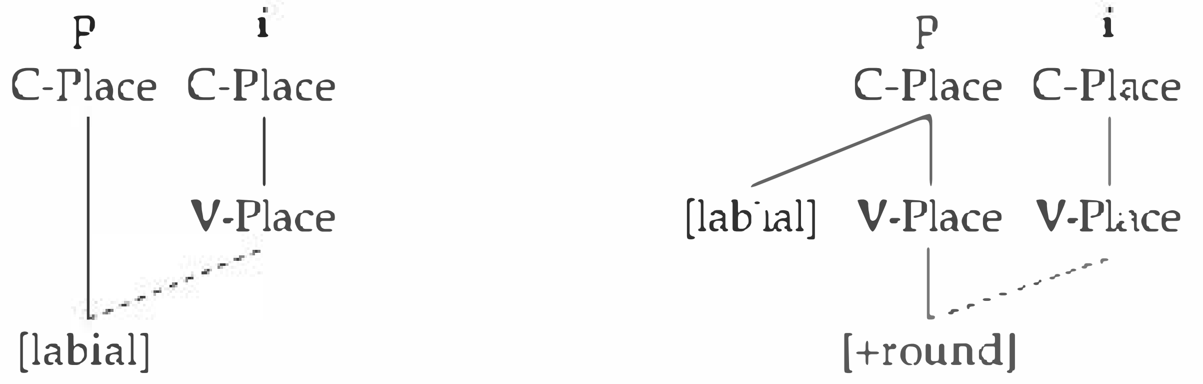
(27) a. /jV-tlob/ → [ji-tlob] b. /jV-tlob / → [ji-tlob]



Similarly Mapila Malayalam (see (6)) involves not spreading of primary [labial] as in (28a) but spreading of the inherent vocalic labial constriction, and lowering of F2, formalized here by means of [+round], (28b).

[24] Discussion of vowel formants and of formant transitions can be found in e.g. Stevens (1998) and Johnson (2003). Another advocate of auditory features in phonology is Boersma (e.g. 1998).

(28) a. /dʒappi/ → [dʒappu] b. /dʒappi/ → [dʒappu]

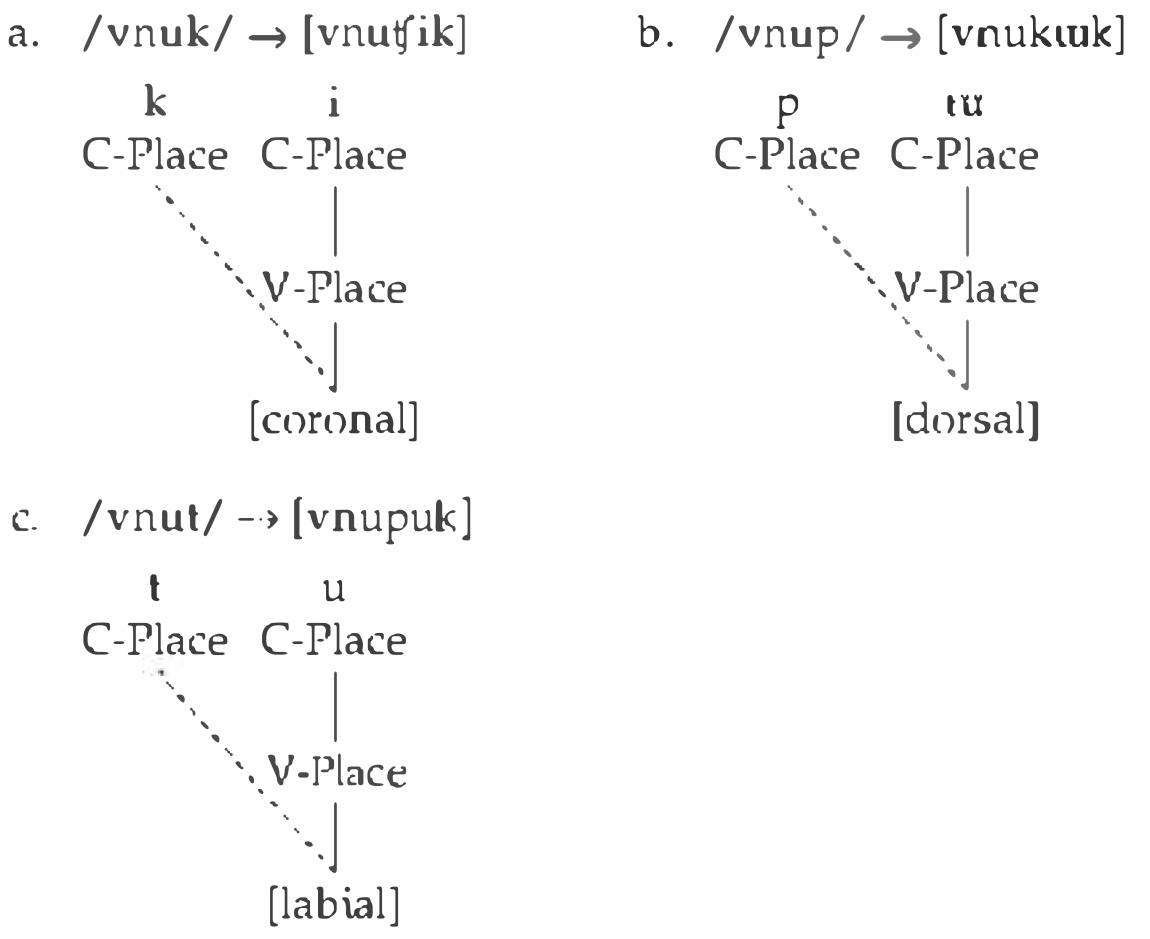


In an analogous fashion, velars can spread inherent [+back] and/or [+high] in cases like Maxakalí (see (9)), and gutturals can spread inherent [+low] in cases like Syrian Arabic (see (10)).

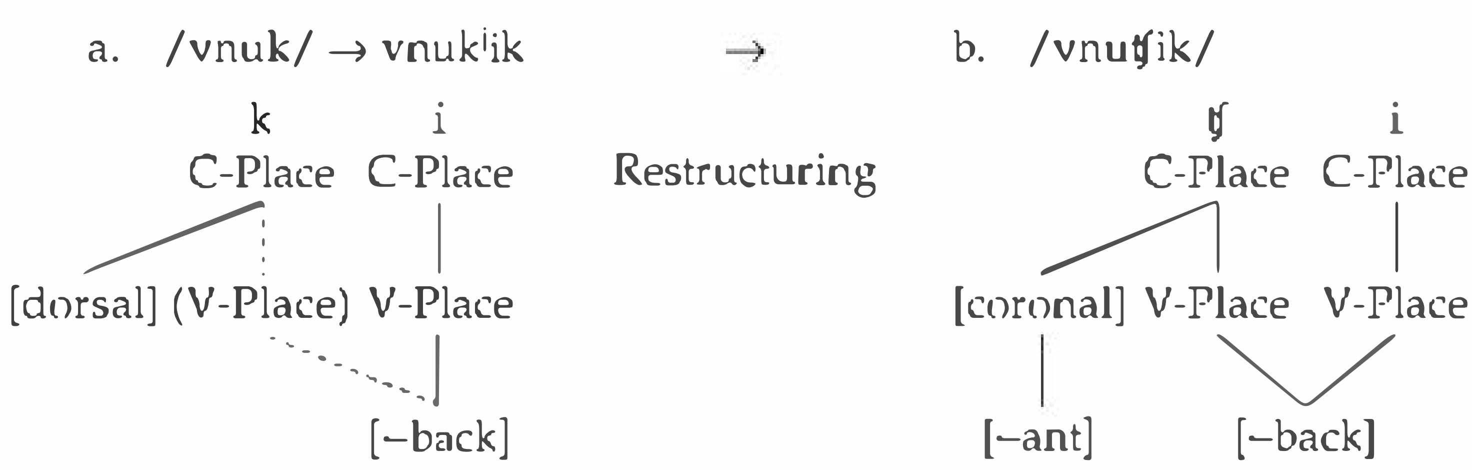
It is uncontroversial that consonants phonetically affect vowels both articulatorily and acoustically, as this approach to "cross-category" interactions assumes. However, incorporating these relatively small phonetic effects into phonology raises questions. One question is prompted by our use of the conventional vowel place features [back], [round], etc. in the account above. How are we justified in specifying the Maltese [t] as [−back] if it isn't palatalized [tʲ], or the Malayalam [p] as [+round] if it isn't rounded [pʷ]? The answer suggested by Ní Chiosáin and Padgett (1993) is that such specifications are realized as bona fide secondary articulations when they are contrastive, but not when they are redundant. In a language such as Russian, where palatalized coronals contrast with non-palatalized ones, [−back] on [t] is realized as [tʲ]; in a language without this contrast, redundant [−back] is realized as [tʲ] (notation borrowed from Ní Chiosáin and Padgett 2001; Flemming 2003), a coronal with only the inherent tongue body fronting described above. This answer to the question predicts that non-palatalized coronals in Russian could not be [−back] even inherently, since this specification is reserved for palatalization. This prediction is correct; in fact, non-palatalized sounds in Russian are velarized, i.e. [tˠ].

The inherent vowel place approach to "cross-category" effects arguably has some advantages over the unified feature approach. First, it helps explain an asymmetry in cross-category assimilations noted in §2: while vowel-to-consonant assimilation occurs, apparent consonant-to-vowel assimilations are strikingly underattested (Ní Chiosáin and Padgett 1993). The most robust example of the latter, as we saw, involves palatalizing mutations as in Slovak, where /k g x ɣ/ become [tʃ dʒ ʃ ʒ], respectively, before any of [j i e æ]. Within a unified features approach, some have argued that these are instances of "coronalization" (Broselow and Niyondagara 1989; Mester and Itô 1989; Pulleyblank 1989; Lahiri and Evers 1991; Hume 1996), as in (29a). A challenge for this claim is the absence of mutations resembling those in (29b) and (29c), or indeed resembling most of the logically possible consonant-to-vowel assimilations, if place features can link to both V-Place and C-Place.[25]

[25] Ní Chiosáin and Padgett (1993) dismiss some claimed cases that are attested as sound changes but have no synchronic reflex. See that work for discussion of other apparent counterexamples.

(29)


Ní Chiosáin and Padgett (1993) argue that consonants simply do not assimilate to vowels in this way, having their primary place displaced, and that the theory should not allow them to. There is a natural explanation within inherent vowel place theory for this asymmetry in the direction of "cross-category" assimilations: consonants can affect vowels because they can have secondary vocalic articulations (whether distinctive or redundant). But vowels by definition do not have C-Place features with which to affect a consonant; they only have V-Place features, features that can only impose secondary vocalic articulations on consonants. Put differently, *all* C–V interactions are interactions between vowel place features. To handle palatalizing mutations, Ní Chiosáin and Padgett, following many others, assume that assimilation only partially derives the output, as in (30a). Further changes from /k^j/ to e.g. [ʧ] must be due to language-particular segmental well-formedness conditions, leading to something like (30b).[26]

(30)


As Clements and Hume (1995: 295–296; see also references therein) point out, however, such restructuring fails to explain why [coronal] in particular results in the context of a front vowel, just the fact that unified feature theory

[26] Though expressed as a rule here, in line with all of this discussion, the idea is commonly expressed in Optimality Theory by means of constraint rankings such as *k^j >> ʧ. This ranking, along with a high-ranking constraint driving [–back] assimilation, will lead to the output [ʧ].

explains. On the other hand, one might seek perceptual explanations for such arbitrary articulatory connections, by means of auditory-based features (see Flemming 1995 for this case in particular), or perhaps via the "p-map" (Steriade 2001).

Another advantage to the inherent vowel place approach is that it allows for more nuance in the ways that consonants can affect vowels (Flemming 1995, 2003). For example, Flemming (2003) (citing Emeneau 1970; Ebert 1996) notes that vowels are backed before retroflex coronals in the Dravidian language Koḍagu. Therefore there are words having central or back rounded vowels before retroflex consonants, as in (31a) and (31b), respectively, but there are no forms with front vowels before them, such as (31c).

(31)	a.	ɯɖi	'the whole'	b.	uɖɯ-	'to put on (sari)'	c.	*iɖ
		kɯːɭɯ	'lower, below'		kuːɭɯ	'cooked rice'		*iːɭ
		ɤɳe	'double'		oɳak-	'to dry'		*eɳ
		kɤːɖɯ	'ruin'		koːɖɤ	'monkey'		*eːɖ

Flemming cites studies showing that retroflexes can be articulated with a retracted tongue body. Unlike other coronals, they tend not to be articulated with a *fronted* tongue body, because this leads to articulatory difficulty. As Flemming points out, unified feature theory predicts that even retroflex coronals should cause vowel fronting, if fronting is [coronal] spreading. But this does not occur.

Third, because it posits that all C–V interactions are within-category, by means of vowel place features, inherent vowel place theory does not require any adjustment of our understanding of tiers or interaction. Vowels do not block consonantal place dissimilations, because they do not have consonant place features. Consonants block vowel dissimilations or assimilations only when they *do* have vowel place specifications.

What about the "weakness" of "cross-category" effects? As we noted at the end of §2, they are weaker than "within-category" effects by a range of diagnostics. Why are they comparatively infrequent? Why are they confined to roughly segmental adjacency? Why do "within-category" effects always win out over "cross-category" ones when both are in theory possible (*/p^ji/ → [p^ju])? Why do "cross-category" effects often seem to target only "underspecified" vowel types? Why do they sometimes need to "gang up" to cause effects?

We might attribute these signs of weakness to the intrinsic weakness of inherent vowel place features. As discussed at the outset of this section, the effects that "plain" consonants have on neighboring vowels are rather brief and slight. The hypothesis here has been that such effects can play a direct role in the phonology. However, feature theory, at least as traditionally conceived, provides no means of encoding this hypothesized difference between, e.g. "strong" (contrastive) and "weak" (inherent, redundant) [+round]. Until this idea is fleshed out, it is only a promissory note of the inherent vowel place approach.

Another weakness of the inherent vowel place approach to C–V interactions is that it has no immediate explanation for the natural classes of consonants and vowels evidenced by vowel strengthenings, as in Porteño Spanish /jelo/ (or /ielo/) → [ʒelo] 'ice', and consonant lenitions, as in Irish /b m/ → [w] (see §4). For processes like these, the unified feature approach has a clear advantage.

7 Conclusion

Consonant–vowel interactions are a rich source of data for phonological theory. This chapter has focused on the ways in which they have influenced the theory of feature make-up and structure, and for reasons of space it has approached even this circumscribed area selectively. Though much of the field has shifted its focus away from these representational questions in recent years, with a concomitant rise in the focus on questions of constraint interaction, this shift has not tended to shed new light on the questions raised in this chapter. It is to be hoped that new trends in the field will eventually allow us to address these questions in a newly productive way.

REFERENCES

Alderete, John. 1997. Dissimilation as local conjunction. *Papers from the Annual Meeting of the North East Linguistic Society* 27. 17–32.

Anderson, John M. & Colin J. Ewen. 1987. *Principles of dependency phonology*. Cambridge: Cambridge University Press.

Anttila, Arto. 2002. Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20. 1–42.

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Bird, Steven & Ewan Klein. 1990. Phonological events. *Journal of Linguistics* 26. 33–56.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Brame, Michael K. 1972. On the abstractness of phonology: Maltese ʕ. In Michael K. Brame (ed.) *Contributions to generative phonology*, 22–61. Austin: University of Texas Press.

Bright, William. 1972. The enunciative vowel. *International Journal of Dravidian Linguistics* 1. 26–55.

Broselow, Ellen & Alice Niyondagara. 1989. Feature geometry and Kirundi palatalization. *Studies in the Linguistic Sciences* 20. 71–88.

Browman, Catherine P. & Louis Goldstein. 1992. "Targetless" schwa: An articulatory analysis. In Gerard J. Docherty & D. Robert Ladd (eds.) *Papers in laboratory phonology II: Gesture, segment, prosody*, 26–56. Cambridge: Cambridge University Press.

Buckley, Eugene. 2000. On the naturalness of unnatural rules. *Proceedings from the 2nd Workshop on American Indigenous Languages*, 1–14. Santa Barbara: Department of Linguistics, University of California, Santa Barbara.

Campbell, Lyle. 1974. Phonological features: Problems and proposals. *Language* 50. 52–65.

Cheng, Lisa L. 1991. Feature geometry of vowels and co-occurrence restrictions in Cantonese. *Proceedings of the West Coast Conference on Formal Linguistics* 9. 107–124.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1976. Palatalization: Linking or assimilation? *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 12. 96–109.

Clements, G. N. 1991. Place of articulation in consonants and vowels: A unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5. 77–123.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 245–306. Cambridge, MA & Oxford: Blackwell.

Clements, G. N. & Engin Sezer. 1982. Vowel and consonant disharmony in Turkish. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part II, 213–255. Dordrecht: Foris.

Colarusso, John. 1992. *A grammar of the Kabardian language*. Calgary: University of Calgary Press.

Coleman, John & John Local. 1991. The "No Crossing Constraint" in autosegmental phonology. *Linguistics and Philosophy* 14. 295–338.

Cowell, Mark W. 1964. *A reference grammar of Syrian Arabic (based on the dialect of Damascus)*. Washington, DC: Georgetown University Press.

Ebert, Karen. 1996. *Koḍava*. Munich: Lincom Europa.

Elorrieta, Jabier. 1991. The feature specification of uvulars. *Proceedings of the West Coast Conference on Formal Linguistics* 10. 139–149.

Emeneau, M. B. 1970. Koḍagu vowels. *Journal of the American Oriental Society* 90. 145–158.

Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles. Published 2002, London & New York: Routledge.

Flemming, Edward. 2003. The relationship between coronal place and vowel backness. *Phonology* 20. 335–373.

Fortescue, Michael. 1984. *West Greenlandic*. London: Croom Helm.

Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22. 179–228.

Gudschinsky, Sarah, Harold Popovich & Frances Popovich. 1970. Native reaction and phonetic similarity in Maxakalí phonology. *Language* 46. 77–88.

Halle, Morris. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26. 1–46.

Hammond, Michael. 1988. On deriving the well-formedness condition. *Linguistic Inquiry* 19. 319–325.

Harris, James W. & Ellen M. Kaisse. 1999. Palatal vowels, glides and obstruents in Argentinian Spanish. *Phonology* 16. 117–190.

Herzallah, Rukayyah. 1990. Aspects of Palestinian Arabic phonology: A non-linear approach. Ph.D. dissertation, Cornell University.

Hulst, Harry van der. 1989. Atoms of segmental structure: Components, gestures and dependency. *Phonology* 6. 253–284.

Hume, Elizabeth. 1990. Front vowels, palatal consonants and the rule of umlaut in Korean. *Papers from the Annual Meeting of the North East Linguistic Society* 20. 230–243.

Hume, Elizabeth. 1994. *Front vowels, coronal consonants and their interaction in nonlinear phonology*. New York: Garland.

Hume, Elizabeth. 1996. Coronal consonant, front vowel parallels in Maltese. *Natural Language and Linguistic Theory* 14. 163–203.

Hyman, Larry M. 1970. How concrete is phonology? *Language* 46. 58–76.

Hyman, Larry M. 1972. *A phonological study of Feʔ-feʔ-Bamileke. Studies in African Linguistics*. Supplement 4.

Hyman, Larry M. 1973. The feature [grave] in phonological theory. *Journal of Phonetics* 1. 329–337.

Ihiunu, Peter & Michael Kenstowicz. 1994. Two notes on Igbo vowels. Unpublished ms., MIT.

Itô, Junko. 1984. Melodic dissimilation in Ainu. *Linguistic Inquiry* 15. 505–513.

Itô, Junko & Armin Mester. 1996. Rendaku 1: Constraint conjunction and the OCP. Paper presented at the Kobe Phonology Forum.

Johnson, Keith. 2003. *Acoustic and auditory phonetics*. 2nd edn. Malden, MA: Blackwell.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1985. The internal structure of phonological elements: A theory of charm and government. *Phonology Yearbook* 2. 305–328.

Keating, Patricia. 1988. Palatals as complex segments: X-ray evidence. *UCLA Working Papers in Phonetics* 69. 77–91.

Keating, Patricia. 1991. Coronal places of articulation. In Paradis & Prunet (1991), 29–48.

Keating, Patricia & Aditi Lahiri. 1993. Fronted velars, palatalized velars, and palatals. *Phonetica* 50. 73–101.

Lahiri, Aditi & Vincent Evers. 1991. Palatalization and coronality. In Paradis & Prunet (1991), 79–100.

Lees, Robert B. 1961. *The phonology of Modern Standard Turkish.* Bloomington: Indiana University.

Lightner, Theodore M. 1972. *Problems in the theory of phonology: Russian phonology and Turkish phonology.* Edmonton: Linguistic Research Inc.

Lombardi, Linda. 2003. Markedness and the typology of epenthetic vowels. Unpublished ms., University of Maryland at College Park (ROA-578).

Maddieson, Ian. 1984. *Patterns of sounds.* Cambridge: Cambridge University Press.

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.

McCarthy, John J. 1986. Features and tiers: The structure of Semitic roots. Paper presented at Brandeis University.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.) *Phonological structure and phonetic form: Papers in laboratory phonology III,* 191–233. Cambridge: Cambridge University Press.

Mester, Armin. 1986. Studies in tier structure. Ph.D. dissertation, University of Massachusetts, Amherst.

Mester, Armin & Junko Itô. 1989. Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language* 65. 258–293.

Ní Chiosáin, Máire. 1991. Topics in the phonology of Irish. Ph.D. dissertation, University of Massachusetts, Amherst.

Ní Chiosáin, Máire. 1994. Irish palatalisation and the representation of place features. *Phonology* 11. 89–106.

Ní Chiosáin, Máire & Jaye Padgett. 1993. Inherent V-Place. Report LRC-93-09, Linguistics Research Center, University of California, Santa Cruz.

Ní Chiosáin, Máire & Jaye Padgett. 2001. Markedness, segment realization, and locality in spreading. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations,* 118–156. Cambridge: Cambridge University Press.

Odden, David. 1991. Vowel geometry. *Phonology* 8. 261–289.

Odden, David. 1994. Adjacency parameters in phonology. *Language* 70. 289–330.

Padgett, Jaye. Forthcoming. Russian consonant–vowel interactions and derivational opacity. In Wayles Brown, Adam Cooper, Alison Fisher, Esra Kesici, Nicola Predolac & Draga Zec (eds.) *Proceedings of the 18th Formal Approaches to Slavic Linguistics meeting.* Ann Arbor: Michigan Slavic Publications.

Paradis, Carole & Jean-François Prunet (eds.) 1991. *The special status of coronals: Internal and external evidence.* San Diego: Academic Press.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Pulleyblank, Douglas. 1988. Vocalic underspecification in Yoruba. *Linguistic Inquiry* 19. 233–270.

Pulleyblank, Edwin G. 1989. The role of coronal in articulator based features. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 25. 379–393.

Rischel, Jørgen. 1974. *Topics in West Greenlandic phonology: Regularities underlying the phonetic appearance of wordforms in a polysynthetic language.* Copenhagen: Akademisk Forlag.

Rose, Sharon. 1996. Variable laryngeals and vowel lowering. *Phonology* 13. 73–117.

Rubach, Jerzy. 1993. *The Lexical Phonology of Slovak.* Oxford: Clarendon Press.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.
Sagey, Elizabeth. 1988. On the ill-formedness of crossing association lines. *Linguistic Inquiry* 19. 109–118.
Schane, Sanford A. 1984. The fundamentals of particle phonology. *Phonology Yearbook* 1. 129–155.
Schane, Sanford A. 1987. The resolution of hiatus. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 279–290.
Schultz-Lorentzen, C. W. 1945. *A grammar of the West Greenland language*. Copenhagen: C. A. Reitzel.
Scobbie, James M. 1991. Attribute value phonology. Ph.D. dissertation, University of Edinburgh.
Selkirk, Elisabeth. 1988. Dependency, place and the notion "tier." Paper presented at the 63rd Annual Meeting of the Linguistic Society of America, New Orleans.
Selkirk, Elisabeth. 1993. [Labial] relations. Unpublished ms., University of Massachusetts, Amherst.
Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 219–250. San Diego: Academic Press.
Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.
Upadhyaya, Suseela P. 1968. Mapila Malayalam (a descriptive and comparative study). Ph.D. dissertation, University of Poona.
Yip, Moira. 1988. The Obligatory Contour Principle and phonological rules: A loss of identity. *Linguistic Inquiry* 19. 65–100.
Yip, Moira. 1989. Feature geometry and cooccurrence restrictions. *Phonology* 6. 349–374.

88 Derived Environment Effects

Luigi Burzio

1 Introduction

Derived environments often exhibit peculiar phonological properties. Notable effects can be identified relative to various senses of "derived" in the expression "derived environment" (DE). In all cases, "environment" refers to some phonological context. Such environment or context can be "derived" in a phonological sense, by virtue of some phonological process having applied to obtain it or, in a morphological sense, by virtue of it being the result of the combination of morphemes or other morphological operation. The following sections will first review the different cases from a descriptive point of view, and then turn to their respective theoretical accounts. §2 and §3 review the basic facts, arguing that there are three subcases overall. §4 turns to contemporary analyses of the two better-known subcases, while §5 reviews pre-Optimality Theory accounts of the same. §6 presents an account of the third subcase.

2 Phonologically derived environments

Environments that are derived by some phonological process can differ from underived environments relative to further phonological processes. An English example of this effect is provided by the following contrasting pair.[1]

(1)

	verb	-able *adjective*
a.	*'remedy*	*re'meːdiable*
b.	*'levy*	*'leviable*

In (1) we assume that in both cases the calculation of the adjective proceeds from the surface form of the verb concatenated with the suffix *-able*. Such calculation

[1] The symbol "ː" is used here and elsewhere in orthographic forms to refer to a long vowel. Stress is also marked in such forms. In (1) the vowel is diphthongized in accordance with the Early Modern English Great Vowel Shift, here the diphthong [ij].

would find a stage in which the position of the stress is changed in (1a) (*'remedy* → *re'medi* . . .), but not in (1b). This phonologically *derived* status of (1a) correlates with the occurrence of vowel lengthening in the adjective in (1a), compared with its absence in (1b). It is clear that it is the restressing that licenses vowel lengthening and not the opposite, because the restressing is independently predictable. Specifically, the stress of both adjectives in (1) simply conforms with the English norm, by which the rightmost stress can be at most antepenultimate, but a syllable like *ble* (= [bḷ], with syllabic *l*) can evade syllable count (Burzio 1994), hence *re.'me.di.a<ble>*, *'le.vi.a<ble>*. The lengthening of (1a) reflects a process of English phonology, which is relatively regular aside from its non-occurrence in cases like (1b), affecting vowels in the context / __ CiV (cf. *Bost*[ow]*nian*, *Can*[ej]*dian*, etc.) except for the vowel *i*, which is immune, e.g. *Palestinian* (see Chomsky and Halle 1968: 47; Halle and Mohanan 1985). Non-restressing cases like (1b) confirming the generalization are *'bury/'burial*, not **b*[ij]*rial*; *Ma'lawi/ Ma'lawian*, not **Mal*[ej]*wian*; *Ken'tucky/Ken'tuckian*, not **Kent*[uw]*ckian*, although exceptions – which I will put aside – exist both ways (Wolf 2008: 302), e.g. restressing but short *Trini'dadian*, *I'talian*, *Chau'cerian*, *He'gelian*, and non-restressing but long *Alab*[ej]*mian*, *Bah*[ej]*mian*.

In addition, both adjectives in (1) constitute environments that are *derived* morphologically, by way of affixation. The latter is evidently not relevant to the phenomenon at hand, else no difference between (1a) and (1b) would be expected. Finally, note as well that characterization of the difference between (1a) and (1b) would not be possible unless it was indeed the *surface* form of the verb that enters into the calculation of the adjective. Direct calculations from bare-bone underlying representations (CHAPTER 1: UNDERLYING REPRESENTATIONS) containing no stress information would predict no difference, since there would then be no *restressing* in either case, just regular assignment of stress. In sum, assuming the surface forms of the verbs are relevant in (1), then the adjective in (1a) would undergo phonological restressing unlike the one in (1b), and thus the latter would represent a case of phonologically *non-derived environment blocking* (henceforth NDEB) relative to the process of CiV lengthening.

The literature documents several other cases of this general type across a significant spectrum of languages. Łubowicz (1999, 2002) reports the cases in (2a) to (2d) below, to which we may add the Finnish case in (2e). The original sources for these data are Rubach (1984) for Polish; Kenstowicz and Rubach (1987) and Rubach (1993, 1995) for Slovak; Bolognesi (1998) for Campidanian Sardinian; Prince (1975) for Tiberian Hebrew; and Kiparsky (1973a, 1993) for Finnish.

(2)		*Processes*	*General process/ DE-only process*
a.	*Polish*		
	i. *kro*[k] / *kro*[ʧ]-*ek* 'step / little step'	k → ʧ	Velar palatalization
	ii. *dron*[g] / *dron*[ʒ]-*ek* 'pole / little pole'	g → ʤ → ʒ	Spirantization of voiced palatal affricates
	iii. *bri*[ʤ] / *bri*[ʤ]-*ek* 'bridge / little bridge'	ʤ → *ʒ (underived)	

b. *Slovak*

i.	*lop*[a]*t-a* / *lop*[a:]*t* 'shovel-NOM SG / GEN PL'	a → a:	Vowel lengthening
ii.	*kazet-a* / *kaz*[ie]*t* 'box-NOM SG / GEN PL'	e → e: → ie	Diphthongization of long [e:]
iii.	*dc*[e:]*ra* 'daughter'	e: → *ie (underived)	

c. *Campidanian Sardinian*

i.	[f]*amilia* / *sa* [v]*amilia* 'family / the family'	f → v	Post-vocalic voicing of obstruents
ii.	*pisci* / *belu* [β]*isci* 'fish / nice fish'	p → b → β	Post-vocalic spirantization
iii.	[b]*ia* / *sa* [b]*ia* 'road / the road'	b → *β (underived)	

d. *Tiberian Hebrew*

i.	ktab'tem / ka:'tab 'we-MASC write' / 'he writes'	a → a:	Pre-tonic, open syllable lengthening
ii.	ʃimi'ka: / ʃe:'m-ot 'your names / names'	i → i: → e:	High long V lowering
iii.	qi:'to:r 'smoke'	i: → *e: (underived)	

e. *Finnish assibilation*

i.	*joke-nä* / *joki* 'river-ESS / NOM'	e → i	*e*-raising, word-finally
ii.	*vete-nä* / *vesi* 'water-ESS / NOM'	te → ti → si	*t*-assibilation before *i*
iii.	*äiti-nä* / *äiti* 'mother-ESS / NOM'	ti → *si (underived)	

In each of the cases in (2), row (i) documents the existence of a process that occurs independently of whether or not the environment is phonologically derived, analogously to the restressing in (1a) above. Then, row (ii) documents the existence of a second process that occurs when the first process has applied, analogously to the CiV lengthening of (1a), while row (iii) further documents the fact that the second process does *not* apply unless the first one has (NDEB), analogously to the failed CiV lengthening of (1b) above. Hence, in Polish, voiced palatal affricates spirantize to fricatives only when they are derived from velar stops before front vowels (CHAPTER 121: SLAVIC PALATALIZATION); in Slovak, long /e:/ diphthongizes to [ie] only if derived via a lengthening process (induced by specific affixes); in the Sardinian case, voiced stops spirantize only when they are derived via post-vocalic voicing; in Tiberian Hebrew, high long vowels lower only when they are derived via a lengthening process; and in Finnish, *t* assibilates to *s* before *i*, but only when the latter is derived from *e*.

In sum, certain phonological changes appear to occur only in conjunction with other specific phonological changes, and not by themselves (NDEB).

3 Morphologically derived environments

Phonological processes may be conditioned not only by other phonological process, but by morphological processes or structure as well. Two subcases need to be distinguished, and are reviewed in turn below. In one subcase, the context of application of the phonological process spans across morphemes, as in *criti*[s]-*ism*, where the velar of *criti*[k] softens before the *i*, across a morpheme boundary. In the other subcase, the phonological process occurs in environments that are morphologically derived, but without reference to the specifics of the morphological structure. For instance, the nouns or adjectives *altern*[ə]*te*, *moder*[ə]*te*, *design*[ə]*te* are presumably derived from the corresponding verbs in *-ate*, but without any overt morphology. Hence the vowel shortening in these cases must depend purely on derived status, without reference to any particular morphological material or boundaries.

3.1 *Boundary contexts*

Kiparsky (1973a, 1993) notes that, in addition to the case in (2) above, Finnish assibilation also displays the pattern in (3).

(3)	a.	*halut-a*	'want-INF'	*halus-i*	'want-PAST'
	b.	*tilat-a*	'order-INF'	*tilas-i*	'order-PAST'
	c.	*äiti*	'mother'		

Here, assibilation turns *t* to *s* before *i*, but only when the latter belongs to a different morpheme, as in each of (3a) and (3b). In particular, the failed assibilation in *ti* of *tilas-i* (3b) shows that being in a morphologically derived form is not sufficient, and that it is necessary for the assibilation environment itself to be created morphologically. The non-assibilating form in (3c) establishes that assibilation does not just single out final syllables, confirming the relevance of the derived environment.

The already noted English *velar softening* (CHAPTER 71: PALATALIZATION), illustrated in (4), appears to be similarly restricted to morphologically derived environments.

(4)	a.	Derived	/k/ → [s]	*critic / critic-ism; electric / electric-ity; opaque / opac-ity*
	b.	Underived	[k]	*kinetic, kidney, kitchen, Viking, . . .*

I note here the item *kinematic-ity*, which constitutes a proper counterpart to Finnish *tilas-i* in (2e), velar softening affecting only the [k] adjacent to the morpheme boundary, and not the initial one. A complicating factor in this classification is that velar softening has exceptions, like *monarch-ist, anarch-ist*, and several others, raising the possibility that items like (4b) may in fact also just be lexical exceptions, rather than being indicative of NDEB. Hence in this case, as in others below, classification will be dependent on the choice of analyses. I will, nonetheless, assume that items like those in (4b) are *not* lexical exceptions, and therefore that the classification of velar softening as operating only across morpheme boundaries like Finnish assibilation is correct. In addition, and with the same caveats, I will assume that each of the phenomena in (5) below also instantiates that same type of NDEB, while the references cited in each case may be consulted for specific analyses.

(5) *Further cases of NDEB, for DEs including morpheme boundaries*

a. *Korean palatalization* (Kiparsky 1993; Iverson and Wheeler 1988)
/hæ tot-i/ → [hæ doj-i] '(sun)rise-NOM' /mati/ 'knot'

b. *Polish velar palatalization* (Łubowicz 2002; Rubach 1984)
/krok-i-ć/ → [krotʃ-i-ć] 'to step' /k'iśel/ 'jelly'

c. *Polish dental palatalization* (Kenstowicz 1994; Rubach 1984)
/serwis-e/ → [serwiś-e] 'auto service-LOC' /serwis/ (NOM)

d. *Pre-coronal laminalization in Chumash* (Kiparsky 1993; Poser 1982, 1993)
/s-tepuʔ/ → [ʃ-tepuʔ] 'he gambles' /stumukun/ 'mistletoe'

e. *Sanskrit (ruki rule) retroflection after r, u, k, i* (Kiparsky 1973a, 1993)
/agni-su/ → [agni-ṣu] 'fire-DAT PL' /kisalaja/ 'sprout'

f. *Indonesian nasal substitution* (Pater 1999)
/məN-pilih/ → [məm-ilih] 'to choose' /əmpat/ 'four'

g. *Finnish cluster assimilation* (Kiparsky 1973a)
/pur-nut/ → [purrut] 'bitten' /horna/ 'hell'

h. *Mohawk epenthesis* (Kiparsky 1973a)
/k-wi'stos/ → [kewi'stos] 'I am cold' /rú:kweh/ 'man'

In each of the cases in (5), the morphologically derived environment affected by the change is compared with an otherwise identical but non-derived environment in which the change fails to occur (see portions in boldface). In all cases, the morphologically derived environment includes a morpheme boundary.

3.2 *Non-boundary contexts*

Turning now to cases where morphologically derived status appears to make a difference without implicating material contributed by the morphology, English vowel shortening, e.g. as in *di'vin-ity vs.* underived *'ivory*, will serve as the prototype, although its exact analysis, given below, will be critical to this role.

Burzio (1993, 1994, 2000a) argued that, while tradition had focused on individual shortening processes, like the "trisyllabic" shortening of *di'vin-ity* or of (6a) below, the actual generalization is in fact found over non-shortening contexts or the "exceptions," while the shortening is otherwise fully general. Given the variety of vowel length changes illustrated by the left-hand cases in (6), including not only shortening but also lengthening, as in (6h), separate characterization of each case would result in a colossal conspiracy. By contrast, such distribution can receive a unitary analysis in terms of vowel length being allophonic in this sector of the lexicon, rather than contrastive, as in the rest of the language. On this view, long vowels would be disallowed in general and produced only under specific contextual demands. The factor responsible for long vowels appears to be stress, and specifically the stress that would be inherited from the respective morphological bases given in (6) in parentheses. A review of each case in (6) will drive this point home.

(6) *Vowel length in the English Latinate-derived lexicon*

		stress preserved		
a.	*'**natur**-al* (*'na:ture*)	yes		
b.	*o'**blig**-aˌtory* (*ob'li:ge*)	yes		
c.	*ˌdefa'**m**-a:tion* (*de'fa:me*)	no		
d.	*ar'ticulaˌ**t**-ory* (*ar'ticuˌla:te*)	no		
e.	*'alternate*$_{N/A}$ (*'alterˌna:te*$_V$)	no		
f.	*'**aspir**-ant* (*as'pi:re*)	no	yes	*de'**si:r**-ous* (*de'si:re*)
g.	*'**generat**-ive* (*'geneˌra:te*)	no	yes	*'legisˌ**la:t**-ive* (*'legisˌla:te*)
h.	*Eˌliza'**be:th**-an* (*E'lizabeth*)	yes	no	*Her'**culean*** (*'Hercules*)

The cases in (6a) and (6b) preserve the stress of their bases, and the short vowel in each case will follow from the fact that such stress patterns are independently known *not to* require a long vowel. The pattern of *'natural* is the same as that of *A'merica*, with stress on an antepenultimate light syllable (short vowel), while that of *o('bliga)ˌtory* is relevantly like that of *(ˌAri)'zona*, the parentheses here marking a binary foot in non-word-final position, evidently also not needing a long vowel/heavy syllable to bear stress. Like the cases in (6a) and (6b), those in (6c)–(6e) also do not need long vowels, but for different reasons. Here, a stress on the vowel in boldface *cannot* be preserved from the base. Given the stress of affixes *-'a:tion*, *-ˌory* (American English), which apparently must take priority, stress preservation is excluded in (6c) and (6d) by the ban on stress clashes. Similarly, given the independent fact that nouns tend to avoid the final stress of verbs (cf. *per'vert*$_V$ / *'pervert*$_N$; *per'mit*$_V$ / *'permit*$_N$), the stress of the base verb is presumably excluded in (6e). Hence, on the "allophonic" hypothesis, vowels are short in each of (6c)–(6e), because nothing motivates the long allophones (CHAPTER 11: THE PHONEME).

Matters are different in the derived forms in (6f)–(6h), where all boldface vowels are in penultimate syllables. Like other languages, English is well known to stress penultimates only if heavy, and antepenultimates otherwise (*ˌA.ri.'zo:.na*, *a.'gen.da* vs. *A.'me.ri.ca*, *'as.te.risk*), if we put aside verbs, which require a separate discussion (Chomsky and Halle 1968; CHAPTER 102: CATEGORY-SPECIFIC EFFECTS). The derived forms in (6f)–(6h) all conform with this generalization. The left/right variation in each case, however, reveals that the general mandate for vowels to be short competes evenly with preservation of stress, some outcomes favoring the former, and others the latter – a case of lexically controlled phonological variation (Burzio 2006). Lengthening as in *Eˌliza'be:than* in (6h) is on this analysis just like the non-shortening of *de'si:rous* in (6f) and other cases, since it allows preservation of the stress of *E'lizabeth*, albeit as a secondary. The alternative **'Eli'zabethan* with a short *e* would lose that stress altogether, given again the ban on stress clashes, as in fact happens in *Her'culean* of (6h) (though the equally expected *ˌHercu'le:an* is also attested). The point at issue is that long vowels occur in the Latinate lexicon *only* to preserve stress, though variably, as in (6f)–(6h), and are banned otherwise, ignoring occasional cases like *o'be:sity*, which are exceptions to the pattern in (6a), but are an effect of different granularity from the variation in (6f)–(6h), which is very robust (giving roughly a 50–50 split; Burzio 1993, 2000a).

The conclusion holds that in this morphologically derived domain vowels are required to be short, except, and variably, under stress demands. It is therefore this general requirement, statable as *V:, that "blocks" in environments that are not so derived, e.g. *'di:nosaur*, *'i:vory*, etc.

The cases listed below would appear to be of the same a-contextual type as argued for the English cases in (6), while, again, the specific analyses may be ultimately critical to the correct classification.

(7) *Further cases of NDEB, for DEs not including morpheme boundaries*

a. *Italian participles* (Burzio 1998)
as.cen.dere 'ascend' *as.ce.so* 'ascended' (less marked syllable)

b. *Turkish disyllabicity condition* (Inkelas and Orgun 1995: 770)
ham 'unripe' **fa-n* '(note) fa-2SG POSS'
fa-dan 'fa (note)-ABL'
(avoidance of marked prosodic structure)

c. *Japanese two-mora requirement* (Itô 1990; Kiparsky 1993)
su 'vinegar' *choko* 'chocolate (truncation)'
**cho* (two-mora requirement satisfied)

d. *Catalan stressed vowel lowering* (Mascaró 1976; Kiparsky 1993: 293)
'sentrə 'center' 'sɛntr-ik 'centric'
direk'to 'director' direk'tɔr-i 'directory' (preferred segment)

e. *Catalan unstressed mid-vowel reduction* (Mascaró 1976; Kiparsky 1993: 294)
bo'ston 'Boston' bustun-'ja 'Bostonian'
'katedrə 'academic chair' katə'drat-ik 'holder of academic chair'
(preferred segment)

f. *French* h-aspiré (Kiparsky 1993: 294)
Hitler 'Hitler' *(h)itlérien* 'hitlerian'
(loss of marked segment)

As in the English case, where the presumably more marked long vowels are avoided, each of the derived cases in (7) would appear to avoid some choice that is relatively more marked either prosodically, as in (7a)–(7c), or from the point of view of segmental inventories, as in (7d)–(7f), as indicated in parentheses. As suggested above, alternative analyses of some of these cases are conceivable. In particular, the cases in (7b) and (7c) are amenable to an account in terms of prosodic requirements imposed by the morphology (Downing 2006). Also, the specific analysis of cases like (7b) will depend on the general solution to the issue of "ineffability" (absence of an output for certain inputs; see van Oostendorp 2009). The lower vowels in (7d) are argued by Wolf (2008: 264f.) to be a general property of oxytonic and paroxytonic stress (though cases like ['sentrə] would require further comment) rather than derived environments. In addition, Wolf (2008: 267f.) attributes cases like (7e) and (7f) to a mechanism under which "roots can lose exceptional properties in morphologically derived contexts." Technical details aside, the latter view is effectively equivalent to the present one, however. In sum, at least some of the cases in (7) are likely to be genuine instances of the proposed class, while others may be more questionable. See also Wolf (2008: 246–248) for cases of NDEB additional to those listed here and in the previous section.

4 Proposed analyses

I now turn to concrete analyses of the different types of NDEB, beginning with contemporary attempts, and turning later to more traditional ones. Here, I will focus in particular on the analyses of Łubowicz (1999, 2002), Burzio (2000a, 2000b, 2002b), and Wolf (2008), all cast within the general framework of Optimality Theory (OT; Prince and Smolensky 1993). However, see Kula (2008) for an account based on Dependency Phonology. A more comprehensive review of the literature on NDEB is provided in Wolf (2008: §4.3).

4.1 Phonologically derived environments

4.1.1 The entailments-based approach

Burzio (2000b, 2002a, 2002b, 2005) argues that a pervasive property of phonological and morphological representations is that they constitute attractors to their neighbors. This attraction effect is achieved formally by means of the hypothesis in (8).

(8) *Representational Entailments Hypothesis* (REH)

Mental representations of linguistic expressions are sets of entailments. For example, a representation consisting of A and B corresponds to the entailments: A ⇒ B, B ⇒ A (if A then B; if B then A).

The "entailments" in (8) are types of violable constraints as in OT, but, unlike OT's "strict" ranking of constraints, underlying the REH in (8) is the assumption that entailments of the same type, like "$x \Rightarrow A$," for any x, directly undergo summation, producing entailments of higher strength or rank. This is analogous to the workings of neural networks, in which activation on a node A is the sum of the activation received from all nodes x to which A is connected.

The workings of attraction can be illustrated with the case of "parasitic" assimilation in Sudanese Arabic in (9), from Kenstowicz (1994: 54).

(9)	a.	*kitaa*[b]	'book'	*kitaa*[f] *Fáthi*	'Fáthi's book'
	b.			*kitaa*[p] *Samíir*	'Samíir's book'

In (9), assimilation in continuancy (/b/ → [f]) is shown to occur only contingently, or "parasitically," on agreement in place (same labial place (9a); for the relation between continuancy and place, see CHAPTER 13: THE STRICTURE FEATURES). Otherwise only voicing assimilation occurs, as in both (9a) and (9b). The initial C of the possessor noun ([f] or [s]) thus appears to act as an attractor on the final C of the head noun [b]. If the latter is close enough to the attractor (same place), it is then drawn closer (same continuancy). Evidently, even the weaker attraction (different place) must be sufficient for voicing assimilation, which we can put aside.

Such attraction effect would follow from the entailments of (8) in the way illustrated in (10).

(10)	A B C D	A ⇒ D	B ⇒ D	C ⇒ D	*Components in* (9)
a.	A B C –D	*	*	*	A: [α consonantal]
b.	A B –C –D	*	*		B: [β sonorant]
c.	A –B –C –D	*			C: [γ place]
d.	–A –B –C –D				D: [δ continuant]

Consider that an attractor consisting of the four components A, B, C, D will generate, relative to its fourth component, D, the three entailments given in the top row in (10). Line (a) then shows how a target representation differing only by that fourth component (–D standing for "minus D" or "not D") would violate all three of those entailments, while subsequent lines show how progressively more distant representations violate fewer and fewer of those entailments. In general, the REH (8) predicts that the number of entailments violated by each component on which a target differs from an attractor will equal the number of components on which target and attractor agree. Overall violation per differing component will therefore be proportional to overall similarity, hence the notion of "attraction." Applicability to (9) above is by way of conversion of the abstract components A, B, C, D into appropriate distinctive features, as shown on the right-hand side in (10), the cases in (9a) and (9b) corresponding in particular to lines (10a) and (10b), respectively. An OT-style account of the contrast in (9) incorporating attraction can be given in terms of the schema "CC-Ident^3[cont] >> IO-Ident[cont] >> CC-Ident^2[cont]," where CC-Ident[cont] refers to the assimilatory pressure in continuancy, and the superscripts give a measure of the pressure in terms of entailments (attraction), translated into OT ranking. Note that no commitment is made here as to the relative strength of individual entailments (only that summation of like entailments will always yield greater strength), and hence no prediction is made that agreement on some other feature (e.g. voicing) must have the same triggering effect as agreement in place in (9), or that any assimilation will necessarily bring about further assimilation by a domino effect.

The case in (9) is by no means unique. It is in fact virtually duplicated by English nasal place assimilation, where the failed assimilation of *in-famous* calls for the assumption that identity in place automatically induces identity in continuancy, just as in (9). This assumption is needed to correctly exclude **im-famous* (same place but different continuancy; cf. *im-possible*), while the alternative **i*[ɱ̃]-*famous* will be excluded by the ban on nasal fricatives in English. Furthermore, Wayment (2009) argues, on a broad empirical basis, that *all* assimilatory phenomena are parasitic in the manner of (9), and thus all involve attraction, as just described. However, sequential proximity also appears to contribute to overall similarity and hence to attraction, so that, when proximity is maximal as in (9), the prerequisite similarity in features may be slight and not immediately detectable, while long-distance phenomena, like long-distance consonant assimilation (Hansson 2001; Rose and Walker 2004; chapter 77: long-distance assimilation of consonants) and vowel harmony (van der Hulst and van de Weijer 1995; chapter 91: vowel harmony: opaque and transparent vowels; chapter 110: metaphony in romance; chapter 118: turkish vowel harmony; chapter 123: hungarian vowel harmony), will predictably exhibit more robust feature-based similarity as a triggering condition, as the references just cited confirm.

Beside assimilations, Burzio argues that attraction underlies further morphophonological phenomena: in particular, that it can reconstruct the "dispersion" account of segmental inventories (Liljencrants and Lindblom 1972; Flemming 1995; chapter 2: contrast), as maximal distance among members of the inventory corresponds to minimal attraction/entailment violation. In addition, under the reduction of contextual neutralization effects to dispersion principles advocated by Steriade (2009) and Flemming (2008), attraction would also speak to those effects in turn. In particular, segmental neutralizations (like coda devoicing) would occur in those environments that attenuate critical perceptual cues, as argued by Steriade (1994, 1999, 2009), and thus compromise distance from the nearest attractor in the inventory. The latter attractor will then exert its influence, neutralizing the contrast. Attraction thus represents a formal alternative to Steriade's (2009) "perceptual map" and, as argued in Burzio (2000b), is a variant interpretation of Wilson's (2000, 2001) "targeted constraints," which are also a formal alternative to Steriade's perceptual map. Attraction has also been argued to underlie morphological syncretisms (also a type of neutralization; see Burzio 2005, 2007; Burzio and Tantalou 2007).

From this point of view, the NDEB of (1) would then be naturally interpreted as another attraction effect, parallel, in fact, to the one in (9). While the relevant relation in (9) is between two segments in the same sequence, the one in (1) is between the adjective's stem and the corresponding verb. Just as in (9), identity in one respect, here stress, results in identity in another, here vowel length, as in (1b). The REH and attraction have in fact been argued to subsume the general OT notion of "faithfulness" (CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS), and at the same time also to derive the set of relations to which faithfulness has been shown to apply (input–output; base–derivative; base–reduplicant; paradigms; similarity of consonants). Similarity can define all such relations, with sequential adjacency/proximity also contributing to similarity, as noted (Wayment 2009). However, morphology also plays a role in further defining faithfulness (§6).

In OT notation, the attraction accounts of (9) and (1) may then be rendered as in the parallel (11a) and (11b), respectively.

(11) a. CC-Ident^{+}[cont] >> IO-Ident[cont] >> CC-Ident[cont]
b. Ident^{+}(V-length) >> *V̆CiV >> Ident(V-length)

The schema in (11a) is the same as the one given above, except that superscript "3, 2" have been replaced with just "+" and no superscript, respectively. In (11b), Ident(V-length) applies between input (verb) and output (adjective) in (1). As in (11a), the "+" version corresponds to the stronger attraction, here due to identity in stress. The second constraint in (11b) promotes a long vowel before "CiV" and prevails in (1a) by competing only with the lower-ranked Ident (weaker attraction), but is ineffective in (1b), where the higher-ranked Ident is involved. The attraction account of (1) as in (11b) seems applicable in transparent ways to the other cases of this type cited in (2) and (3) above.

4.1.2 *The "local conjunction" approach*

An account of NDEB based on the formal device of local conjunction (LC) of constraints in OT (CHAPTER 62: CONSTRAINT CONJUNCTION) is proposed in Łubowicz (1999, 2002). When deployed in a case like (1), such an account would feature the hierarchy in (12), which can be compared with the one in (11b).

(12) [*V̆CiV & Ident(stress)] >> Ident(V-length) >> *V̆CiV

The conjunction in (12) combines one faithfulness constraint, Ident(stress), and one markedness constraint, *V̆CiV. Simultaneous violation of both conjuncts would thus invoke the higher-ranked conjunction, forcing the repair as in the stress-unfaithful (1a), while the markedness constraint alone would be low-ranked, hence failing to induce the repair, as in (1b). As with the account in the previous section, stress constraints must exclude un-restressed *'*remediable* (Burzio 1994).

It is easy to see that this approach also covers the cases in (2a)–(2d), as Łubowicz shows, and possibly the one in (2e) (but see below), given appropriate choices of markedness and faithfulness constraints that make up the conjunction. The LC solution seems ingenious and has interesting typological properties, predicting that not only conjunctions of markedness and faithfulness constraints, but conjunctions of two markedness constraints, as well as conjunctions of two faithfulness constraints, should prove equally useful. These predictions seem to be supported to the extent that conjunctions of markedness constraints have frequently been proposed in the OT literature (see McCarthy 2002: 18f.; Łubowicz 2002: note 4; and, for general reviews, Fukazawa 1999), while conjunctions of faithfulness constraints have also been proposed, in particular to account for counterfeeding chain shifts (Kirchner 1996; CHAPTER 73: CHAIN SHIFTS).

The LC approach cuts the empirical domain differently from the attraction-based approach. While it may relate NDEB to the other "conjunctive" effects just noted, it does not relate them to other phenomena in the domain of attraction, like the parasitic character of assimilations shown in (9). The reason a schema along the lines of (12) will be silent on (9) is that in the latter case the crucial difference, i.e. same place [. . . b#f . . .] (9a) *vs.* different place [. . . b#s . . .] (9b) is not one created by some process, but rather one already present in the input. Hence there is no difference in terms of IO-faithfulness that could be recruited for a LC like the one in (12).[2] To make the LC approach applicable to cases like (9), we might import from the attraction framework the notion that faithfulness may also hold across segments within the same input sequence, but further arguable attraction effects (like dispersion or syncretism) would remain recalcitrant. Other cases, like those in (13), from Burzio (2002b), may, like the one in (9), also involve differences without an actual process, and may thus also pose a challenge to the LC approach.

(13)	*base*	*derivative: more similar*	*derivative: less similar*	*interacting dimensions*
a.	*com'pare*	*com'parable*	*'comparable*	stress; semantics
b.	*di'vide*	*di'vidable*	*di'visible*	vowel length; segmentism
c.	*ap'ply*		*'applicable*	vowel length; segmentism
	de'ny	*de'niable*		
d.	*'larynx*	*'larynxes*	*la'rynges*	stress; segmentism

The case in (13a) features, in the "less similar" column, an idiosyncratic semantic change, as the derived adjective *'comparable* means "roughly equal" rather than

[2] The same predicament would face McCarthy's (2003) "Comparative Markedness" (CM) approach, which distinguishes markedness violations that are present in the input (old markedness) from those that are not (new markedness). As Łubowicz (2003) shows, this approach has similar effects to those of the LC approach. Conceivably, a higher-ranked "new" *V̆CiV could in particular apply to (1a) given the changed stress, but not to (1b). In (9), however, nothing is independently "new" in (a) any more than in (b), yielding no account in these terms. Further difficulties with CM are noted in Wolf (2008: §4.2.3).

"able to be compared," while the cases in (13b)–(13d) exhibit idiosyncratic segmental changes. In each case, such changes cluster with other changes, in stress, vowel length, or both. These further changes are then roughly predictable, though space prevents a full discussion here. For instance, the form *la'rynges* displays regular stress (heavy penultimate) rather than the stress of its base. By comparison, the same changes "block" in the absence of the first type of change, as shown in the "more similar" column. Such cases of NDEB are covered by attraction, which responds to difference/distance regardless of its source. By contrast, the LC approach will require an extension of the formal notion of faithfulness to the domain of suppletion and semantic idiosyncrasy.

The two approaches reviewed so far may also differ in the degree of locality that they impose on possible interactions. The LC approach enforces the locality required by the notion of "local conjunction." For instance, LC would be applicable to (1a), due to the fact that both changes, in stress and vowel length, occur in the same vowel. The theory of LC is not fully explicit, however, on what exactly counts as a local domain (Smolensky 1997), and we may note that in fact not all cases of phonological NDEB involve the same segment. For instance, the two relevant changes in the Finnish case in (2e) *vete* → *veti* → *vesi* occur in different segments. See, however, Łubowicz's (2002: note 29) appeal to an alternative analysis that would bring this case in line. An apparently similar challenge is also posed by Sanskrit retroflexion of s after the disjunction {*r, u, k, i*} (Kiparsky 1973a; Kenstowicz 1994: 202; CHAPTER 119: REDUPLICATION IN SANSKRIT).

As for the predictions of the REH on the exact range of interaction of changes, they are also not very clear at the present stage, but perhaps some inference can be drawn from the fact that the REH aims to characterize inventories as sets of attractors, combined with the independent fact that inventories are attested for segments and morphemes. If these are the attractors, then attraction effects should be observable at both of these levels. Then the cases in (2a)–(2d) and perhaps (1), would instantiate cases where attraction occurs between corresponding segments, while those in (2e) and (13) may instantiate attraction between corresponding allomorphs (note that several of the latter cases involve more significant structural changes than those in (2a)–(2d)).

4.1.3 *Serial Optimality Theory*

Wolf (2008) develops an approach to NDEB based on the serial OT of McCarthy (2007), in which different potential sequences of operations ("candidate chains") are evaluated by constraints on derivations. This approach receives its primary motivation from phonological opacity (counterfeeding and counterbleeding effects; see CHAPTER 74: RULE ORDERING). Given, for instance, an opaque derivation like that of Serbo-Croatian /okrugl/ → okrugal → [okrugao] 'round' (Kenstowicz 1994: 90f.), in which vocalization of /l/ to [o] counterbleeds the epenthesis of [a] that would break up a final cluster, this approach would postulate a "precedence" constraint PREC(DEP, IDENT[consonantal]) prescribing that a violation of IDENT[consonantal] (i.e. vocalization) must be preceded by a violation of DEP (i.e. epenthesis), whenever the latter violation would be harmony-improving (i.e. in final CC environments). The case in (1) would then be handled in this perspective via the schema in (14).

(14) PREC(IDENT(stress), IDENT(V-length)) >> *V́CiV >> IDENT(V-length)

The top-ranked constraint in (14) will demand that a violation of IDENT(V-length) (i.e. lengthening) be preceded in the derivation by a violation of IDENT(stress) (i.e. restressing), thus blocking lengthened but un-restressed **le:viable*. While using different theoretical means, this requirement is parallel to that imposed by IDENT+(V-length) in (11b) on the basis of the attraction approach (expressing the stronger attraction/IDENT requirement if stress is unchanged), while the rest of the schema in (14) remains the same as that of (11b).

Major differences lie again in the way this approach relates to phenomena other than NDEB. As noted, it links NDEB directly with phonological opacity, while this link is more indirect in the alternative approaches. In the attraction approach, it is natural to see segmental neutralizations (like the /l/–/o/neutralization of Serbo-Croatian) as attraction effects due to phonetic similarity. In turn, neutralizations (and perhaps other processes) have been shown to yield opacity effects within the theory of Targeted Constraints of Wilson (2000, 2001), which is arguably interpretable as a formalization of segment-level attraction effects (Burzio 2000b). As for the LC approach, it has so far been shown to provide an account of only a subset of the opacity effects, in particular, and as already noted, that of counterfeeding chain shifts, by means of a LC of faithfulness constraints (Kirchner 1996).

Like the LC approach, the serial OT approach is potentially challenged by cases where representational distance does not arise through a specific phonological process. In particular, the schema of (14) will not be applicable to the parasitic assimilation of (9), where there is no derivational step that would precede place assimilation and could account for the contrast. Likewise, (14) may not extend to cover the cases in (13), where the relevant (attraction-weakening) differences are suppletive or semantic.

4.2 *Morphologically derived environments in boundary contexts*

4.2.1 The entailments-based approach

Burzio (2000a) analyzes the Finnish case in (3b) above, repeated in (15a), as in (15b).

(15) a. *tilat-i* → *tilas-i* 'order'-PAST
b. FAITH[ti] >>*TI (assibilation) >> FAITH[t]

In (15a), the first *t* in *tilat-i* is subject to higher-ranked FAITH[ti] and thus resists the effects of the assibilation constraint, while the second *t* is subject only to lower-ranked FAITH[t], and thus undergoes the assibilation. This analysis assumes that faithfulness only applies to the form of morphemes, not to morpheme combinations. Hence the output form *tilas-i* in (15a) would only be faithful, independently, to the stem *tilat-*, and to the past tense affix *-i*, but not to their concatenation, making the top-ranked constraint in (15b) inapplicable to heteromorphemic *t-i*.

While the analysis in (15b) predates the introduction of the representational entailments of (8) above, the latter can be recruited to improve it. The reason is that the constraint FAITH[ti] of (15b) will now consist of the entailment $i \Rightarrow / t$ __ (*i* entails a preceding *t*, or *i* must be preceded by *t*), an entailment generated by any surface occurrence of the stem *tilat-* or its allomorph *tilas-*. The reason there is no heteromorphemic counterpart to that entailment is that affixal -*i* occurs in heterogeneous environments, attached to stems of all sorts. Any entailment of a preceding *t* would be contradicted by entailments generated by other stems that do not end in *t*, and thus effectively suppressed under algebraic summation of entailment. Of course one must assume that the *i* involved in such entailment is not just the phoneme *i* of Finnish. Rather, it must be the *i* of *tilat-*, which is in turn entailed by the rest of the representation of the stem *tilat-*, including its semantics. It is that *i* which will then entail a *t* by transitivity of entailments, while suffixal -*i* will not. See Wolf (2008: 326f.) for a review of other attempts in the literature, some of which bear resemblance to the one being described.

The remaining question is still which sequences of segments may count, which now takes the form of "what entailments matter, exactly?" For instance, why shouldn't an entailment from *a* to *t* in (15a) prevent the second *t* from spirantizing while allowing the first, hence yielding **silat-i*? Here, the REH (8) has been shown to yield one other critical effect besides attraction: "binding," tying together components that are relatively similar to one another. While a formal demonstration is beyond the goals of this chapter, an intuitive grasp can be attained by considering similarity as a sharing of subcomponents. Given two co-occurring components A, B, the REH will prescribe that each of A and B will entail its own internal structure, as well as each other. But when A and B are similar, the former effect will contribute to the latter, resulting in a stronger mutual entailment between A and B. See Burzio (2005: §4.5) and Wayment *et al.* (2007) for more discussion.

Now, if all assimilations are parasitic on similarity, as suggested above, then the "*t-i* → *s-i*" process of Finnish will suggest, along with articulatory considerations, that there is some level of similarity between *t* and *i*, with a consequent binding effect. A tautomorphemic sequence *ti* would then be simultaneously subject to two effects. One would be attraction, with the potential for assimilation, and the second the "binding," or enhanced mutual entailment, of *t* and *i*, opposing assimilation and evidently sufficient to block it. On the other hand, heteromorphemic *t-i* would experience only attraction, thus leading to assimilation. The reason is that "binding" only describes an enhancement effect over the entailment $i \Rightarrow / \ t \ \_\_$, an entailment which, however, is effectively false for the heteromorphemic case, for the reasons discussed. While the OT analysis thus remains as in (15b), with Faith[ti] describing the enhanced entailment (binding) and *TI describing attraction, the "grounding" just provided narrows the range of applicability of sequential constraints of type Faith(xy), and hence the range of pathological predictions re segmental inventories.

4.2.2 *The "local conjunction" approach*

Łubowicz (1999, 2002) proposes a "local conjunction" analysis of this case as well. When applied to the Finnish case in (15a), that analysis would be as in (16).

(16) [*TI (assibilation) & R-Anchor(Stem, σ)] >> Ident[cont] >> *TI (assibilation)

The constraint R-anchor(Stem, σ) in (16) requires that the right edge of the stem line up with a syllable boundary, and is violated in (15a). Hence the repair in (15a) is triggered by the fact that the conjunction of the latter constraint with the assibilation constraint dominates the relevant faithfulness constraint, IDENT[cont]. No repair affects the initial *ti* sequence, because in that case violation of the assibilation constraint is not in the same local domain as the violation of the right-anchor constraint, making the local conjunction inapplicable, while the assibilation constraint by itself is ranked lower than IDENT[cont] and hence ineffective.

The prediction that syllable boundaries play the role characterized by (16) seems disconfirmed by several of the cases in (5) above, however. This is true in particular in the following cases: Chumash pre-coronal laminalization, which alongside the example in (5d) also features /s-is-lusisin/ → [ʃ-iʃ-lu-sisin] (from Poser 1982, 1993, with presumed syllabification [ʃiʃ.lu . . .], with no misalignment); Sanskrit retroflexion (5e), as in /agni-su/ → [agni-ṣu]; Finnish cluster assimilation (5g), as in /pur-nut/ → purrut; and Mohawk epenthesis (5h) /k-wi'stos/ → [kewi'stos]. It is possible that these cases could be analyzed as processes that affect only affixal material due to a lower-ranked AFFIX-FAITH and that block in underived environments simply because those environments are stems, hence not for the reasons provided by (16). Yet this alternative does not seem viable for the case in (5f), repeated in (17).

(17) *Indonesian nasal substitution* (Pater 1999)

/məN-pilih/ → [məm-ilih] 'to choose' /əmpat/ 'four'

Here the stem is affected as much as the affix, and yet a faithful candidate *[mən-pilih] would feature no misalignment (though in any event a "left" rather than a "right-anchor" constraint would be needed).[3] Pater (1999) in fact analyzes the asymmetry in (19) not in terms of LC, but rather in term of a ROOT LINEARITY constraint that aims to preserve sequential relations within a root. The mapping /əmpat/ → *[əmat] 'four' would violate such a constraint with respect to the input sequence /mp/, while a comparable heteromorphemic sequence is not in the scope of the constraint. This type of analysis is in fact essentially subsumed by the above entailment-based discussion (/m/entails /p/).

4.2.3 Serial OT

Wolf (2008: ch. 4) provides an account of the now familiar Finnish test case in (15a) based on the ranking schema in (18):

(18) PREC(insert-affix, IDENT[cont]) >> *TI (assibilation) >> IDENT[cont]

Again, the schema in (18) differs from the entailment-based one in (15b) only by the top-ranked constraint, parallel to the one Wolf deploys in (14) above to handle phonologically derived environments. In both of Wolf's analyses, the top-ranked constraint imposes a specific order of processes, here demanding that a violation of IDENT[cont] (i.e. assibilation) be preceded by affixation. Such constraint is obviously satisfied by the heteromorphemic assibilation *t-i* → *s-i*. It would,

[3] Łubowicz (2002: 265) acknowledges this prediction with regard to simple nasal assimilation, which would not be expected to "block" since morpheme and syllable boundaries line up, finding no counterexamples, but does not discuss nasal substitution. However, English nasal assimilation does in fact seem to exhibit NDEB, judging from *Finland, Henry, only*, compared with *illegal, irrational*. But these cases too could perhaps be attributed to lower-ranked AFFIX-FAITH.

however, be violated by a tautomorphemic assibilation yielding **silas-i*. The reason is essentially that in the latter case affixation does not provide any material critical to the assibilation, and is thus taken not to count as establishing the required order in Wolf's formal system.

In sum, this approach provides a unitary account of two of the cases of NDEB in terms of the PREC constraints of Wolf's framework. We will see in §6 below, however, that it appears not to extend to the third.

5 Pre-Optimality Theory accounts

5.1 *Strict cyclicity*

This subsection reviews the most influential of the early accounts of NDEB, the one based on "strict cyclicity," as developed by Mascaró (1976) as part of the conception of the phonological "cycle." Other accounts, carefully reviewed in Kiparsky (1993), include the "(Revised) Alternation Condition" and specific applications of the "Elsewhere Condition," briefly discussed below.

The traditional motivation for the phonological cycle comes from the observation that underlying representations (URs; CHAPTER 1: UNDERLYING REPRESENTATIONS) are insufficient for correct phonological derivations, and that reference to surface forms is also necessary (CHAPTER 85: CYCLICITY), as in the familiar English example in (19), from Chomsky and Halle (1968: 117).

(19) a. *cond*[ɛ]*nsation* cf. *con'd*[ɛ]*nse*
b. *comp*[ə]*nsation* cf. *comp*[ə]*nsate* **com'p*[ɛ]*nse*

The two nouns in (19) would have fully parallel URs, and yet exhibit different degrees of reduction in the bracketed vowels, a difference that seems predictable only by reference to the corresponding verbs in parentheses. The "cycle" thus required the phonology to apply first to inner morphological layers, first calculating *con'd*[ɛ]*nse* within *cond*[ɛ]*nsation*, and then move on (see Kenstowicz 1994: 204 for exact derivations, and Cole 1995; CHAPTER 85: CYCLICITY for a full review of the motivations for the cycle).

On the other hand, in the parallel approach that OT embraces, it becomes possible to argue that the reference to surface forms that cases like (19) make necessary is also sufficient, thus simply dispensing with the traditional URs, rather than requiring an additional notion like that of the cycle (Burzio 1996, 2000a). On this view, the lexicon is constituted of full surface forms, whose well-formedness, including their morphological relatedness, is calculated in parallel by a grammar that effectively just "checks" them, rather than deriving them step by step.

The ability to refer directly to surface forms has been forcefully advocated in the OT literature by way of the notion of output-to-output faithfulness (Benua 1997; McCarthy 2005; and many others), which has then been used to account for the effects of the derivational cycle like those in (16). However, many practitioners have continued in the tradition of considering the lexicon as being constituted of morphemes, which give rise to URs when assembled together.

In contrast to cyclicity, "strict cyclicity" effects, to which NDEB effects were in turn related, can perhaps be illustrated with the simple English series in (20), but see Kenstowicz (1994; §5.3) and Mascaró (1976) for the Catalan data that were actually utilized.

(20) a. *de'si:re* b. *de'si:r-able* c. *de,si:r-a'bil-ity*

Assume for present purposes that *-able* is one of the affixes that trigger vowel shortening, as shown for instance by *admi:re / admir-able*. As discussed above, shortening is only variable rather than systematic in stressed penultimate syllables, whence unshortened (20b) *de('si:ra)<ble>*, assuming an extrametrical syllable *<ble>*, as in (1) above. The point of note is that further derivatives of *de'si:rable* such as (20c) maintain this choice despite the fact that *-ity* is itself a shortening affix, witness *divin-ity* (cf. *divi:ne*), or *promiscu-ity* (cf. variant *pro:miscuous*), and the fact that metrical environments identical to that of (20c) (medial foot (σ σ)) shorten quite regularly, as in *pro(,nunci)'ation* (cf. *pronounce*). The conclusion drawn from facts of this sort was thus that certain processes "block" in environments that are not "properly" derived. The one in (20c) would be one of them, since the shortening environment is already present in (20b) (let us say), attachment of *-ity* thus contributing nothing further. But if shortening blocks for those reasons in (20c), it may well block for the same reasons in (20a), or "trisyllabic" *i:vory*, which are not derived at all. In addition, since it is the definitional property of cyclic rules to apply *when* there is a new morphological environment, it will only be a matter of strengthening this definition to *only when* (thus making the cycle "strict"), to derive the blocking of both (20c) and (20a) or *i:vory*. Along with such cases, Strict Cyclicity would provide presumptive accounts of the cases of §3.2 above (non-boundary contexts), assuming all relevant processes can refer to the added morphological structure so as to activate the cycle and block elsewhere, and would similarly also account for "boundary context" cases of §3.1 above, like Finnish *tilas-i* of (3b). As for the cases that appear to be phonologically but not morphologically derived, like Finnish /vete/ → [vesi] of (3), a clause like (21b) below was from the inception added to the already established (21a).

(21) *Applicability of cyclic rules*

a. Contexts that are newly created morphologically.
b. Contexts that are newly created phonologically.

This definitional fiat extended the account to phonologically derived cases such as (2) and (3) above (I leave (13) aside), but it did so at a cost. The two forms of NDEB are now predicted to be co-extensive, an incorrect conclusion, as Łubowicz (2002: 271) points out. While there are cases, like Finnish assibilation and Sanskrit *ruki* retroflexion, that can be argued to occur in both types of derived environments, others do not. The problem is perhaps best illustrated by the English cases in (1). The case in (1a) *re'me:di-able* was argued above to be phonologically derived via restressing. Under (21), this would be because of clause (b). But the simultaneous presence of clause (a) would now incorrectly predict lengthening in (1b) **'le:vi-able* as well, since both cases involve affixation of *-able* and hence are both "derived" in one way or another. This problem persists under Kiparsky's (1982) attempt to reduce Strict Cyclicity to the Elsewhere Condition (Anderson 1969; Kiparsky 1973b; also referred to as the Pāṇini Principle), according to which processes that refer to more specific contexts trump processes that refer to more general ones. This principle was conjoined with the assumption that lexical items (morphemes) undergo identity rules. In a case like Finnish /tilat-i/, cf. above, the assibilation rule would be blocked relative to tautomorphemic /ti/ since the identity rule would apply to /tilat/, a more specific context than just /ti/. In the case of heteromorphemic /t-i/, however, the context of the assibilation rule would not be contained within the context of the identity rule, allowing that rule to apply with no inhibition, yielding [tilasi]. Similarly, in Finnish /vete/ → [vesi], the assibilation rule would find no obstacle once /e/ raised to [i], since the identity rule would apply only to /vete/, which does not contain the assibilation context /ti/. However, again, a case like *levi-able* of (1b) above would not be accounted for, since the context for CiV lengthening obtains here thanks to combination of morphemes, hence is beyond the reach of an identity rule for *levy*. Nonetheless, the Elsewhere Condition account appeared to solve the puzzle posed by the stress rule which, while seemingly cyclic in light of cases like (19), did not block in underived environments. This could now be attributed to the fact that, in the case of regular stress systems, lexical items would not contain stress information, which could therefore not be referred to by the identity rules.

Still working within a rule-based system, Kiparsky (1993) rejects the Strict Cyclicity account of NDEB, not only because of the incorrect predictions of the disjunction in (23), but also because various other hallmarks of cyclicity failed to correlate with NDEB effects. For instance, one diagnostic of cyclicity would be sequential orders like P_1, P_2, P_1, where a process P_1 is found to apply both before and after P_2. The cycle would enable such orders so long as the two occurrences of P_1 could be placed in different cycles. Other properties attributed to cyclic rules, though essentially by stipulation, were application at "lexical" as opposed to phrasal levels, and their contrastive, as opposed to allophonic, character. Kiparsky shows clearly that NDEB does not correlate with these attributes, and thus proposes the alternative I review next.[4]

5.2 *Underspecification*

Kiparsky's (1993) proposal can be illustrated for the Finnish cases as in (22), where upper-case *T* is assumed to be underlyingly not specified for continuancy, while lower-case *t* is fully specified as [−continuant] (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION).

(22)	a.	tilaT-i			→	tilas-i	'order-PAST'	(cf. *tilat-a* 'order-INF')
	b.	veTe	→	veTi	→	vesi	'water-NOM'	(cf. *vete-nä* 'water-ESS')

Correct derivations in (22) are ensured by assuming that the assibilation process can fill in the value [+continuant] in the context / __ *i*, yielding [s], but not change fully specified *t*. At the same time, one must also assume that [−continuant] can be filled in by a later default rule to any representation that may remain unspecified after the assibilation rule has had a chance, so as to yield *t* rather than

[4] Note that parallel OT does not predict any of the formerly stipulated distinctions between lexical and post-lexical processes. A discussion of this issue is beyond present goals, but see below for some discussion of phonology–morphology interaction.

s in the parenthesized forms on the right, as in /tilaT-a/ → [tilat-a]; /veTe-nä/ → [vete-nä].

On the one hand, this proposal bears some similarity to the one in (15b) above and its entailment-based version, in that immunity of the first /t/ comes from a greater amount of information associated with it. On the other, however, this approach does not appear to be fully workable. If the forms on the right in (22) were the sole determinants of the URs on the far left, then, indeed, the latter URs could be correctly obtained. In the course of acquisition, the *i* of *tilat-a* would force full specification of the preceding *t*, lest the expected form be **silat-a*, given knowledge of the assibilation. Elsewhere, however, *t* (as given in upper case) could remain underspecified, correctly yielding assibilation when a following *i* shows up. This dynamic is parallel to the one based on the entailments, where a following *i* would also confer additional resilience to a *t*. However, as argued in Burzio (2000a), the assumption that the parenthesized forms are privileged sources for the URs cannot be maintained. In *divine* / *divin-ity*, the presumed UR *div*/iː/*ne*/, with a long vowel, must be inferred from the base, but in *dam(n)* / *damn-ation*, the UR /dæmn/, with /n/, would only be inferrable from the derivative. On the other hand, in [pærənt] / [pərɛntəl], the full set of underlying vowels is only inferrable from base and derivative combined. The fact of the matter is that, in general, there is no independent principle or a priori restriction on what surface forms can contribute to a UR in a theory that has URs. It totally depends on where neutralization processes occur down the line. This means that hypothetical **tilat-i* could also contribute to its UR, yielding full specification of the second *t* incorrectly, making the account of (22) circular (the initial premises rest on the final results). This liability is not shared by the entailments. As we have seen, affixal -*i* would not entail a preceding *t* even in a hypothetical *tilat-i*, the reason being that, as a past tense affix, -*i* will entail to its left whatever results from entailment summation over all of its stems, most of which do not end in *t*.

The inadequacy of the underspecification account is even more apparent if one attempts to extend it to other cases of phonologically derived environments beside (22b). For example, to handle *remediable* in (1a), this account would have to underlyingly underspecify the lengthened vowel while fully specifying (as short) the one of *leviable* (1b). But, again, there is no independent basis for such an asymmetry. The fundamental reason for this inadequacy is that, in general, an underspecification account will predict asymmetries based on some coherent theory of what can be marked *vs.* default values. Hence, it cannot under any circumstance predict the true generalization, which appears to be whether or not some change has occurred in the same segment or morpheme, as in each of (1), (2), and (13).

6 Morphologically derived environments in non-boundary contexts

I turn now to the case in §3.2 above, whose prototype was taken to be English vowel shortening.

Burzio (2000a) argues that such effects simply result from the parallel interaction of morphology and phonology in the absence of any level of UR. Consider in particular that regular/productive morphological systems generally exert an inhibitory effect on phonological processes, as shown in (23).

(23) a. *'effort-less-ness* (exceptional stress)
b. *beep-ed* [bijpt] (exceptional syllable size)

Stress patterns such as the one in (23a) are unattested among morphologically underived items, as are syllables like the one in (23b), where a long vowel is followed by two consonants. These effects would be types of "DEB," namely reversals of NDEB. Here, otherwise regular processes "block" exactly in the derived environments. Burzio (2002a) argues in this connection that one can simply interpret the selectional properties of the relevant affixes, expressible for example as in (24), as types of constraints.

(24) *-less* ⇒ / Noun __ (*-less* attaches to a noun)

If there are no URs, then the context "Noun" in (24) can only refer to surface forms, affix *-less*, thus demanding identity between its stem and any such form. Then, cases like (23a) will be accounted for by taking (24) to dominate the constraints responsible for regular stress, the irregular one of (23a) simply coming from identity with that of *'effort*, as imposed by (24), and similarly for the past tense affix in (23b) and syllabification constraints.

If morphology and phonology compete this way, so that (23a) and (23b) result from the morphology winning, then effects in the opposite direction should result when the morphology loses, and this would be the case of vowel shortening, as illustrated in (25).

(25) a. *natur-al* (cf. *na:ture*)
b. *V: >> *-al* ⇒ / Noun __ (*-al* attaches to a noun)

In (25b), a general markedness constraint banning long vowels outranks the morphological constraint demanding identity with the independent noun *na:ture*, resulting in a short vowel. The difference between the high-ranked selectional constraint in (24) and the low-ranked one in (25b) reflects the general difference between (roughly) Germanic and Latinate affixes, termed respectively "Level 2" and "Level 1" in Kiparsky's (1982) Lexical Phonology framework. This ranking-based characterization of the two morphological systems is independently supported by the fact that the two systems also differ with respect to morphological idiosyncrasy (Burzio 1994, 2002a; Benua 1997), as well as productivity, which can also be correlated with rank, though more indirectly (see Burzio 2006). Latinate affixes like the one in (25) tolerate massive amounts of morphological idiosyncrasy, as in *arbore-al* (cf. absence of **arbore*), *crimin-al* (cf. *crime*, not **crimin*). By contrast, the Germanic affixes exhibit virtually no idiosyncrasy like hypothetical **arbore-less* (cf. *tree-less*) or **crimin-less* (cf. *crime-less*).

Hence there exists an inverse correlation between phonological and morphological regularities (Burzio 2002a), as Germanic affixes exhibit tight morphological regularity along with abundant phonological irregularity as in (23), while the Latinate ones reverse both effects, exhibiting much morphological irregularity along with regular phonology (CHAPTER 103: PHONOLOGICAL SENSITIVITY TO MORPHOLOGICAL STRUCTURE). This includes regular stress aside from a specific range of cases discussed below, as in *pa'rent-al* (not *'*parent-al*, which would parallel *'effort-less*), regular syllabification, the shortening of (25), and other processes like the velar softening of (4) above (*electri*[s]-*ity*), all absent from the Germanic class (cf. *froli*[k]-*ing*, *crime-less*). The overriding generalization is thus in terms of the requirement that a stem be identical to an independent surface form, referred to as output-output (OO) faithfulness in the literature (Benua 1997), which is strong for one class of affixes, but weak for the other (Burzio 1994). Two ingredients are critical to a successful account. One is that morphology (interpretable as OO-faithfulness) must be constraint-based to compete with phonology this way. The other is that there must be no UR. The first ingredient, constraint-based morphology, can be taken to result from the REH, as argued in Burzio (2002b). Constraints like those in (24) and in (25b) and their ranking can be viewed as the result of summation of identical entailments across the lexicon. The second ingredient, absence of a UR level, is a natural hypothesis for a constraint-based system. To see how it is critical to the analysis, consider first that, in the Germanic/Level 2 case of (23), the UR would be simply superfluous. One only needs to assume a high-ranked OO-faithfulness (expressed here by (24)) to express the identity of each stem to the independent words. The effect of a lower-ranked faithfulness to a UR (termed input-output (IO) faithfulness) would be cancelled by the higher-ranked OO-faithfulness. In the case of (25), however, a hypothetical UR is not just superfluous, but false. To account for the long vowel in *na:ture*, one must assume the ranking "IO-Faith >> *V," the standard OT schema for marked choices. But if the same UR was input to both *na:ture* and *natur-al*, as is the case by the standard definition of UR as the common input to all allomorphs of the same morpheme, then *natur-al* should also have a long vowel. Just supplementing IO-Faith with OO-Faith is therefore not sufficient in this case. Rather, IO-Faith as faithfulness to a UR must be removed from the scene altogether, and the only principled way to do so is to drop the already unnecessary as well as insufficient UR. Hence, *na:ture* will be faithful to its own input with whatever rank the language at large has. The form *natur-al* will also be faithful to its input, but that input (except for *-al*) is the word *na:ture* and not a UR, and that ranking is determined by the particular morphological system, not by the language at large. Affixes like *-al* in (25), which are relatively unproductive and prone to idiosyncrasy, evidently establish relatively weak/low-ranked associations.

The notion of "cyclic" derivation would therefore have seemed right here in a way, requiring that *na:ture* be derived from a UR, while *natur-al* would be derived from *na:ture*. But such a notion has turned out to be right only when it reproduces (in a more complicated way) the effects of just dropping the UR from the theory, which forces reference to surface forms directly. In other respects, the cycle, and its cluster of attribute properties, prove to have been incorrect, as Kiparsky (1993) shows for the association "cyclic = blocking in NDE." The presumed further association "cyclic = lexical" (see discussion in Kenstowicz 1994: 195f.) also fails to hold. Here the notion "lexical" defines the presence of idiosyncrasy, while the property "cyclic" identifies, by its core definition, preservation of phonological structure, as in *cond*[ɛ]*nsation* (19a) above. As argued for (23), it turns out that such preservation is massive with Germanic/Level 2 affixes (no changes compared with the base word). This would make them "cyclical" if one took those definitions seriously. But idiosyncrasy is absent in the presence of those affixes, which would make them "non-lexical" (i.e. more like phrasal constructs). Conversely, Latinate/Level 1 affixes would be part of the "lexical" morphophonology by virtue of the noted idiosyncrasies, but the preservation effects they generate are in fact tenuous, *cond*[ɛ]*nsation* notwithstanding, making them only marginally "cyclic." As argued in Burzio (1991, 1993, 1994), such effects consist of only minimal distortions of the standard footing, as in e.g. *phe*(ˌ*nome*)'*nology* (to match *phe'nomenon*), or *a'merica*(ˌ*nist*Ø) (to match *A'merican*). As in underived items, adjacent stresses are ruled out, for example as in ˌ*cata'strophic*, ˌ*infor'mation*, ˌ*consul'tation* (*ca'tastrophe*, *in'form*, *con'sult*). This contrasts with the more robust distortion of cases like (23a) '*effort-less-ness*. As for the apparent preservation of stress in *cond*[ɛ]*n'sation* of (19a), it is argued in Burzio (1994: 185) that in fact it only concerns the details of vowel reduction rather than stress itself, though the relation with *con'dense* remains relevant.

This means that, when faithfulness to a base word specifically refers to stress, it is relatively low-ranked for the Latinate affixes, consistently with (25b) above, but not so low-ranked as to be totally ineffective. This accentual faithfulness, referred to as Metrical Consistency (MC) in Burzio (1994), is what accounts for long vowels in this sector of the lexicon, as argued in §3.2 above. Consider here the contrast between (26a) and (26b), along with the partial hierarchy in (26c).

(26) a. *de'si:rous*, *ad'he:rent*, *ex'tre:mist*, *di'vi:sive*, '*medi*ˌ*ta:tive*, . . .
 b. '*blasphemous*, '*aspirant*, '*hypnotist*, '*relative*, '*generative*, . . .
 c. MC, *V: >> IDENT(V-length)

The hierarchy in (26c) is taken to be dominated in turn by constraints mandating well-formed metrical feet which exclude, along with adjacent stresses and other degeneracies, stress on light penultimate syllables. This means that, when a long vowel and the stress of the base word both end up in a penultimate syllable in a derivative, as in all of the cases in (26a) and (26b), the two leftmost constraints in (26c) cannot be simultaneously satisfied. Either the vowel will have to fail to shorten, as in (26a), or the stress of the base will be lost, as in (26b). The variation between the cases in (26a) and those in (26b) then reveals that the grammatical system is indeterminate on the relative ranking of the two leftmost constraints in (26c), allowing lexical information to choose outcomes (see Burzio 2006). The effect observed for some of (13) above may also be at work, however. That is, items that are semantically very close to their base, as perhaps those in (26a) are to a greater extent than those in (26b), may end up accentually faithful as well, with no shortening.

In sum, the main class of exceptions to vowel shortening finds the principled account in (26c) leading to the conclusion that vowel shortening is – on its own – perfectly general (further scattered exceptions aside) in the Latinate lexicon. On the analysis in (26c), this would then be a case of what is referred to in the OT literature as "The Emergence of the Unmarked" (TETU; McCarthy 2002: 129f.; CHAPTER 58: THE EMERGENCE OF THE UNMARKED).

Other approaches to NDEB do not cover this type of case. Wolf (2008) refers to putative cases of this sort, in which the phonological process does not appear to depend on the specifics of the morphological operation, as "pseudo" DE effects, and suggests the existence of analyses that would bring them in line with the non-pseudo cases. Successful re-analysis is critical to his approach, which predicts non-existence of pseudo DE effects. The reasons for this prediction by Wolf's serial OT approach are effectively the same as the reasons excluding assibilation of tauto-morphemic /ti/ in the Finnish case discussed above. Specifically, and for example, if vowel shortening as in *div*[ɪ]*n-ity* was due to an a-contextual

"*V:" constraint, then the same output will be produced whether shortening applies before or after affixation of *-ity*. But then a mechanism of "chain merger" would remove any ordering between the two operations, so that a putative precedence constraint PREC(insert-affix, IDENT[cont]) will always be violated if shortening applies, thus keeping all vowels long if high-ranked, while allowing all to shorten, including in **div*[ɪ]*ne*, if low-ranked.[5] However, while Wolf (2008: §4.3.5) does review the above "TETU" analysis of vowel shortening, he provides no alternative to it. He also concedes the existence of other cases of markedness reduction under affixation, such as those in (7e) and (7f) above, which the TETU analysis can handle, but which he attributes to failure of a feature to "percolate" up the morphological structure (2008: 268), an additional mechanism, in his perspective.

REFERENCES

Anderson, Stephen R. 1969. West Scandinavian vowel systems and the ordering of phonological rules. Ph.D. dissertation, MIT.

Benua, Laura 1997. Transderivational identity: Phonological relations between words. Ph.D. dissertation, University of Massachusetts, Amherst.

Bolognesi, Roberto. 1998. The phonology of Campidanian Sardinian: A unitary account of a self-organizing structure. Ph.D. dissertation, University of Amsterdam.

Burzio, Luigi. 1991. On the metrical unity of Latinate affixes. In Germán Westphal, Benjamin Ao & Hee-Rahk Chae (eds.) *Proceedings of the Eastern States Conference on Linguistics* 8, 1–22. Reprinted (1995) in Hector Campos & Paul M. Kempchinsky (eds.) *Evolution and revolution in linguistic theory: Essays in honor of Carlos Otero*, 1–24. Washington, DC: Georgetown University Press.

Burzio, Luigi. 1993. English stress, vowel length and modularity. *Journal of Linguistics* 29. 359–418.

Burzio, Luigi. 1994. *Principles of English stress*. Cambridge: Cambridge University Press.

Burzio, Luigi. 1996. Surface constraints versus underlying representations. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, vol. 1, 123–141. Salford: ESRI.

Burzio, Luigi. 1998. Multiple correspondence. *Lingua* 104. 79–109.

Burzio, Luigi. 2000a. Cycles, non-derived-environment blocking, and correspondence. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: Phonology, syntax, and acquisition*, 47–87. Oxford: Oxford University Press.

Burzio, Luigi. 2000b. Segmental contrast meets output-to-output faithfulness. *The Linguistic Review* 17. 368–384.

Burzio, Luigi. 2002a. Missing players: Phonology and the past-tense debate. *Lingua* 112. 157–199.

Burzio, Luigi. 2002b. Surface-to-surface morphology: When your representations turn into constraints. In Paul Boucher (ed.) *Many morphologies*, 142–177. Somerville, MA: Cascadilla Press.

Burzio, Luigi. 2005. Sources of paradigm uniformity. In Downing *et al.* (2005), 65–106.

Burzio, Luigi. 2006. Lexicon versus grammar in English morphophonology: Modularity revisited. *Korean Journal of English Language and Linguistics* 6. 437–464.

Burzio, Luigi. 2007. Phonologically conditioned syncretism. In Fabio Montermini, Gilles Boyé & Nabil Hathout (eds.) *Selected proceedings of Décembrettes* 5, 1–19. Somerville, MA: Cascadilla Press.

[5] Thanks to Matt Wolf (personal communication) for assistance on this point.

Burzio, Luigi & Niki Tantalou. 2007. Modern Greek accent and faithfulness constraints in OT. *Lingua* 117. 1080–1124.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Chung, Sandra. 1983. Transderivational relationships in Chamorro phonology. *Language* 59. 35–66.

Cole, Jennifer. 1995. The cycle in phonology. In Goldsmith (1995), 72–113.

Downing, Laura J. 2006. *Canonical forms in prosodic morphology*. Oxford: Oxford University Press.

Downing, Laura J., T. A. Hall & Renate Raffelsiefen (eds.) 2005. *Paradigms in phonological theory*. Oxford: Oxford University Press.

Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles. Published 2002, London & New York: Routledge.

Flemming, Edward. 2003. The relationship between coronal place and vowel backness. *Phonology* 20. 335–373.

Flemming, Edward. 2008. Asymmetries between assimilation and epenthesis. Unpublished ms., MIT.

Fukazawa, Haruka. 1999. Theoretical implications of OCP effects on features in Optimality Theory. Ph.D. dissertation, University of Maryland at College Park.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.

Halle, Morris & K. P. Mohanan. 1985. Segmental phonology of Modern English. *Linguistic Inquiry* 16. 57–116.

Hansson, Gunnar Ólafur. 2001. Theoretical and typological issues in consonant harmony. Ph.D. dissertation, University of California, Berkeley.

Hargus, Sharon & Ellen M. Kaisse (eds.) 1993. *Studies in Lexical Phonology*. San Diego: Academic Press.

Hulst, Harry van der & Jeroen van de Weijer. 1995. Vowel harmony. In Goldsmith (1995), 495–534.

Inkelas, Sharon & Cemil Orhan Orgun. 1995. Level ordering and economy in the lexical phonology of Turkish. *Language* 71. 763–793.

Itô, Junko. 1990. Prosodic minimality in Japanese. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26(2). 213–239.

Iverson, Gregory K. & Deirdre Wheeler. 1988. Blocking and the elsewhere condition. In Michael Hammond & Michael Noonan (eds.) *Theoretical morphology*, 325–338. San Diego: Academic Press.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kenstowicz, Michael & Jerzy Rubach. 1987. The phonology of syllabic nuclei in Slovak. *Language* 63. 463–497.

Kiparsky, Paul. 1973a. Phonological representations: Abstractness, opacity, and global rules. In Osamu Fujimura (ed.) *Three dimensions of linguistic theory*, 57–86. Tokyo: Taikusha.

Kiparsky, Paul. 1973b. "Elsewhere" in phonology. In Stephen R. Anderson & Paul Kiparsky (eds.) *A Festschrift for Morris Halle*, 93–106. New York: Holt, Rinehart & Winston.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.

Kiparsky, Paul. 1993. Blocking in nonderived environments. In Hargus & Kaisse (1993), 277–313.

Kirchner, Robert. 1996. Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry* 27. 341–350.

Kula, Nancy C. 2008. Derived environment effects: A representational approach. *Lingua* 118. 1328–1343.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulations of vowel quality systems: The role of perceptual contrast. *Language* 48. 839–862.

Łubowicz, Anna. 1999. Derived environment effects in OT. *Proceedings of the West Coast Conference on Formal Linguistics* 17. 451–465.

Łubowicz, Anna. 2002. Derived environment effects in Optimality Theory. *Lingua* 112. 243–280.

Łubowicz, Anna. 2003. Local conjunction and comparative markedness. *Theoretical Linguistics* 29. 101–112.

Mascaró, Joan. 1976. Catalan phonology and the phonological cycle. Ph.D. dissertation, MIT. Distributed 1978, Indiana University Linguistics Club.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John J. 2003. Comparative markedness. *Theoretical Linguistics* 29. 1–51.

McCarthy, John J. 2005. Optimal paradigms. In Downing *et al.* (2005), 170–210.

McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.

Oostendorp, Marc van. 2009. The phonology of the unspeakable. Paper presented at the 6th Old World Conference on Phonology (OCP 6), Edinburgh.

Pater, Joe. 1999. Austronesian nasal substitution and other NÇ effects. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*, 310–343. Cambridge: Cambridge University Press.

Poser, William J. 1982. Phonological representations and action-at-a-distance. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part II, 121–158. Dordrecht: Foris.

Poser, William J. 1993. Are strict cycle effects derivable? In Hargus & Kaisse (1993), 315–321.

Prince, Alan. 1975. The phonology and morphology of Tiberian Hebrew. Ph.D. dissertation, MIT.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Rose, Sharon & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80. 475–531.

Rubach, Jerzy. 1984. *Cyclic and Lexical Phonology: The structure of Polish*. Dordrecht: Foris.

Rubach, Jerzy. 1993. *The Lexical Phonology of Slovak*. Oxford: Clarendon Press.

Rubach, Jerzy. 1995. Representations and the organization of rules in Slavic phonology. In Goldsmith (1995), 848–866.

Smolensky, Paul. 1997. Constraint interaction in generative grammar II: Local conjunction or random rules in Universal Grammar. Paper presented at the Hopkins Optimality Theory Workshop/University of Maryland Mayfest 1997.

Steriade, Donca. 1994. Positional neutralization and the expression of contrast. Unpublished ms., University of California, Los Angeles.

Steriade, Donca. 1999. Lexical conservatism in French adjectival liaison. In J.-Marc Authier, Barbara E. Bullock & Lisa A. Reed (eds.) *Formal perspectives on Romance linguistics*, 243–270. Amsterdam & Philadelphia: John Benjamins.

Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Studies in honor of Paul Kiparsky*, 151–179. Cambridge, MA: MIT Press.

Wayment, Adam. 2009. Assimilation as attraction: Computing distance, similarity, and locality in phonology. Ph.D. dissertation, Johns Hopkins University.

Wayment, Adam, Luigi Burzio, Donald Mathis & Robert Frank. 2007. Harmony versus distance in phonetic enhancement. *Papers from the Annual Meeting of the North East Linguistic Society* 37(2). 253–266.

Wilson, Colin. 2000. Targeted constraints: An approach to contextual neutralization in Optimality Theory. Ph.D. dissertation, Johns Hopkins University.

Wilson, Colin. 2001. Consonant cluster neutralisation and targeted constraints. *Phonology* 18. 147–197.

Wolf, Matthew. 2008. Optimal interleaving: Serial phonology–morphology interaction in a constraint-based model. Ph.D. dissertation, University of Massachusetts, Amherst.

76 Structure Preservation: The Resilience of Distinctive Information

Carole Paradis
Darlene LaCharité

1 Introduction

All languages have a phonemic inventory, including a set of distinctive vowels and consonants, i.e. linguistic sounds that contribute to the meaning of a word. For instance, *chip* [ʧʰɪp] contrasts with *cheap* [ʧʰip] in English, on the basis of the vowel quality; in the first case, the high front vowel is lax, whereas in the second one it is tense. We therefore say that /ɪ/ and /i/ are two distinct phonemes (segments) in English (CHAPTER 11: THE PHONEME) and that [tense] is a distinctive feature (CHAPTER 17: DISTINCTIVE FEATURES) for high vowels in this language.[1] While phonemic inventories are built in agreement with the principles of Universal Grammar (UG), the exact composition of a phonemic inventory varies from one language to another. Along with the suprasegmental inventory, the phonemic inventory is a good part of what allows a listener to identify a language at first glance and to distinguish it from other languages. We expect speakers to resist either dropping phonemes or phonemic contrasts from their language's inventory, or introducing new phonemes and phonemic contrasts – although this constitutes the bread and butter of language change – since the automatic consequence of such moves is a different system. We believe that resistance to change cannot be due simply to inertia – it is not passive. In this chapter we will try to show that resistance to change is, above all, a question of contrast/category pattern resilience in the mind of the speaker, which is expressed intralinguistically (i.e. resistance to change due to the passage of time, dialect contact, etc.) and also interlinguistically (between L2 and L1, as will be illustrated in §3 with respect to loanwords). We will link contrast resilience to the traditional notion of Structure Preservation, providing a history of this notion in generative grammar in §2, and considering in §3 the question of whether it is still pertinent now that phonological rules have given way to constraints. We will also address the relation between Structure Preservation and phoneme/structure resilience in loanword adaptation from the point of view of L1 and L2. We conclude in §4.

[1] Even if /i/ and /ɪ/ were to be distinguished by vowel length instead of tenseness, as proposed by some authors, the point made here would stand.

2 The history of Structure Preservation

It has long been noted that, intralinguistically, languages (or, more properly, their speakers) resist phonemic change before succumbing to and accepting a new phonemic contrast. Changes to a given phonemic inventory follow defined steps, which are gradual, and characteristically occur over a long period of time (CHAPTER 2: CONTRAST). Although such sound changes can sometimes occur relatively rapidly, it is not unusual for them to take centuries to complete. Broadly speaking, a small phonetic detail becomes sufficiently large over time that what begins by distinguishing phonetic variants ends up being categorical, i.e. phonemic (see Harris 1990 and Bybee 2008 for a detailed description of these steps). Clearly, though, the forces of change are counterbalanced by resistance to change, or intralingual change would typically proceed at a much faster rate and produce much more dramatic results than it usually does (CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME).

The lexicon is the crucial place where the battle between the forces of change and resistance to change takes place. In Lexical Phonology, the resistance to using non-phonemic sounds or sound combinations at the lexical level was expressed through the notion of Structure Preservation (SP). In Kiparsky (1982, 1985), SP regulated the application of phonological rules, constituting a ban on the introduction of phonemes at the lexical level that are not part of the underlying inventory.

(1) *Structure Preservation* (Kiparsky 1985: 88)

If a certain feature is non-distinctive in a language we shall say that it may not be specified in the lexicon. This means that it may not figure in non-derived lexical items, nor be introduced by any lexical rule, and therefore may not play any role at all in the lexical phonology.[2]

The model assumed by Kiparsky is basically that in Figure 76.1:

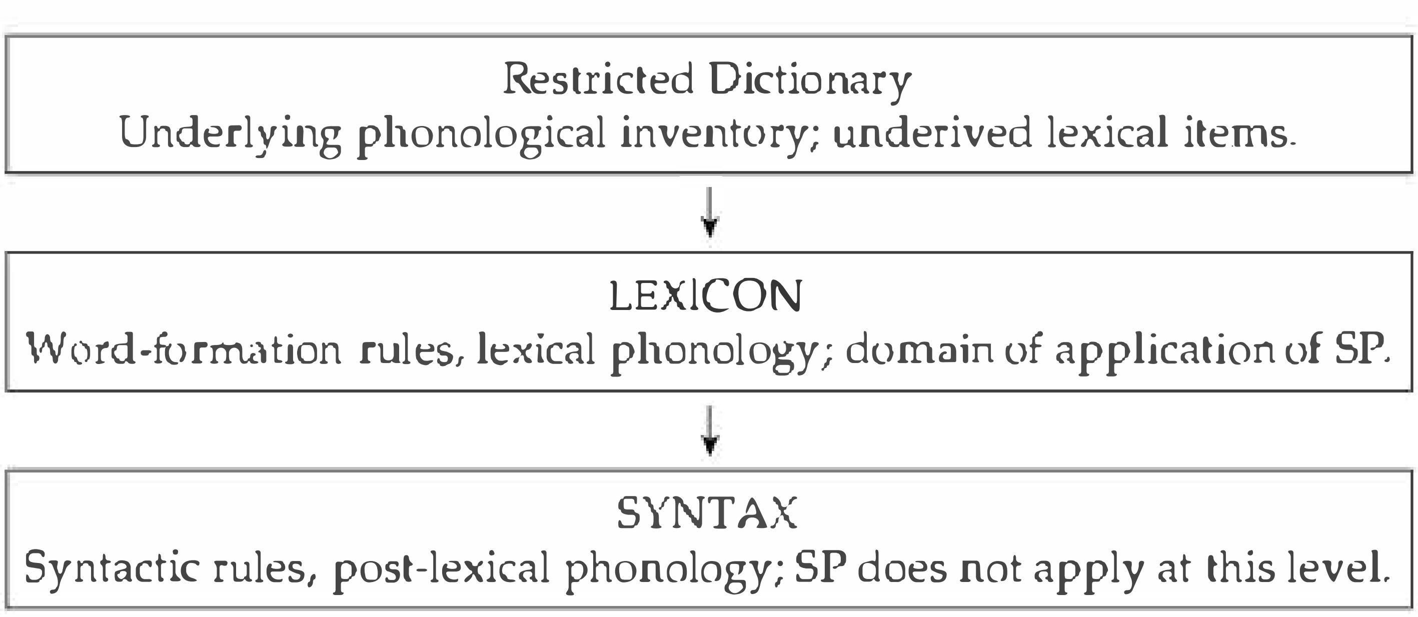


Figure 76.1 Lexical Phonology (Kiparsky 1982)

[2] Kiparsky does not present this constraint formally. The constraint given here is a description of SP as presented in the text by Kiparsky (1985: 88).

As Harris (1987: 255) puts it, "the lexical segment inventory of a language (the output of the lexical rules) must be isomorphic with the underlying inventory." Bybee (2008: 111) adds: ". . . alternations that are restricted to the word level involve only contrastive features. Segments or feature combinations that are non-contrastive must be introduced by postlexical rules . . ."[3]

Mohanan (1986) considers the formulation of SP given in (1) to be too restrictive. According to Mohanan, the Malayalam and English facts cannot be explained if SP is interpreted as in (1), so he softens it, saying instead, "the alphabet used for syntactico-phonological representations is the lexical alphabet" (1986: 174). The lexical alphabet refers to the phoneme inventory at the lexical level, which is the result of lexical (as opposed to post-lexical) application of phonological rules.[4] In Figure 76.1, Mohanan's lexical alphabet would be generated in the lexicon module by phonological rules that apply at this level. Of particular present relevance, the lexical alphabet in the view of Mohanan and Mohanan (1984) and Mohanan (1986) can contain distinctions that are absent from the underlying inventory (found in the restricted dictionary in Figure 76.1). The Malayalam case, detailed in Mohanan and Mohanan (1984), focuses primarily on contrasts in the system of nasals. The crux of the issue is that to achieve an elegant analysis of apparently complicated surface distributional restrictions on stops and nasals in Malayalam, one needs to assume that at the underlying level there are three nasals (bilabial, alveolar, retroflex), but that at the lexical level there are seven (bilabial, dental, alveolar, palato-alveolar, retroflex, palatal, velar). At the heart of their analysis is the clearly lexical application of two phonological rules (one that changes post-nasal voiced stops to nasals, and another that changes intervocalic [−continuant] velars to palatals when preceded by front vowels), which produces nasals with places of articulation that are not underlying for that class of sounds.

Still, the most widespread interpretation of SP in Lexical Phonology remains essentially the same: phonological rules are not expected to generate new phonemes or phonemic contrasts at the lexical level, nor are phonemes expected to undergo absolute neutralization at this level (CHAPTER 80: MERGERS AND NEUTRALIZATION). Any operations that introduce features that are not distinctive underlyingly are predicted to be necessarily post-lexical. For instance, French has a rich vocalic system that includes the mid-back lax and tense vowels /ɔ/ and /o/ (e.g. *hotte* [ɔt] 'hood' *vs.* *haute* [ot] 'high'). Although both vowels are frequent, /ɔ/ is prohibited word-finally in French (*/ɔ/#), at the lexical level. If a morphological operation produces a word-final /ɔ/ in the course of a derivation in French, it is systematically turned into [o]. Various morphological operations generate such a result; they include abbreviation (e.g. *Caroline* [kaʀɔlin] → *Caro* [kaʀo], *condominium* [kɔ̃dɔminjɔm] 'condominium' → *condo* [kɔ̃do]), gender inflection (e.g. *sotte* [sɔt] 'silly (FEM)' *vs.* *sot* [so] 'silly (MASC)'), verbal and adjectival derivation (e.g. *roter*

[3] Sproat (1985: 454) says that Structure Preservation could be interpreted as a restriction on contrasts that are the output of lexical rules, rather than a restriction on underlying representations. However, as Harris (1987: 259) points out, "given its inherent circularity [this interpretation] of Structure Preservation is hardly worthy of serious consideration."

[4] Mohanan and Mohanan (1984: 590) and Mohanan (1986: 12) consider that phonological rules are all part of a single independent phonological module and that they interact with either the lexical or post-lexical level or both, according to their domain specifications. Application in one domain or the other is subject to different restrictions. Notably, lexical application is subject to SP, provided that SP is interpreted in the less stringent manner indicated above.

[ʀɔte] 'to belch' *vs. rot* [ʀo] 'belch' (N)), reduplication (e.g. *dormir* [dɔʀmiʀ] 'to sleep' > *dodo* [dodo] 'sleep (N, child language)'), etc. The result is always the same: /ɔ/# → [o]#. Nonetheless, the restriction */ɔ/# in French does not apply at the post-lexical level. For instance, in Quebec French, final /a/s are systematically pronounced either as [ɑ] or [ɔ] (e.g. *chocolat* [ʃɔkɔlɑ], [ʃɔkɔlɔ] 'chocolat'; *matelas* [matlɑ], [matlɔ] 'mattress') or something in between, despite the fact that the lexical restriction */ɔ/# also applies in this variety of French. We know that [ɔ] and [ɑ] are variants of /a/ in such cases because derivatives such as *chocolaté* [ʃɔkɔlate] 'with chocolat' and *matelassé* [matlase] 'padded' indicate that the underlying vowel is /a/. Gradient and unstable rules such as /a/ → [ɑ] or [ɔ] in Quebec French are typically post-syntactic rules that are predicted not to occur at the lexical level (Mohanan 1986: 174). The existence of clearly necessary categorical constraints, such as */ɔ/#, alongside the existence of forms that clearly do not obey them is the kind of case that SP is intended to explain: a phonetic process can apply at the post-syntactic level in spite of the fact that its effect contradicts that of a phonotactic constraint at the lexical level.

Another classic example of a lexical (hence structure-preserving) rule, is velar softening in English (and also in French), where /k/ yields [s] before a high front vowel (e.g. *electric* [ilɛktrɪk] *vs. electricity* [ilɛktrɪsɪti]). As reiterated by Bybee (2008: 112), this rule does not apply between words, is unproductive, lexically restricted, and morphologically conditioned. In contrast to this lexical alternation, /k/ in English, as in French, has a palatal variant [c] before a front vowel, as in *key* /ki/ → [ci], *kiss* /kɪs/ → [cɪs], etc. (in French, *qui* /ki/ 'who' → [ci], *quitter* /kite/ 'to leave' → [cite], etc.). The emergence of the palatal variant in both languages is automatic, productive, and neither lexically nor morphologically restricted. SP embodies the claim that such an assimilation rule could not apply at the lexical level because, in both French and English, it would introduce at this level a sound, [c], that is not part of the phonemic inventory of either language (see Kiparsky 1985 for a discussion of many other assimilation and harmony processes that are non-structure-preserving and which he shows are post-lexical).

However, SP has been challenged on a variety of fronts. Its domain of application has been debated vigorously. Kiparsky (1982, 1985) proposes that lexical rule application is subject to SP but post-lexical application is not. However this neat division of territory between structure-preserving and non-structure-preserving rule application has proven to be doubtful. For example, Kaisse (1990), Rice (1990), and Hyman (1993) agree that SP is not necessarily turned off in the post-lexical component. In other words, post-lexical rules can be subject to SP. On the other hand, Harris (1987: 256) argues, mainly on the basis of a certain type of vowel harmony in southern Bantu languages, that "failure to preserve structure cannot be reliably considered proof of a rule's postlexical status." That is, lexical rule application is not necessarily subject to SP.

Harris (1987, 1989, 1990) discusses other allophonic processes that must be lexical, but that are not structure-preserving (see also discussions in Mohanan 1995 and Steriade 1995), but one of the best-known problematic cases is the distribution of [ç] and [x] in modern German. According to Hall (1989), there is no underlying contrast between velar and palatal fricatives in German; the feature [back], though distinctive for vowels in German, is not distinctive for fricatives. The [ç] *vs.* [x] contrast results from a rule of fricative assimilation that spreads the backness feature from a vowel to a following voiceless high fricative. Crucially, fricative assimilation applies lexically and it produces a phoneme/phoneme contrast that is not underlying. Hall (1989: 1) concludes that the rule of fricative assimilation is "a blatant counterexample to SP." We will address this case more thoroughly later.

Macfarland and Pierrehumbert (1991) propose an alternative view, which is intended to salvage the integrity of SP. In their view, non-distinctive features introduced at the lexical level stem from spreading, resulting in doubly linked structures. In the German case just discussed, the [back] feature of the vowel spreads to the feature matrix of the following fricative /X/, which is unspecified for backness. This results in the feature [back] being simultaneously linked to the vowel and the following fricative consonant. By virtue of their double linking, such structures are technically exempt from SP (as well as from a condition they call the marking condition). Though the solution might work for this and some other problematic cases that challenge SP, it does so at the cost of seriously weakening the SP constraint. Iverson (1993) proposes instead an approach that re-examines the relationship among some of the constellation of properties originally intended to distinguish between lexical and post-lexical rules or rule application, namely SP, and the restriction of applying to derived environments. In classical Lexical Phonology, a lexical rule had certain properties, two of which were that it preserved structure and that it applied in derived environments. SP is a consequence of a rule's lexical status in that view. Iverson turns this relationship on its head (1993: 265); if a given rule preserves structure, then it observes the derived environment constraint. This arguably explains the clustering of properties previously considered to be diagnostics of a rule's lexical or post-lexical status, but it implies that SP is a property of some, but not necessarily all, lexical rule applications. Indeed, Iverson concludes (1993: 270) that "... structure-building applications of lexical rules need not (though may) be structure-preserving."

As the previous discussion suggests, SP, as formulated by Kiparsky (1982, 1985) was inextricably linked to the overall architecture and other principles (e.g. the Strict Cycle Condition and the Derived Environment Constraint) and theoretical tools (e.g. underspecification) of Lexical Phonology. To reiterate, SP was part of a set of properties that distinguished lexical from post-lexical rule applications. Like many other notions of Lexical Phonology, SP was found to be problematic for a variety of reasons. For instance, even if we accept the view of Mohanan and Mohanan (1984: 589) that phonological rules whose domain of application is lexical yield the "lexical alphabet" – to be distinguished from the underlying one, found in the restricted dictionary in Figure 76.1 – it seems clear, as indicated by the work of a number of phonologists working on several different languages, that we cannot uphold the position that lexical rule application is necessarily structure-preserving while post-lexical rule application is not. For example, as mentioned previously, Iverson (1993) argues that lexical rules are not necessarily structure-preserving, while Rice (1990) argues that post-lexical rules may be. However, this is only one problem facing SP. Another is that SP has resisted formulation, interpretation, or application in any way that can be universally applied to yield felicitous results. Different attempts, including reformulation (e.g. Mohanan 1986: 174; Borowsky 1989: 148), reinterpretation (e.g. Macfarland and Pierrehumbert 1991: 179; Iverson 1993: 265), or restriction of its application to some, but not all, lexical levels – often on a language-specific basis (e.g. Borowsky 1986, 1989) – yield no universally satisfactory outcome. Moreover, such attempts often have extremely damaging consequences for the theory and the SP principle. For example, the effect of Mohanan and Mohanan's (1984) analysis of Malayalam and their distinction between an underlying and a lexical alphabet is to allow rules to introduce contrasts that are not underlying. This is clearly at odds with Kiparsky's (1982) view of the role that SP plays. A closely related problem is that, under no formulation or interpretation, in any language through which the principle has been test-driven to any extent, has SP been found to be exceptionless (see the discussion of Bybee 2008 below).

What has the outcome of the challenges to SP been? Sproat (1985), who rejects Lexical Phonology's approach to word formation altogether, considers SP completely dispensable, along with the rest of the theory. However, few phonologists would go this far. Although some, such as Mohanan (1989: 609) and Hall (1992: 233), have given up on reformulating SP, or tweaking the conditions of its application, and concluded that it is not a linguistic universal, they still consider it a cross-linguistic tendency. As Steriade (2007: 146) asserts, "... Structure Preservation cannot be abandoned altogether. ..."

Bybee (2008) has picked up the idea of SP as a cross-linguistic tendency rather than a true synchronic generalization or principle of language. She proposes to interpret the constraint as a result of "paths of change," saying "Three well-documented universal paths of change occur in parallel and lead to the synchronic situation that is described as Structure Preservation" (Bybee 2008: 114). She also says that it is some sort of restatement of the older structuralist principle of "separation of levels" where phones are distributed by phonetic criteria and phonemes by lexical and morphological ones.[5] In Bybee's view, because SP is an emergent property of recurring mechanisms of language change (that feed and complement each other), counterexamples to this constraint are unavoidable and expected. The thinking behind this is that since the transition from phonetic to phonemic status is gradual, there will always be linguistic sounds that are introduced in a language lexicon with the initial status of phone, either native or foreign, which will later acquire the status of phoneme. However, while some instances make the transition from variant to phoneme, others instances do not, or at least not at the same time. More concretely, Bybee explains that purely phonetic sounds can gradually be disassociated from their phonetic conditioning and become associated with particular lexical or morphological conditions. An example discussed extensively in the structuralist and generativist literature, and already pointed out in this section, is the case of German [x] and [ç] (see Bybee 2008: 112 for a synopsis and Hall 1992 for more detailed discussion). In brief, [ç] and [x] were originally variants, with [ç] occurring after a front vowel in German. When the German diminutive suffix *-ichen* [içən] lost its conditioning front vowel and the shortened suffix *-chen* [çən] started to appear after a back vowel, [x] and [ç] (arguably) became distinctive (e.g. *Kuhchen* [kuːçən] 'little cow' *vs. Kuchen* [kuːxən] 'cake').[6] The distinctiveness of /ç/ *vs.* /x/ was reinforced by the fact that /ç/ could also occur at the beginning of loanwords in some German dialects, where

5 Except that SP avoids the duplication problem that classical phonemics (structuralists) faced. Indeed, in classical phonemics, a generalization had to be stated twice, once at the level of phonemes and once again at the level of phones, because of the separation of levels.

6 However, Macfarland and Pierrehumbert (1991: 171) do not recognize this as a minimal pair because "*Kuchen* is a monomorphemic noun [as opposed to *Kuhchen* 'little cow']." They maintain that there are no true minimal pairs distinguishable only by [ç] *vs.* [x] in German.

the initial phonetic conditioning (the preceding front vowel) is obviously absent.[7] To take another example, this time from English, the non-anterior voiced fricative /ʒ/, which initially occurred in the Early Modern English Period as a result of stress-conditioned palatalization (/zj/ > [ʒ]; e.g. *pleasure* [pleʒəɹ] from French *plaisir*), has begun to be allowed in word-final and even word-initial position under the influence of more recent French borrowings such as *rouge* [ruʒ], *beige* [beʒ], *garage* [gəɹɑʒ], *massage* [məsɑʒ], *camouflage* [kæməflɑʒ], *luge* [luʒ], *genre* [ʒɑ̃r], *joie de vivre* [ʒwɑdeviv], etc. (see Millward 1996: 252–253). Similar to the German situation, the appearance of /ʒ/ in these environments cannot be due to phonetic conditioning. In short, the picture that emerges is that what originated as phonetic variants in the German and English examples might have become phonemes (albeit ones with sometimes restricted distribution) due to, among other things, the pressure of loanwords.

This is one path of change; concurrent with that are two others. The second is that small phonetic changes tend to become larger ones over time, leading to a greater phonetic distance between the original sound and its variant. If we take the [ç] ~ [x] alternation, the variant [ç], which is unstable and phonetically close to /x/, is becoming more stable and more clearly distinct phonetically from /x/ over time (see Bybee 2008: 113). The third related path of change discussed by Bybee is loss of productivity as phonetic processes become lexicalized. To sum up, variants come to occur at the lexical/morphological level because of the diachronic tendency of phonetic changes to become linked to particular lexical items or morphological processes, creating a shift from the purely phonetic to the lexical level. As a result of being linked to particular morphological or lexical conditions, the phonetic conditions that originally give rise to the variant can lose their automatic productive power. Once the link between a sound and its (phonetic) conditioning environment is broken, the sound is "liberated," as it were, and free to enjoy wider phonotactic/syllabic distribution, giving it phonemic as opposed to purely phonetic status (see also Harris 1990: 93). SP has exceptions because such change does not affect the entire vocabulary at once, but rather proceeds via normal processes of lexical diffusion (see e.g. Phillips 2006 on lexical diffusion and its links to various sound-based phenomena).

Is that the end of the story? Kiparsky (2008) clearly disagrees with the diachronic view. He summarizes the situation as follows:

> An increasingly popular research program seeks the causes of typological generalizations in recurrent historical processes, or even claims that all principled explanations for universals reside in diachrony. Structural and generative grammar has more commonly pursued the reverse direction of explanation, which grounds the way language changes in its structural properties. (Kiparsky 2008: 52)

Kiparsky points out (2008: 27) that, once spelled out, historical explanations like those proposed by Neogrammarians or, more recently, those working in the diachronic view (e.g. Bybee 2008) often turn out to appeal implicitly to tendencies that are themselves in need of explanation. In other words, there must be principles

[7] According to Marc van Oostendorp (personal communication), there might remain some sort of phonetic conditioning in loanwords, however, since [ç] can occur only before a front vowel word-initially.

governing the nature and extent of change, which are ultimately responsible for the tendencies that are observed. He proposes criteria to distinguish true universals, which constrain language change, from typological generalizations, which result from language change, and adds: "The issue goes well beyond the simple question how cross-linguistic generalizations originate. It is about the nature of those generalizations themselves" (Kiparsky 2008: 27). He also raises the possibility (2008: 25) that functional explanations for language change might have become biologized within UG itself, through language use, thereby constraining change also via acquisition.

Though the argument we present below does suggest that a guiding principle of grammars is the pressure to preserve structure, which cannot be simply a side-effect of sound change over time, our goal in this chapter is not to argue whether SP is a basic principle of UG, as opposed to an emergent property of converging processes of language change. Rather, we hope to show, using primarily the phonological treatment of loanwords, that distinctive information is, indeed, highly resistant to destruction or alteration at the lexical level, not only intralinguistically, but interlinguistically too. Whatever problems SP has faced, or continues to face, there is no doubt that distinctive phonological information is resistant to change, so some notion of Structure Preservation is still needed, even under current constraint-based approaches, both derivational and non-derivational. The essence of our argument is that if there were no notion of Structure Preservation synchronically active in grammars, we could not explain why a borrower works so hard to preserve distinctive information from a foreign system (L2) in his/her own language (L1). Why should he/she care in the first place?

3 A broader perspective of Structure Preservation

3.1 *Structure Preservation in loanword adaptation*

> Languages . . . which have undergone striking changes in their lexicons through the additions of thousands of borrowed words can no doubt be expected to trouble phonologists for some time. (Kaisse 1990: 141)

Because borrowing normally takes words conforming to the sound patterns and restrictions of one language (the source language, L2) and makes them conform to those of another (the borrowing language, L1), loanwords routinely present the need to modify or destroy phonological information (cf. also CHAPTER 95: LOANWORD PHONOLOGY). A priori, borrowing includes three phenomena that seem to challenge the notion of Structure Preservation. These are the modification of sounds, the deletion of sounds, and, apparently paradoxically, the importation of sounds at the lexical level. Of these three phenomena, deletion and importation are, on first impression, the most problematic. However, as we will see, phoneme deletion seldom occurs and importation is respectful of L2's phonological integrity. We believe this, along with other facts to be discussed, makes loanwords especially relevant to the study of Structure Preservation, provided one accepts an enlargement of its scope. We will henceforth use the full form, "Structure Preservation," to refer to this larger conception of the constraint. SP will refer to the notion of Structure Preservation as defined by, and linked to, Lexical Phonology.

If, instead of seeing the Structure Preservation constraint as just a ban on the introduction of non-phonemic sounds or distinctions at the lexical level, we interpret it as a form of pressure to preserve any contrastive information (features, phonemes, phonemic patterns, unpredictable syllabic information, etc.), then observing the way loanwords are adapted becomes extremely relevant. This is what we propose to discuss here. As we will show in the next sections, especially in §3.2 where the statistics from a large loanword database, that of the CoPho Project,[8] are presented, L2 distinctive information is very seldom squarely destroyed in L1. Instead, L2 distinctive information, when it is not imported, is normally phonologically "adapted" in the borrowing language, with as few adjustments as possible, i.e. minimally. If there were no constraint on synchronic grammars to preserve structure, then there should be no reason for phoneme deletion to be so scarce in loanwords and for adaptations to be minimal.

The notion of minimal adaptation is closely tied to the generally agreed idea that in loanword adaptation, languages normally seek to replace unacceptable foreign sounds with those that are "closest." There is disagreement over how closeness is defined: a matter of some contention in the field of loanword adaptation is whether closeness is determined primarily on phonological grounds, as we maintain, or whether it is determined mainly phonetically (CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). Our present purpose is not to debate this issue, but rather to present statistics from large corpora of loanwords in several languages that indicate that L2 distinctive information is routinely maintained to the maximum allowed by the L1 phonological constraints. In the parlance of the Theory of Constraints and Repair Strategies (TC),[9] this is attributable mainly to two principles, the Preservation Principle and the Minimality Principle, which conspire, we believe, to produce this result. As we will endeavor to show, both principles could be instantiations of the pressure to preserve structure in the larger sense that we propose here.

3.2 *Adaptation, deletion, and Structure Preservation*

The data used to illustrate the effects of preservation in loanwords are taken from the CoPho Project's loanword database, which includes general corpora of French borrowings in Canadian English, Moroccan Arabic, Kinyarwanda, and Lingala, and English borrowings in Calabrese Italian, Japanese, Mexican Spanish, Quebec French, Parisian French, etc. The main findings yielded by the analysis of the CoPho database are summarized in Table 76.1.

The first relevant point to note about the figures in Table 76.1 is that L2 distinctive phonological information is systematically adapted in L1 (34,070/50,092 cases, i.e. 68 percent), as opposed to being deleted (3.3 percent of cases). Phonological adaptation, which is simply called "adaptation" here, is the modification/replacement (i.e. repair) of an L2 sound or structure to comply with one or more L1 phonological constraints. Adaptation is linked to Structure Preservation insofar as it is geared to ensuring that the L1 contrastive system remains unchanged (see

[8] CoPho stands for constraints (Co) in phonology (Pho). The project is supervised by Carole Paradis at Laval University, Quebec City.
[9] TC (previously TCRS) was originally proposed by Paradis (1988).

Table 76.1 The CoPho Project loanword database of phonemic and supraphonemic malformations (updated August 2009)

General information				*Malformations*				
Corpora	*Loans*	*Forms*	*Total*	*Phonological cases*[a]				*Non-phonological cases*
				Total	*Adaptations*[b]	*Importations*	*Deletions*	*Total*
English borrowings in:								
Old Quebec French	485	597	489	398 81.4%	298 74.9%	78 19.6%	22 5.5%	91 18.6%
Parisian French	901	2,576	3,153	2,749 87.2%	1,570 57.1%	987 35.9%	192 7%	404 12.8%
Quebec City French	949	2,416	2,434	2,183 89.7%	1,479 67.7%	602 27.6%	102 4.7%	251 10.3%
Montreal French	949	2,248	2,285	2,099 91.9%	1,262 60.1%	747 35.6%	90 4.3%	186 8.1%
Mexican Spanish I	1,045	1,514	3,137	3,008 95.9%	1,583 52.6%	1,317 43.8%	108 3.6%	129 4.1%
Mexican Spanish II	1,034	2,342	5,645	4,490 79.5%	2,836 63.2%	1,569 34.9%	85 1.9%	1,155 20.5%
Japanese	1,167	2,991	7,760	7,373 95%	6,778 91.9%	492 6.7%	103 1.4%	387 5%
Calabrese Italian	2,161	5,191	14,740	14,438 98%	6,182 42.8%	7,821 54.2%	435 3%	302 2%
French borrowings in:								
Canadian English	674	1,667	1,034	748 72.3%	555 74.2%	137 18.3%	56 7.5%	286 27.7%
Moroccan Arabic	1,127	2,685	4,275	3,979 93.1%	3,104 78%	568 14.3%	307 7.7%	296 6.9%
Kinyar-wanda	756	2,130	4,639	4,207 90.7%	4,119 97.9%	26 0.6%	62 1.5%	432 9.3%
Lingala	672	1,917	3,734	3,408 91.3%	3,396 99.6%	2 0.1%	10 0.3%	326 8.7%
Fula	532	1,081	1,118	1,012 90.5%	908 89.7%	45 4.5%	59 5.8%	106 9.5%
Total for all corpora	12,452	29,355	54,443	50,092 92%	34,070 68%	14,391 28.7%	1,631 3.3%	4,351 8%

[a] Percentages of phonological and non-phonological cases are calculated on the total number of malformations.
[b] Percentages of adaptations, non-adaptations, and deletions are calculated on the total number of phonological cases.

LaCharité and Paradis 2005, Paradis and Tremblay 2009, and Paradis and LaCharité, forthcoming, for discussion), and is the norm.

Table 76.1 also shows that if foreign sounds are not adapted, they are normally imported, i.e. left unadapted, as opposed to being deleted. Importations/non-adaptations account for 14,391/50,092 (28.7 percent) of the phonological cases (importation is discussed in §3.5). Deletions, which might be seen as *prima facie* counterexamples to the idea that structure is preserved in loanword adaptation, are rare in the database overall. Deletions that have been classed as phonological are those that can be explained by the phonological principles of the theory; the others have been classed as non-phonological precisely because they cannot be predicted on phonological grounds. In the phonological cases, the rate is well below 10 percent in any individual corpus and is only 3.3 percent in the corpora overall (1,631/50,092 phonological cases). Non-phonological cases represent only 4,351/54,443 cases (8 percent). Therefore, whether we consider only phonological cases, or we include non-phonological cases (8 percent) as well, deletion of a phoneme is uncommon in the CoPho loanword database. Moreover, not all non-phonological cases involve deletions, and those that do may not always be best explained as such. As shown in Paradis and LaCharité (2008), who address the treatment of non-phonological cases in three corpora of old and recent Quebec French, phoneme deletion is uncommon even in non-phonological cases. Very often, it results from analogy, real or false (e.g. QF [lɪpsɪŋ] instead of [lɪpsɪŋk] for English *lip-sync*; the absence of /k/ in the QF borrowing is not a case of phoneme deletion per se, but rather a case of false analogy to the English verb *to sing*). Lexical truncation, such as QF *tan* for English (*sun*) *tan*, is also sometimes responsible for the disappearance of L2 phonemes (here *sun*); cf. also lexical truncation in QF *parking* from English *parking lot* and French *pull* from English *pullover*. In our view, these lexical truncations should not be seen as phoneme deletion, since deletion does not occur on the basis of phonemes, but of lexical items. Paradis and LaCharité (2008, forthcoming) suggest that these non-phonological processes, along with hypercorrection, phonetic approximations, etc., are responsible for many so-called "divergent repairs" and "unnecessary repairs" (see CHAPTER 95: LOANWORD PHONOLOGY).

We attribute the rarity of deletion to the Preservation Principle in (2) (see e.g. Paradis *et al.* 1994; Paradis and LaCharité 1997).

(2) *Preservation Principle*

Phonemic information is maximally preserved, within the limits of constraint conflicts.

The Preservation Principle is a TC mechanism first proposed by Paradis *et al.* (1994) and used extensively to analyze the CoPho loanword database (see e.g. Paradis and LaCharité 1997).[10] However, we are not the only ones working in loanword adaptation to have seen the need for such constraints; for example, Calabrese (2005)

[10] TC is not restricted to loanwords; we do not maintain that the processes observed in loanwords are independent of general phonology. Nonetheless, it might be that the Preservation Principle is more evident in loanword adaptation than in native phonology, where the influence of morphology and residual historical processes play a greater role.

invokes comparable mechanisms, the Principles of Economy and Last Resort.[11] As previously stated, deletions are *prima facie* violations of (2). However, not only are L2 phoneme deletions rare in the CoPho database, but most are also highly predictable. As we will try to show, deletion is, for the most part, phonologically predictable, and most phonologically predictable deletion can be reconciled with the notion of Structure Preservation.

There are two main scenarios in which phoneme deletion occurs. The first scenario involves deletion of a guttural – sounds characterized by a Pharyngeal node – by languages that do not use this primitive in the representation of the sounds of their native inventories. A language that does not employ this primitive cannot adapt a guttural (see Paradis and LaCharité 2001 for a detailed discussion on the treatment of gutturals in loanwords). For instance, neither French nor Italian have a guttural in their phonemic inventories,[12] so neither language is equipped to adapt a phonemic guttural such as English laryngeal /h/. Instead, they delete it (e.g. English *hamburger* [hæmbəɹgəɹ] yields Quebec French (QF) [_ambəɹgəɹ] and Italian [_amburgər]). The systematicity of guttural deletion is indicated by the figures of the three contemporary QF corpora (see Paradis and LaCharité 2001: 264). There are, overall, 173 cases of /h/ in English loanwords in QF; deletion applies in 163 cases (94.2 percent). The remaining ten cases are importations in the Montreal French corpus. The figures for the Calabrese Italian corpus reinforce this point and further illustrate the fact that guttural deletion accounts for the preponderance of deletions in the CoPho database. In the Calabrese Italian corpus of English loanwords, there are 296 cases of /h/ in the English input. In only 23/296 cases (7.8 percent) is English /h/ imported; the rest of the time it is deleted, meaning that there are 273 /h/-deletions in the Calabrese Italian corpus. Since there are only 278 deletion cases in that corpus overall, this means that /h/-deletion accounts for 98.2 percent of them (273/278). The vast majority of the 3.3 percent of deletions in the CoPho database concern guttural consonants in languages that do not exploit the Pharyngeal node. Such deletions would not be a violation of Structure Preservation from the point of view of L1, because the borrowing languages do not have a native guttural contrast to preserve and are not phonologically equipped to preserve that of L2, as argued in Paradis and LaCharité (2001).

The second general source of phoneme deletion involves the loss of a coda /r/ in borrowings by languages that do not allow (rhotic) codas (CHAPTER 30: THE REPRESENTATION OF RHOTICS).[13] This is the case in Japanese, which allows only N or the first part of a geminate in codas (see Itô 1986 for details of the coda condition in Japanese). Coda consonants in English borrowings in Japanese are systematically adapted by vowel insertion; this has the effect of moving the problematic coda consonant to the onset of the following new syllable (e.g. English

[11] Calabrese's (2005: 20) principles of Economy ("Use the minimal amount of maximally relevant units") and Last Resort ("Use a maximally relevant operation minimally") are highly reminiscent of TC's Preservation and Minimality Principles, and serve the same structure-preserving function.

[12] French [ʁ] and [ʀ] are just two of the numerous variants for the coronal /r/ in French; in contrast with /ʁ/ in Arabic, these sounds are not phonemic in French.

[13] Another predictable, but statistically marginal, source of deletion results from violation of the Threshold Principle (Paradis and LaCharité 1997). This principle offers an explanation for several "atypical" deletion cases, including vowel-initial deletion in French polysyllabic loans introduced in Moroccan Arabic (see Paradis and Béland 2002 for an in-depth discussion of this case and, more generally, Paradis and LaCharité, forthcoming).

optimism [ɑptəmɪzəm] > Japanese [opɯtimizɯmɯ]). However when the coda is a rhotic, instead of having vowel insertion, merger of the rhotic with the preceding vowel occurs (804/804 cases; e.g. English *order* [ɔɹdəɹ] and *corner* [kɔɹnəɹ] > Japanese [oːdaː] and [koːnaː]).[14] Deletion of /r/ also causes vowel lengthening in Thai, where it is not permitted in the coda. For instance, English *care* [kɛɹ], *carbon* [kɑɹbən], *cartoon* [kɑɹtun], *party* [pɑɹti], *poker* [pokəɹ], and *star* [stɑɹ] are pronounced [kʰɛː], [kʰaːbɔn], [gaːtuːn],[15] [paːti], [pokəː], and [sətaː], respectively.[16] Deletion of /r/ is not limited to loanwords that come from English. It also occurs in loanwords from French. For instance, French *Argentine* [aʀʒɑ̃tin], *arrière* [aʀjɛʀ], *beurre* [bœʀ], *carte* [kaʀt], *carton* [kaʀtɔ̃], *orchidée* [ɔʀkide], and *radar* [ʀadaʀ] yield Khmer [aːzɑ̃tin],[17] [aːjaː], [ɓəː], [kaːt], [kaːtɔ̃], [oːkiːɗeː], and [raːɗaː], respectively. In some cases, the French rhotic is replaced with /a/ or a glottal stop, as in [oːpaːlua] and [pɛʔmiː] from French *haut-parleur* [oparlœʀ] and *permis* [pɛʀmi]. As can be seen, the deletion of /r/, which is prohibited in coda position in Khmer, causes vowel lengthening even in closed syllables, as in [kaːt] from French *carte*.[18] This latter set of examples in Khmer shows that apparent /r/-deletion is not influenced by the pronunciation of /r/ in the donor language, since French and English have very different rhotics. Vowel lengthening suggests that /r/ might not really be deleted but rather fused with the preceding vowel, when it is not replaced with /a/ or a glottal stop. This is why we have not incorporated these cases in the deletion column of the statistics in Table 76.1. If we are correct in viewing vowel lengthening as /r/-adaptation rather than /r/-deletion, it does not contradict the idea of Structure Preservation that is invoked here. However, even if we did consider these cases of deletion, the deletion rate would still remain very low (2,435/50,091 – 4.9 percent instead of 3.3 percent).

In Paradis and LaCharité (forthcoming), we attribute /r/-deletion to the fact that /r/ is vowel-like and can easily be fused with the preceding vowel, whether it results in vowel lengthening or not. We envision that, perhaps, as in the case of /h/-deletion in English loanwords in French, Italian, Portugese, etc., the answer lies in the phonological structure of /r/. Rhotics with a variety of phonetic realizations are prone to deletion, cross-linguistically, and they exhibit several phonological behaviors that are not yet well understood. For example, in many different languages, where the rhotics exhibit diverse phonetic realizations, a coda /r/ is deleted, or merged with, transformed into, or replaced by, a vowel. To cite just a few of many possible examples, in German, where /r/ is phonetically uvular, coda /r/ can lower to something akin to a low vowel, so that *tür* 'door' is realized as [tyːɐ] in the singular (Wiese 1996) but *türen* [tyrən] in the plural, that is with the full rhotic, where it is in onset position. In Quebec French, coda /r/, which can be realized as a uvular or a coronal, is often deleted word-finally in informal speech (e.g. *bonjour* /bɔ̃ʒur/ 'good day' → [bɔ̃ʒu(ː)]). During the Middle Ages, /r/-deletion prevailed for so long in French that /r/ almost disappeared

[14] Tones are omitted here, because they are irrelevant.

[15] [g] is a variant of unaspirated /k/ in Thai.

[16] Data gathered during fieldwork in Thailand in February and March 2010.

[17] Even though French is no longer spoken by young people in Cambodia and Laos, Lao and Khmer speakers, both young and old, almost always import the French nasal vowels in French loanwords, which are very numerous in both languages.

[18] Data gathered during fieldwork in Cambodia in March 2010.

as a coda phoneme (Zinc 1986). The deletion of /r/ also applies in many Spanish dialects (Moreno de Alba 1988; Rojas 1988; e.g. *mar* 'sea' → [ma] in Caribbean Spanish and *cliquear* from English *to click*, which is realized as [klikea] in Spanglish). Interestingly, English *short* is realized [ʧo] in Spanglish when it is singular but [ʧores] in the plural, i.e. with the full rhotic when it is no longer in coda position, thus indicating that the rhotic is present in the L1 lexical representation of the borrowing. Examples of this type are common cross-linguistically. It could be that a post-vocalic rhotic is actually part of a diphthong, as proposed by several phonologists (e.g. see Nikiema and Bhatt 2003 for their analysis of post-vocalic /r/-deletion in Haitian Creole).

Nonetheless, we know there are some languages where coda /r/ is ill-formed, but its deletion does not yield vowel lengthening. Although the general CoPho loanword database does not include such languages, we would view these cases as true deletions and, as such, challenges to the idea that structure is preserved in loanword adaptation. Our targeted corpus on aspiration in Mandarin Chinese (MC) (see also Hall-Lew 2002 on this) shows that this is what happens in English loans in Mandarin Chinese.[19] The English coda rhotic, which is disallowed in Mandarin Chinese, is dropped without yielding systematic vowel lengthening (e.g. English *laser* [lezəɹ], *cigar* [sɪgɑɹ], *cartoon* [kɑɹtun], and *sardine* [sɑɹdin], which yield MC [lej ʂə_], [sjə tsja_], [kʰa tʰuŋ] and [ʂa_ tiŋ], respectively). The coda rhotic is not adapted in /l/, as in onset position (e.g. English *radar* [ɹedɑɹ] and *trust* [tɹʌst] > MC [ləj ta] and [tʰ(w)ə las]). The net result with respect to a discussion of Structure Preservation is that one must consider /r/-deletion/fusion to be very common across languages in both native and borrowed words, and that there seems to be a phonological explanation for many such cases that avoids conflicts with Structure Preservation.

However, even when /r/-deletion does not lead to lengthening, it may not be a problem for Structure Preservation from the point of view of loanword adaptation, because in many cases such deletions stem from a native process. For example, it is common for rhotics to be deleted when they are included in a complex onset (CHAPTER 55: ONSETS). For instance, Quebec French *trois* [tʀwa] 'three', *Lacroix* [lakʀwa] (a proper name), and *fruit* 'fruit' [fʀɥi] are often pronounced [twa], [lakwa], and [fɥi] in casual speech. In Thai, /r/ in complex onsets is only pronounced in very formal speech (on television, for instance). In less formal/casual speech it might be replaced with /l/, but most of the time the liquid disappears altogether. For instance, the famous shopping center of Bangkok, *Maboonkrong*, is pronounced with the rhotic only in very formal speech. Otherwise, it is pronounced [maːbuːŋkʰɔŋ], with no rhotic; this is what taxi drivers say, with [maːbuːŋkʰlɔŋ], a more prestigious pronunciation with the lateral, being used much less frequently. The same happens with the Thai city *Trat*, which is systematically pronounced [tat]; cf. also Thai [pʰroːm] 'carpet', which is pronounced [pʰoːm] except in very formal speech (this information on /r/-deletion, as well as on /l/-deletion, is readily available in any grammar of Thai). Native /r/-deletion in complex onsets is frequent in Asian languages, so it is not surprising to see /r/-deletion apply to

[19] A targeted corpus, as opposed to a general one, is a (normally smaller) corpus of loanwords collected to test a particular hypothesis (e.g. aspiration in MC, Hindi, Thai, and Lao, palatalization in Russian, etc.). Therefore, all borrowings in a targeted corpus contain a particular sound or contrast of interest to the hypothesis being tested.

their loanwords (e.g. French *crème* [kʀɛm] 'cream' which yields [kem] in Vietnamese, according to CHAPTER 95: LOANWORD PHONOLOGY; cf. also English *credit card* [kɹɛdɪt kɑɹd] and *brake* [bɹek], which are pronounced [kʰ(l)eːdit gaːt] and [b(l)ɛːrk] in Thai, respectively, and French *programme* [pʀɔgʀam] 'program' and *groupe* 'group' [gʀup], which yield Lao [pog(l)am] and [g(l)up], respectively). Rhotic deletion in cases such as English loanwords in Thai should not be interpreted as a repair of an ill-formed L2 structure or as an "unnecessary loss," since it stems from a very productive native process related to speech register/dialectal differences.

To sum up this discussion, the vast majority of deletions are phonologically predictable and thus, despite initial impressions, pose little threat to the idea that input structure is preserved in loanword adaptation. However, the real story is that, together, predictable and unpredictable deletion affect only a small percentage (less than 5 percent) of input phonemes in the CoPho loanword database. We conclude from this that the loss of L2 phonemes is strongly avoided in loanword adaptation.[20] In the case of ill-formed sounds, feature adjustments apply systematically; in the case of ill-formed clusters, that are perceived as unsyllabifiable by L1, phoneme insertion is the norm. For instance, French *drapeau* [dʀapo] yields [darapo] in Fula, not *[dapo] or *[rapo] (see also French *force* [fɔʀs] > Fula [fɔrsɔ], not *[fɔs] or *[fɔr], and French *ministre* [ministʀ] > Kinyarwanda [ɪninisitiri], not *[mini]). This pattern consistently predominates in the general corpora of the CoPho loanword database, as well as in more recently assembled targeted corpora such as the Kashmiri one. When an English borrowing contains a cluster that is disallowed in Kashmiri, the sequence undergoes vowel insertion, not consonant deletion, despite the fact that consonant deletion would solve the problem equally well. For instance, English *silk* [sɪlk], *snow* [sno], and *flag* [flæg] result in Kashmiri [silik], [sono], and [fəlag], and not in *[sik], *[so/no], or *[fag], for example.[21]

Why does L1 resort to phoneme insertion instead of phoneme deletion when it has to handle a problematic L2 cluster? We attribute this to the Preservation Principle in (2), which seeks to safeguard contrastive information, and which can be seen as a constraint in the Theory of Constraints and Repair Strategies to preserve structure. However, TC is a derivational constraint-based theory. One might immediately wonder whether Optimality Theory (OT), a non-derivational (non-serial) filter-based theory, can dispense with the need for SP. The crux of the issue is that standard OT posits that, underlyingly, anything goes (cf. Richness of the Base, following Prince and Smolensky 1993). The patterns that emerge from the lexicon are the result of universal surface filters, which are ranked on a language-specific basis. In short, the basic architecture and tenets of classical OT, with constraints acting as filters, suggest that there should be no particular underlying phoneme or structure inventory to protect. Itô and Mester (2001: 265) examine the possibility that some of the devices and principles of Lexical Phonology might have outlived their usefulness and have no place in a putatively non-serial framework such as OT. Is SP one such device? Itô and Mester argue for recognizing, within OT, the need for stratal organization, with lexical outputs

[20] Phoneme deletion outside the context of malformations, that is when the phoneme and structure that contains it are both permissible in L1, is also very rare (see Paradis and Prunet 2000). As shown by Paradis and LaCharité (2008) and Paradis and LaCharité (forthcoming), these rare cases result mostly from analogy, morphological truncation, phonetic approximation, and hypercorrection.

[21] Data gathered during fieldwork in North India in April 2009.

being structure-preserving, which they define (2001: 289) as "limitation to a restricted inventory of elements and structures. . . ." Bermúdez-Otero and McMahon (2006) work within the framework of stratal OT and maintain, *contra* Itô and Mester, that ". . . the issue of Structure Preservation does not arise in Stratal OT . . ." (2006: 396). However, even if one agrees with that, and rejects Itô and Mester's point of view, OT analyses still rely on some notion of preservation, in the form of faithfulness constraints, which occupy a high-ranked – though not necessarily undominated – place in most OT analyses (CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). All this clearly suggests that OT requires some notion of contrast preservation, an issue that some OT analyses have confronted directly (e.g. Krämer 2006). Our goal here is to point out that no current constraint-based or filter-based theory completely does away with the need for some notion closely related to Structure Preservation. Indeed, it seems more likely that, for all phonologists, it will be important to reconsider the idea of Structure Preservation and to determine its mandate in the context of our particular theories. The remainder of the discussion is framed in the TC model, because Structure Preservation has been more directly addressed in this framework, but we assume that all phonological theories need to confront the same observations concerning what appears to be preserved, as evidenced in loanword adaptation. In other words, our focus will be on the facts, not the theory used to handle them.

The preceding discussion has shown that L2 phonemes are adapted rather than deleted, that the repair of illicit clusters via epenthesis is preferred over their repair via deletion, and that when deletion does occur, it is largely predictable on phonological grounds. The study of loanword adaptation reveals a further implication of Structure Preservation: the violation of L1 constraints is generally solved with as little loss of phonological information as possible. Thus, an ill-formed L2 phoneme is not deleted if a feature can be added or deleted to solve the problem; an ill-formed syllabic structure is not deleted if insertion of a phoneme or, in the case of a constraint conflict, the loss of a phoneme will suffice, etc. Another key observation in loanword adaptation (we see this as another side-effect of Structure Preservation), for which any theory must account, is the limited range of adaptations that predominate cross-linguistically. This issue is addressed in LaCharité and Paradis (2005). For example, English /æ/ is systematically adapted as /a/, not as /i/, /e/, /o/, or /u/, in the CoPho database.[22] In Mexican Spanish, adaptation of */æ/ to /a/ occurs in 354/360 cases (98.33 percent); in French it occurs in 1,405/1,405 cases (100 percent), in Japanese it occurs in 536/536 (100 percent) cases, and in Calabrese Italian 1,121/1,214 cases (92.3 percent). As another example, English /ɪ/ is predictably adapted as /i/. In Mexican Spanish, */ɪ/ adapts to /i/ in 387/388 adaptation cases (99.7 percent); in Japanese, adaptation of */ɪ/ to /i (iː)/ occurs in 631/649 adaptation cases (97.2 percent); in Calabrese Italian, this adaptation occurs in 1,588/1,588 adaptations (100 percent). Even when more than one adaptation for a given sound is attested, either cross-linguistically or within a single language, the range of results is small and predictable. For

[22] In Quebec French, there are cases where English /æ/ surfaces as [ɛ] in loans such as *band*, *gang*, and *pantry*. We believe that this is because these words are often pronounced with the variant [ɛ] in English (e.g. [bɛnd], [gɛŋ], and [pɛntɹi]). In these cases, we say that the English variant is imported. It sounds more "anglophone" i.e. more "in," to pronounce these words with [ɛ], although they can be pronounced with [a] too.

example, English /v/ is adapted as /b/, /f/, or /w/ cross-linguistically. English /ʌ/ is adapted as either /a/ or /o(ɔ)/. Why should a borrowing language not simply replace any illicit sounds arbitrarily, or with default/high-frequency sounds, if there were no pressure to remain close to the input? Even though particular sounds are illicit from the point of view of the borrowing language, as much as possible is salvaged or, conversely, as little as possible is lost. For instance, in the common cross-linguistic adaptation of */v/ to /b/, only the continuant value changes; in the adaptation to /f/, which is also found cross-linguistically in loanword adaptation, only the voicing value is modified, whereas in the adaptation to /w/, a slightly less frequent but nonetheless common adaptation, it is the sonorant value which is targeted. The adaptation of /v/ to /w/ is systematic in Fula (French *civil* [sivil] 'civilian' > Fula [siwil]; Paradis and LaCharité 1997) and in several Asian languages, including Thai in word-initial and intervocalic positions (e.g. English *vitamin* [vajtəmɪn]/[vɪtəmɪn] and *travel agent* [tɹævəl edʒənt] > Thai [wittamin] and [t(r)awəl ejɛn], respectively; word-finally it is adapted as /p/ for phonotactic reasons; e.g. English *serve* [səɹv] > Thai [səːp]). Within the context of TC, this has been attributed to the Minimality Principle in (3).

(3) *Minimality Principle*

a. A repair strategy must apply at the lowest phonological level to which the violated constraint refers.
b. Repair must involve as few strategies (steps) as possible.

The lowest phonological level referred to in (3a) is determined by the phonological level hierarchy (metrical level > syllabic level > skeletal level > root node > feature), an independently required organization of phonological information. Clearly, the Minimality Principle (whose effects are addressed in Paradis and LaCharité 1997) is intrinsically related to the notion of preservation. If preservation were not an issue, then why should repair not often, or even routinely, operate at a higher-than-needed level, guided by some notion of "better safe than sorry"?

3.3 Preservation of L2 phonemic contrast patterns in L1: English loanwords in Chinese and Hindi

Not only are individual L2 phonemes conserved, to the greatest extent possible within the limits allowed by the L1, but L2 phonemic contrast patterns are also maintained to the greatest extent possible permitted by the phonology of L1. For example, Chinese does not have a voicing distinction among stops; it does, however, distinguish stops on the basis of another laryngeal feature, aspiration. In the adaptation of English loanwords in Mandarin Chinese (MC), English voiceless stops (/p t k/) systematically yield aspirated voiceless stops (/p^h t^h k^h/), and English voiced stops (/b d g/) are systematically replaced by unaspirated voiceless ones (/p t k/; see Paradis and Tremblay 2009 for an in-depth discussion of this issue, with figures and statistics). For example, English *pizza* [pitsə], *hippies* [hɪpiz], and *tank* [tæɪŋk] yield MC [p^hi sa], [si p^hiʂ], and [t^han k^hɤ], respectively, whereas English *Boeing* [boɪŋ], *radar* [ɹedɑɹ], and *golf* [gɑlf] are adapted as MC [pəin], [ləj ta], and [kaw ər fu]. This pattern of adaptation is not restricted to English loans; it also applies to French loans in MC (e.g. French *Pierre Cardin* [pjɛʀkaʀdɛ̃] and *Chirac* [ʃiʀak] >

MC [pʰi ər kʰa tan] and [si la kʰə]), despite the fact that voiceless stops are not aspirated in French as they are in English before a stressed vowel. The same type of pattern transfer is found in other Chinese dialects, such as Cantonese. This indicates that Chinese borrowers are aware of the systematic distinction between voiced and voiceless stops in English, and that adaptation seeks to preserve this L2 distinctive pattern, using the contrastive resources provided by the L1.

Comparable facts are found in Hindi. Hindi has a voicing distinction for stops. Thus English voiced and voiceless stops yield Hindi voiced and voiceless stops, respectively (e.g. English *bellboy* [bɛlbɔj], *baggage* [bægədʒ], *coffee* [kɑfi], and *frock* [frɑk] > Hindi [bɛlbɔj], [bagedʒ], [kɔfii], and [frɔk]). However, English voiced and voiceless alveolar stops are adapted as retroflex stops /ʈ/ and /ɖ/, which contrast with dental stops /t/ and /d/ in Hindi (e.g. English *agreement* [əgɹimənt], *beauty parlor* [bjuti pɑɹləɹ], *badminton* [bædmɪntən], and *baking powder* [bekɪŋ pawdəɹ] > Hindi [agrimənʈ], [bwuʈi pɑɽləɽ], [baɖminʈən], and [bekɪŋ pawɖəɽ], respectively), while English interdentals /θ/ and /ð/ are adapted as plain dental stops, that is /t/ and /d/, respectively (e.g. *Thatcher* [θætʃəɹ] and *brother* [bɹʌðəɹ] yield Hindi [tatʃər] and [brɔdər]). Again, the L2 contrast pattern is preserved in L1, using the contrastive resources provided by the latter. Adaptation of interdentals to fricatives would yield a greater loss of information because Hindi has only /s/, not /z/ (except in borrowings, especially from Arabic). The voicing contrast of English interdentals would then be lost. On the other hand, if English alveolar stops were adapted as phonetically more expected dental /t/ and /d/ in Hindi, there would not be any slot left for the adaptation of the interdentals, which would have to merge with the English alveolar stops in Hindi.

3.4 Preservation of L2 syllabic contrasts in L1: French loanwords in Russian

The adaptation of French loanwords in Russian suggests that unpredictable syllabic structure might also be preserved in loanword adaptation. French diphthongs have to be marked underlyingly, as they are unpredictable. Pairs such as *oiseau* [waso] 'bird' *vs. watt* [wat] 'watt' show this. In *oiseau*, *wa* is a diphthong (e.g. *l'oiseau* [lwazo] 'the bird'), making the word vowel-initial, whereas in *watt* it is an onset–nucleus sequence, as it is in English, making *watt* consonant-initial (e.g. *le watt* [lə wat]; see Kaye and Lowenstamm 1984 on diphthongs in French). The presence or absence of an onset is shown by, among other things, the choice of singular and plural definite articles. Before vowel-initial words the singular definite article is [l], as with *l'arbre* [laʀbʀ] 'the tree'. Moreover, the plural definite article triggers liaison (*les oiseaux* [lɛ zwazo] 'the birds', as with *les arbres* [lɛ zaʀbʀ] 'the trees'). Preceding a consonant, the definite articles are *le* [lə] and *les* [lɛ], respectively (*le watt* [lə wat] 'the watt', not *[lwat], and *les watts* [lɛ wat] 'the watts', not *[lɛ zwat], as with *le bateau* [lə bato] 'the boat' and *les bateaux* [lɛ bato] 'the boats').

In French loanwords in Russian, /wa/ is adapted as a bisyllabic sequence of /u+a/ when it is part of a diphthong (e.g. French *voile* [vwal] 'veil', *mémoire* [memwaʀ] 'memory', and *couloir* [kulwaʀ] 'corridor' > Russian [vual], [mʲemuarʲi], and [kuluarʲi]), whereas when /w/ constitutes an onset, it is systematically adapted as /v/ (French *watt* [wat] > Russian [vat];[23] English *whisky* [wɪski] and

[23] This loan was introduced via French, even though it originates from English.

tramway [træmwej] > Russian [vʲiskʲɪ]) and [tramvaj]). These examples might suggest that the difference in adaptation is due to the fact that /wa/ in French loans is preceded by a consonant, whereas in English loans it is not. However, English borrowings such as *sweater* [swɛtəɹ], *swap* [swɑp], and *swing* [swɪŋ], which yield Russian [svʲitɛr], [svop], and [svʲiŋ],[24] not *[suʲitɛr] or *[suatɛr], etc., invalidate this hypothesis. The fact that /wV/ is treated differently when it is a diphthong than when it is an onset–nucleus sequence is interesting, because it suggests that where syllabic affiliation is unpredictable – when it is contrastive and would have to be indicated underlyingly – it is preserved. This interesting question remains to be investigated more thoroughly.

3.5 *Importation and Structure Preservation*

Non-adaptations in loanwords – that is the importation of foreign phonemes in words borrowed from another language (L2) – present a challenge to SP since they consist in the introduction of new phonemes at the lexical level. Two cases that were previously mentioned are /ç/ in German and /ʒ/ in English. One might object that the German case as presented here is oversimplified and does not present an uncontroversial picture of the facts (consider, for example, the contradiction between Hall's 1989 position and the scenario advanced by Bybee 2008), or that, in English, the phonemic status of /ʒ/ is not well established, given that its distribution is restricted to intervocalic position, except in loanwords (see Iverson and Salmons 2005: 210 on /ʒ/ in English). However, the German case seems to be problematic for SP no matter which view one takes; either a sound/sound distinction ([ç] *vs.* [x]) that does not exist at the underlying level is introduced at the lexical level (Hall's view) or non-native /ç/ has become phonemic in German over time, under the influence of loanwords (Bybee's view). As for the voiced palatal fricative in English, even if /ʒ/ were validly considered a phonetic variant intervocally in native English words, the fact that it is tolerated (unadapted) at the end and now at the beginning of borrowings indicates that it is a phoneme in English, though a marginal or peripheral one, in the terminology of Itô and Mester (1995).

The challenge goes beyond German and English: the literature on loanwords reports abundant cases of importation (see e.g. Ulrich 1997: 432 on the importation of an English coda palatal in Lama, and Mohanan and Mohanan 2003 on the importation of English /f/ in Malayalee English). In the case of particular phonemes, importation can even be the norm. In the Moroccan Arabic corpus of the Project CoPho loanword database, /p/ is widely imported (320/454 cases, 70.5 percent) (e.g. French *pape* [pap] 'pope' > Moroccan Arabic [pap] instead of expected [bab]). Another example is /ʃ/ in the CoPho corpus of English loanwords in Mexican Spanish, which is imported in 102/138 cases (74 percent) (e.g. English *shorts* [ʃoɹts] and *carwash* [kaɹwaʃ] > Mexican Spanish [ʃɔɹts] and [karwɔʃ], not [ʧɔrts] and [karwɔʧ], as expected). While some foreign sounds are only occasionally, or never, left unadapted, others are imported more often than they are adapted. In some language situations, such as Spanish loanwords in Guarani, importations from Spanish are systematic, i.e. Spanish phonemes are never adapted (see Oñederra 2009 for a similar situation with Spanish loans in Basque).

[24] [svɪŋk] also exists as a variant; it is perceived by some Russian speakers as more "English," possibly because of hypercorrection.

We must remember, though, that in phonological situations involving language contact, including loanword adaptation, two languages are in play. Our hypothesis is that under certain sociolinguistic conditions, such as when borrowers are highly bilingual and society generally tolerant of importations (such as when the L2 enjoys widespread prestige), the preservation of the L2 system also becomes an issue, one that can be at odds with the preservation of the L1 system. In cases of adaptation, preservation of the L2 system becomes subordinated to the preservation of the L1 system; in cases of importation, the reverse occurs. This, and the fact that loanword adaptations are generally minimal, supports the view that the structural integrity of L2 is rarely left out of consideration altogether, so it is not unlikely that such concern may sometimes come to predominate. If this interpretation is correct, then identification and preservation of contrastive information is important in both L1 and L2, despite inevitable conflicts between the demands of each of the two linguistic codes.

4 Conclusion

As shown in §2, SP, as referred to in Lexical Phonology, was regularly challenged by the facts, even though most phonologists agree that it plays some role, i.e. nonphonemic sounds are generally not generated in the lexicon. Languages tend to preserve their phonemic integrity at this level. Nonetheless, the numerous exceptions to SP reported by many different authors, providing evidence from numerous different languages, might give the impression that Structure Preservation is either misguided or just an artifact of some other principles/processes with no intrinsic validity. There remains little doubt these days that SP as conceived in Kiparsky (1982, 1985) is too restrictive, not to mention its being linked to a network of other assumptions and principles that have themselves been seriously challenged. In fact, even the notions of phonological rules and their application has been subjected to a major rethinking. SP limited the power of phonological rule application, but modern frameworks eschew rules in favor of constraints; if rules have not been abandoned altogether, they have certainly lost their driving force. In a derivational constraint-based theory, such as TC, rules are context-free and functionally motivated, being limited to repairing constraint violations. Thus, their power is intrinsically more circumscribed than was the case of *SPE*-type rules that were, in and of themselves, the motivation for phonological change (i.e. they were essentially descriptive devices with little or no explanatory power). In a filter-based theory, such as OT, filters, including any that favor the preservation of input structure or contrasts (i.e. faithfulness filters), are ranked on a language-specific basis. Given that feature of the theory, it is not obvious how OT would deal with a cross-linguistic tendency to preserve input structure/contrasts (see Paradis 1996 on this issue). If the faithfulness constraint Max-X (which prevents deletion of a phoneme or feature that is in the input; previously Parse) is shown to play a consistently high-ranked (though, as already mentioned, not necessarily undominated) role in OT analyses, then OT too might need to appeal to a mechanism that accounts for Structure Preservation. Therefore, the real question is: does Structure Preservation have any kind of intrinsic validity for phonological theory? We maintain that it does.

In §3, we used the adaptation of loanwords to underline the continued need for some notion of Structure Preservation. However, we see this principle as having a broader scope than that defined for SP in Lexical Phonology. Not only do languages tend to preserve their own phonemic inventories at the lexical level as much as possible (in the spirit of SP in Lexical Phonology), but they also tend to maximally preserve the phonemic contrasts and contrast patterns of the languages from which they borrow words. Thus, the resistance to change is, above all, a question of contrast/category pattern preservation, which is expressed interlinguistically (i.e. between L2 and L1, as was illustrated in this chapter with the treatment of loanwords), as well as intralinguistically (as was illustrated with the [ɔ/o] alternation case in French and Velar Softening in the English and French cases). As mentioned at the outset of the chapter, it is not just a question of inertia. Speakers work hard to preserve L1 or L2 phonological patterns.[25]

If there were no (universal) pressure to preserve an input's contrastive information, then why would deletion be so rare in loanword adaptation? Why would it not occur randomly in some 50 percent of the cases? Moreover, when deletion does occur, why is it so largely predictable on the basis of phonology? Among adaptations, why are the changes to ill-formed sounds and structures so consistently predictable in terms of minimality, and why is there such a limited range of adaptations found cross-linguistically? This is because distinctive information is as resilient and resistant to change in L1 as it is in L2 in the mind of borrowers. When L2 wins, the result is an importation (a non-adaptation), i.e. the introduction of a new phoneme or structure in L1, as discussed in §3.5. Extensive language contact is required for this to happen, though. When L1 wins, which is more generally the case in the first stages of borrowing, we obtain an adaptation, whose goal is to produce a form that meets the phonological demands of the borrowing language's phonology. This means that some L2 contrastive information will inevitably, though minimally, be sacrificed, because the preservation of the L2 contrastive information is often at odds with preservation of the contrasts of the L1 phonological system. However, in focusing on phoneme modification (i.e. adaptation), we risk undervaluing the fact that, to the greatest extent possible, an adaptation retains most properties of the source form. In this chapter, the properties referred to have included distinctive phonemic information, as illustrated with loanwords from French and English in many different languages (Japanese, Khmer, Thai, Fula, Kinyarwanda, Kashmiri, etc.), phonemic contrast patterns from English loans in Chinese and Hindi, and syllabic contrast pattern from French loans in Russian. This is not intended to be exhaustive though. Other types of contrastive information are expected to show similar resilience. In this chapter, we have tried to emphasize that contrast resilience extends to L2; it is not limited to L1. L1 adapters feel strongly concerned about preserving L2 contrastive information; in the case of importations this is to the detriment of their own (L1) contrast system, which is forced to change.

Ultimately, we suggest that what might salvage SP, after all, is to consider it in a much broader perspective in order to deepen our understanding of its purpose and functioning. It will then be easier to circumscribe its effects in native and borrowed words and formulate it more formally, even if it is in terms of a statistically significant tendency instead of an absolute generalization.

[25] Structure Preservation obviously does not have the same impact in L2 acquisition as in loanword adaptation. Its influence is necessarily reduced in L2 acquisition, since L2 learners (especially beginners) are not as knowledgeable about the L2 code as are the borrowers (see Paradis and LaCharité 1997 on the borrowers' bilingualism issue) and thus cannot be as protective of a code with which they are not sufficiently acquainted.

ACKNOWLEDGMENTS

We would like to thank K. P. Mohanan for his comments on a draft of this chapter and for several discussions about Structure Preservation. We are grateful to Joan Bybee for sending comments on this notion, along with a scanned version of a relevant document. We also benefited from discussions with Zhiming Bao and Haruo Kubozono on phoneme preservation/deletion in loanwords. Finally, we would like to thank the editors, in particular Marc van Oostendorp and Keren Rice, and the reviewers for their comments. We remain solely responsible for the views expressed here as well as for any remaining errors or omissions. Research for this chapter was made possible by SSHRCC grant #410-2008-1128 to the fiirst author and by SSHRCC grant #410-2007-0566 to the second author.

REFERENCES

Bermúdez-Otero, Ricardo & April McMahon. 2006. English phonology and morphology. In Bas Aarts & April McMahon (eds.) *The handbook of English linguistics*, 382–410. Cambridge, MA & Oxford: Blackwell.

Borowsky, Toni. 1986. Topics in the lexical phonology of English. Ph.D. dissertation, University of Massachusetts, Amherst. Published 1990, New York: Garland.

Borowsky, Toni. 1989. Structure preservation and the syllable coda in English. *Natural Language and Linguistic Theory* 7. 145–166.

Bybee, Joan. 2008. Formal universals as emergent phenomena: The origins of structure preservation. In Good (2008), 108–121.

Calabrese, Andrea. 2005. *Markedness and economy in a derivational model of phonology*. Berlin & New York: Mouton de Gruyter.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.

Good, Jeff (ed.) 2008. *Linguistic universals and language change*. Oxford: Oxford University Press.

Hall, T. A. 1989. Lexical Phonology and the distribution of German [ç] and [x]. *Phonology* 6. 1–17.

Hall, T. A. 1992. *Syllable structure and syllable-related processes in German*. Tübingen: Niemeyer.

Hall-Lew, Lauren A. 2002. English loanwords in Mandarin Chinese. B.A. thesis, University of Arizona.

Hargus, Sharon & Ellen M. Kaisse (eds.) 1993. *Studies in Lexical Phonology*. San Diego: Academic Press.

Harris, John. 1987. Non-structure-preserving rules in Lexical Phonology: Southeastern Bantu Harmony. *Lingua* 73. 255–292.

Harris, John. 1989. Toward a lexical analysis of sound change in progress. *Journal of Linguistics* 25. 35–56.

Harris, John. 1990. Derived phonological contrasts. In Susan Ramsaran (ed.) *Studies in the pronunciation of English: A commemorative volume in honour of A. C. Gimson*, 87–105. London: Routledge.

Hyman, Larry M. 1993. Structure preservation and postlexical tonology in Dagbani. In Hargus & Kaisse (1993), 235–254.

Inkelas, Sharon & Draga Zec (eds.) 1990. *The phonology–syntax connection*. Chicago: University of Chicago Press.

Itô, Junko. 1986. Syllable theory in prosodic phonology. Ph.D. dissertation, University of Massachusetts, Amherst. Published 1988, New York: Garland.

Itô, Junko & Armin Mester. 1995. Japanese phonology. In Goldsmith (1995), 817–838.

Itô, Junko & Armin Mester. 2001. Structure Preservation and stratal opacity in German. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 261–295. Cambridge: Cambridge University Press.

Iverson, Gregory K. 1993. (Post) lexical rule application. In Hargus & Kaisse (1993), 255–275.

Iverson, Gregory K. & Joseph C. Salmons. 2005. Filling the gap: English tense vowel plus final /s#/. *Journal of English Linguistics* 33. 207–221.

Kaisse, Ellen M. 1990. Towards a typology of postlexical rules. In Inkelas & Zec (1990), 127–143.

Kaye, Jonathan & Jean Lowenstamm. 1984. De la syllabicité. In François Dell, Daniel Hirst & Jean-Roger Vergnaud (eds.) *Forme sonore du langage: Structure des representations en phonologie*, 123–159. Paris: Hermann.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Kiparsky, Paul. 2008. Universals constrain change; change results in typological generalizations. In Good (2008), 23–53.

Krämer, Martin. 2006. The emergence of the comparatively unmarked. *Proceedings of the West Coast Conference on Formal Linguistics* 25. 236–244.

LaCharité, Darlene & Carole Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry* 36. 223–258.

Macfarland, Talke & Janet Pierrehumbert. 1991. On ich-Laut, ach-Laut and Structure Preservation. *Phonology* 8. 171–180.

Millward, C. M. 1996. *A biography of the English language*. 2nd edn. Fort Worth: Harcourt Brace.

Mohanan, K. P. 1986. *The theory of Lexical Phonology*. Dordrecht: Reidel.

Mohanan, K. P. 1995. The organization of the grammar. In Goldsmith (1995), 24–69.

Mohanan, K. P. & Tara Mohanan. 1984. Lexical phonology of the consonant system in Malayalam. *Linguistic Inquiry* 15. 575–602.

Mohanan, Tara. 1989. Syllable structure in Malayalam. *Linguistic Inquiry* 20. 589–625.

Mohanan, Tara & K. P. Mohanan. 2003. Towards a theory of constraints in OT: Emergence of the not-so-unmarked in Malayalee English. Unpublished ms., National University of Singapore (ROA-601).

Moreno de Alba, José G. 1988. *El español en América*. Mexico City: Fondo de Cultura Económica.

Nikiema, Emmanuel & Parth Bhatt. 2003. Two types of R deletion in Haitian Creole. In Ingo Plag (ed.) *The phonology and morphology of creole languages*, 43–69. Tübingen: Niemeyer.

Oñederra, Miren Lourdes. 2009. Early bilingualism as a source of morphophonological rules for the adaptation of loanwords: Spanish loanwords in Basque. In Wetzels & Calabrese (2009), 193–210.

Paradis, Carole. 1988. On constraints and repair strategies. *The Linguistic Review* 6. 71–97.

Paradis, Carole. 1996. The inadequacy of faithfulness and filters in loanword adaptation. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, 509–534. Salford: ESRI.

Paradis, Carole & Renée Bèland. 2002. Syllabic constraints and constraint conflicts in loanword adaptations, aphasic speech and children's errors. In Jacques Durand & Bernard Laks (eds.) *Phonetics, phonology and cognition*, 191–225. Oxford: Oxford University Press.

Paradis, Carole & Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33. 379–430.

Paradis, Carole & Darlene LaCharitè. 2001. Guttural deletion in loanwords. *Phonology* 18. 255–300.

Paradis, Carole & Darllene LaCharité. 2008. Apparent phonetic approximation: English loanwords in old Quebec French. *Journal of Linguistics* 44. 87–128.

Paradis, Carole & Darlene LaCharité. Forthcoming. Loanword adaptation: From lessons learned to findings. In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.) *Handbook of phonological theory*. 2nd edn. Malden, MA & Oxford: Wiley-Blackwell.

Paradis, Carole & Jean-François Prunet. 2000. Nasal vowels as two segments: Evidence from borrowings. *Language* 76. 324–357.

Paradis, Carole & Antoine Tremblay. 2009. Nondistinctive features in loanword adaptation: The unimportance of English aspiration in Mandarin Chinese phoneme categorization. In Wetzels & Calabrese (2009), 211–224.

Paradis, Carole, Caroline Lebel & Darlene LaCharité. 1994. Adaptation d'emprunts: Les conditions de la préservation segmentale. *Toronto Working Papers in Linguistics* 12. 108–134.

Phillips, Betty S. 2006. *Word frequency and lexical diffusion*. New York: Palgrave Macmillan.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Rice, Keren. 1990. Predicting rule domains in phrasal phonology. In Inkelas & Zec (1990), 289–312.

Rojas, Nelson. 1988. Fonología de las líquidas en el español cibaeño. In Robert Hammond & Melvyn Resnick (ed.) *Studies in Caribbean Spanish dialectology*, 103–111. Washington, DC: Georgetown University Press.

Sproat, Richard. 1985. On deriving the lexicon. Ph.D. dissertation, MIT.

Steriade, Donca. 1995. Underspecification and markedness. In Goldsmith (1995), 114–174.

Steriade, Donca. 2007. Contrast. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 139–157. Cambridge: Cambridge University Press.

Ulrich, Charles H. 1997. Loanword adaptation in Lama: Testing the TCRS model. *Canadian Journal of Linguistics* 42. 415–463.

Wetzels, W. Leo & Andrea Calabrese (eds.) 2009. *Studies in loan phonology*, 211–224. Amsterdam & Philadelphia: John Benjamins.

Wiese, Richard. 1996. *The phonology of German*. Oxford: Clarendon Press.

Zinc, Gaston. 1986. *Phonétique historique du français*. Paris: Presses Universitaires de France.

77 Long-distance Assimilation of Consonants

Sharon Rose

1 Introduction

There are numerous patterns in languages in which consonants assimilate at a distance for some acoustic or articulatory property. When vowels and consonants intervening between the assimilating consonants show no observable effect of the assimilating property, such patterns are labeled "consonant harmony." Other terms such as "consonant agreement" have been used (Rose and Walker 2004) in order to distinguish them from cases of harmony involving both vowels and consonants, such as emphasis harmony (Shahin 2002; CHAPTER 25: PHARYNGEALS) or nasal harmony (Walker 2000a; CHAPTER 78: NASAL HARMONY). Consonant harmony has played a central role in debates concerning harmony patterns in general (Rose and Walker, forthcoming) with respect to several issues: locality of interaction, transparency or blocking in long-distance assimilation, and directionality. In this chapter, the main typological patterns of consonant harmony are outlined, highlighting the challenges that the typology presents, including a discussion of harmony domains and directionality. Two main theoretical approaches to consonant harmony are then explored: analyses involving *spreading* an assimilating feature or extending a gesture across all segments within a string, and analyses advocating distinct *correspondence* relationships between consonants independently of intervening segments. The role of contrast in determining harmony interaction is examined within both of these frameworks. Finally, experimental approaches to consonant harmony are discussed, showing how they shed light on the analysis of consonant harmony.

2 Typology of long-distance assimilation of consonants

Long-distance assimilation of consonants or "consonant harmony" can be defined as:

(1) *Consonant harmony*

Assimilation for an articulatory or acoustic property between two or more non-adjacent consonants, where intervening segments are not noticeably affected by the assimilating property.

An example is given in (2) from Tahltan, an Athabaskan language (Shaw 1991). The 1sg subject prefix /s-/ (2a) is realized as [θ] when a dental fricative or affricate follows (2b), or as [ʃ] when a lamino-post-alveolar fricative or affricate follows (2c). Intervening consonants and vowels, including other coronal consonants, are transparent to the harmony:

(2) *Tahltan coronal harmony*

a.	ɛsk'aː	'I'm gutting fish'
	nɛstɛɬ	'I'm sleepy'
b.	xaʔɛθt'aθ	'I'm cutting the hair off'
	ɛθduːθ	'I whipped him'
c.	ɛʃdʒɪni	'I'm singing'
	jaʃtɬ'ɛʧ	'I splashed it'

Consonant harmony can involve morpheme alternations, as in (2), but may also occur as a morpheme structure constraint (CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS), requiring consonants within a root to share featural properties. In Ngizim (Chadic) roots, non-implosive obstruents must have the same voicing property (3a), unless the linear order of the obstruents is voiced . . . voiceless (3b) (Schuh 1997):

(3) *Ngizim laryngeal harmony*

a.	kùtə́r	'tail'	
	tàsáu	'find'	
	zə̀dù	'six'	(Hausa /ʃídà/)
	gâazá	'chicken'	(Hausa /kàazáa/)
b.	bàkú	'roast'	
	gùmʧí	'chin'	

The asymmetrical nature of the restriction points to a harmonic process. Voiceless . . . voiced combinations are not sanctioned, and Hausa words with such sequences are realized as voiced . . . voiced in Ngizim. The Ngizim harmony is therefore a regressive harmony, in which voiceless consonants assimilate to voiced, but not vice versa.

The definition of consonant harmony provided in (1) excludes other types of long-distance harmony that also involve assimilation spanning several segments, including both vowels and consonants, such as nasal harmony (Piggott 1988, 1992, 2003; Piggott and van der Hulst 1997; Walker 2000a, 2000b, 2003; CHAPTER 78: NASAL HARMONY) or post-velar or emphasis harmony (Younes 1993; Watson 1999; Zawaydeh 1999; Shahin 2002; CHAPTER 25: PHARYNGEALS). See also CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS on consonant–vowel interactions in general. In these harmony systems, assimilation affects both vowels and consonants, and certain types of segments can block harmony. Transparency of consonants is observed only under very restricted conditions. Conversely, transparency is routine in consonant harmony, whereas blocking is rare.

There are several types of long-distance consonant assimilation identified in typological studies of consonant harmony, laid out in detail in Hansson (2001a) and summarized in Rose and Walker (2004). The main types are outlined in the following sections.

2.1 *Laryngeal harmony*

Laryngeal harmony requires consonants to agree in aspiration, glottalic airstream, or voicing. Laryngeal distinctions are characterized by the features [spread glottis], [constricted glottis], and [voice], respectively, although different feature specifications are possible. Laryngeal harmony is most frequently observed in morpheme structure constraints (MacEachern 1999).

Laryngeal harmony is found in Chaha, a Gurage Semitic language of Ethiopia (Rose and Walker 2004), in which oral coronal and velar stops in roots match for both [constricted glottis] and [voice]:

(4) *Chaha laryngeal harmony*

a.	jɨ-t'ək'ɨr	'he hides'		
	ji-t'əβk'	'it is tight'		
	ji-t'ək'ik'	'it is being crushed'	cf. Endegegn (Gurage)	ɨ-dəkk'
	ji-t'ərk'	'it is dry'	cf. Masqan (Gurage)	ji-dərk'
b.	jɨ-kətf	'he hashes (meat)'		
	jɨ-kəft	'he opens'		
	jɨ-təks	'he sets on fire'		
c.	jɨ-gədɨr	'he puts to sleep'		
	jɨ-dərg	'he hits, fights'		
	jɨ-gəda	'he draws liquid'	cf. Amharic	ji-k'ada-l

Cognates in related languages show laryngeal mismatches, giving insight into the direction and implementation of the harmony. Harmony was regressive, and either ejectives or voiced stops could trigger harmony. Exceptions to laryngeal harmony involve non-adjacent combinations of an ejective and a voiced stop, e.g. [ji-gəmt'] 'he chews off'.

Voicing and aspiration harmony is found in (non-click) stops in disyllabic roots of Zulu (Bantu), as in (5a) (Khumalo 1987; Hansson 2001a). Zulu contrasts plain stops (which may be realized as ejective), voiced stops (described as "depressors," as they can lower tone),[1] and aspirated stops. Loanwords (5b) are adapted to conform to laryngeal harmony.

(5) *Zulu laryngeal harmony*

a.	ukú-peta	'to dig up'	
	úku-p^{h}át^{h}a	'to hold'	
	uku-guba	'to dig'	
b.	í-k^{h}ôt^{h}o	'court'	
	um-bídi	'conductor'	< English *beat*

Ngizim voicing harmony was illustrated in (3). Kera (Chadic) appears to have voicing alternations in affixes conditioned by voiced stops or affricates in the stem (Ebert 1979; Rose and Walker 2004), e.g. [kə-sár-káŋ] 'black (coll.)' *vs.*

[1] Zulu voiced stops may be phonetically voiceless (Traill *et al.* 1987), so this is not a clear case of "voicing" harmony.

[gə-dʒàr-gáŋ] 'colorful (coll.)'. However, Pearce (2005) argues that voicing is conditioned by a neighboring low tone rather than the voiced stop in the stem, so this does not constitute a case of voicing harmony. Hansson (2004) argues that in Yabem, a Huan Golf language of Papua New Guinea, voicing restrictions arose from tonal patterns, and only superficially resemble consonant harmony.

Laryngeal harmony is often restricted to apply between subclasses of obstruents. Harmony operates between pulmonic obstruents in Ngizim, whereas in Chaha and Zulu, it applies between stops with differing airstream mechanisms, but excludes fricatives. In Kalabari Ijo (Jenewari 1989; Hansson 2001a) and Bumo Izon (Efere 2001; Mackenzie 2005, 2009), Ijoid languages of Nigeria, plain voiced stops and implosives may not co-occur in roots. Other cases of laryngeal harmony require homorganicity or complete identity between consonants. In Bolivian Aymara, laryngeal harmony for aspiration and ejectivity occurs between homorganic stops, so they are identical, e.g. [k'ask'a] 'acid to the taste', whereas no harmony occurs between heterorganic stops: e.g. [t'aqa] 'flock, herd' (Hardman *et al.* 1974; Davidson 1977; de Lucca 1987; MacEachern 1996, 1999). Similar effects are found in Mayan languages, such as Chol (Gallagher and Coon 2009), Modern Yucatec (Straight 1976), and Tzutujil (Dayley 1985; Gallagher 2010).

In conclusion, laryngeal harmonies are attested in numerous languages, most typically those that exhibit a three-way contrast in laryngeal features. Laryngeal harmony is usually root-restricted, may be subject to homorganicity requirements, and appears to be regressive for those cases in which directionality can be identified.

2.2 *Coronal harmony*

Coronal harmonies involve articulations both for tongue tip/blade posture (apical *vs.* laminal) and tongue position (dental, alveolar, post-alveolar). Sibilant harmony is the most commonly attested type of consonant harmony and requires sibilant coronal fricatives and affricates to match for tongue tip/blade posture and location. It is widely attested in Native American languages, particularly in Athabaskan and Chumash languages, but it also occurs in Basque, Berber, Bantu, Cushitic, and Omotic languages. An example of sibilant harmony in Tahltan was illustrated in (2).[2] In Sidaama, a Cushitic language of Ethiopia (Kawachi 2007), the causative suffix /-is/ (6a) is realized as [iʃ] when palato-alveolar fricatives or affricates appear in the preceding stem (6b).

(6) *Sidaama sibilant harmony*

a.	dirr-is	'cause to descend'
	hank'-is	'cause to get angry'
	raʔ-is	'cause to become cooked'
b.	miʃ-iʃ	'cause to despise'
	ʃalak-iʃ	'cause to slip'
	ʧ'uf-iʃ	'cause to close'

[2] Tahltan harmony may be only partially sibilant, since it is not clear the fricatives [θ] and [ð] are sibilant. They are described as predorsal alveolar in Nater (1989).

Sibilant harmony operates across vowels and non-sibilant consonants, including other coronals. In (6b), the intervening segments do not block and do not participate in the harmony.

In some languages, such as Ineseño Chumash (Applegate 1972; Poser 1982; McCarthy 2007), both alveolar and post-alveolar sibilants may trigger harmony. The rightmost sibilant determines the tongue tip-blade realization of all sibilants in the stem. In (7a) and (7c), the 3rd singular subject prefix is /s-/, but it is realized as [ʃ] if there is a palatal sibilant to its right, (7b) and (7d). In contrast, the dual marker /iʃ-/ (7e) is realized as [is] if followed by an alveolar sibilant (7f).

(7) *Ineseño Chumash sibilant harmony*

a.	/s-ixut/	[sixut]	'it burns'
b.	/s-ilakʃ/	[ʃilakʃ]	'it is soft'
c.	/ha-s-xintila/	[hasxintila]	'his gentile'
d.	/ha-s-xintila-waʃ/	[haʃxintilawaʃ]	'his former gentile'
e.	/p-iʃ-al-nanʔ/	[piʃananʔ]	'don't you two go'
f.	/s-iʃ-tiʃi-jep-us/	[sistisijepus]	'they two show him'

Dental harmony is found in Nilotic languages such as DhoLuo (Stafford 1967; Yip 1989; Tucker 1994), Anywa (Reh 1996), Mayak (Andersen 1999), and Päri (Andersen 1988). It operates between dental and alveolar stops, including nasals if a contrast exists in the language, and may be triggered by either. In Päri (Andersen 1988; Hansson 2001a), dental harmony is respected in roots (8a). Root-final stops that are the product of final mutation combined with affixation match the dental or alveolar property of the initial stop (8b).

(8) *Päri dental harmony*

a.	n̪ɔt̪	'sucking'		
	àtwá:t̀	'adult male elephant'		
b.	dè:l	'skin'	dè:nd-á	'my skin'
	t̪ùol	'snake'	t̪ùon̪d̪-á	'my snake'

In Mayak (Andersen 1999), harmony is triggered by an alveolar and optionally affects suffixes of the shape /-Vt̪/, as in (9). Intervening stops that are noncontrastive for the dental–alveolar distinction are transparent to the harmony.

(9) *Mayak dental harmony*

a.	leɣ-it̪	'tooth'
	wʌð̆-it̪	'buttock'
	ʔi̪n-ʌt̪	'intestine'
b.	tiɗ-ʌt̪ ~ tiɗ-ʌt	'doctor'
	kɛt-ɪn-ɛt̪ ~ kɛt-ɪn-ɛt	'star'

Retroflex harmony is reported for several languages. In Gimira (Benchnon), an Omotic language of Ethiopia (Breeze 1990), retroflex harmony restricts combinations of sibilants, requiring them to agree for retroflexion. In Malto (Dravidian) (Mahapatra 1979; Hansson 2001a), retroflex harmony operates between oral stops.

In Australian languages such as Arrernte (Arandic) (Henderson 1998; Tabain and Rickard 2007), apical alveolar and retroflex stops match for retroflexion in a root. Retroflex harmony is also reported in Kalasha (Indo-Aryan) (Trail and Cooper 1999; Arsenault and Kochetov, forthcoming), where it operates between stops, between fricatives, or between affricates in a root, but combinations of different manners of articulation may disagree for retroflex. Kalasha contrasts dentals, retroflex, and palatals; dentals and palatals also tend not to co-occur, so this is a general coronal harmony.

(10) *Kalasha retroflex/palatal harmony*

a.	stops	dental	tʰedi	'now'
		retroflex	ʈoʈ	'apron'
b.	fricatives	dental	sastirek	'to roof a house'
		palatal	ʑoɕi	'spring festival'
		retroflex	ʂuʂik	'to dry'
c.	affricates	dental	tsètsaw	'squirrel'
		palatal	tɕʰatɕi hik	'to take care of'
		retroflex	ɖʐaʈʂ	'spirit beings'

Other cases of coronal harmony involving alveolar stops and alveo-palatal affricates are reported in Hansson (2001a), and include Aymara (de Lucca 1987), Kera (Ebert 1979), and Pengo (Dravidian) (Burrow and Bhattacharya 1970). In each case, harmony rules out /t . . . ʧ/ sequences, but allows the reverse, /ʧ . . . t/.

In terms of directionality, Hansson (2001a) points out two main directionality effects with respect to sibilant harmony. First, sibilant harmony shows a strong tendency to be regressive. In some cases, harmony is triggered by the rightmost sibilant, regardless of its location within a root or affix, as in Chumash (7) or Navajo (11) (McDonough 1991). The 1sg subject prefix /-iʃ/ is variably realized as [is] or [iʃ], depending on whether /s/ follows.[3]

(11) *Navajo sibilant harmony*

/j-iʃ-mas/	jïsmas	'I'm rolling along'
/dz-iʃ-ɬ-taːɬ/	ʤiʃtaːɬ	'I kick him (below the belt)'
/dz-iʃ-l-ts'in/	dzists'in	'I hit him (below the belt)'

Hansson (2001a, 2001b) relates the regressive bias of sibilant harmony to speech production. In speech production studies, anticipatory errors and assimilations are more common than are perseverative (Dell *et al.* 1997). This is modeled in a serial order theory of speech production, whereby one segment activates a consonant being planned and anticipates its production. There are cases of progressive sibilant harmony, as in the Sidaama case in (6), but in such cases, a suffix alternates in agreement with a root. The same pattern holds for dental harmony; no strong evidence for regressive patterns in dental harmony or retroflex harmony has been detected.

3 Navajo actually has examples of progressive sibilant harmony in the prefix string. See McDonough (1990, 1991) and Hansson (2001a: 193–198) for discussion.

The second directionality effect concerns the nature of the trigger. While some cases of sibilant harmony are like Navajo in that either alveolar or post-alveolar consonants can trigger harmony, other languages only allow /s/ to become [ʃ] and not the reverse. Hansson (2001a: 472) cites sixteen cases of the /s/ → [ʃ] pattern, but only one case of /ʃ/ → [s]. Hansson connects this effect to speech planning and the *palatal bias* effect reported in speech error research (Shattuck-Hufnagel and Klatt 1979). The palatal bias effect refers to the higher frequency with which alveolar consonants act as targets of speech errors by palatals.

2.3 *Nasal harmony*

Nasal consonant harmony is attested primarily in Bantu languages. Nasal stops harmonize with voiced stops and oral approximants. If voiceless stops harmonize, they do so only if voiced stops harmonize. In Kikongo (Dereau 1955; Ao 1991; Odden 1994), a nasal stop in a verb root causes a [d] in the active perfect suffix (12a) or [l] in the applicative suffix to be realized as [n] (12b).

(12) *Kikongo nasal harmony*

a.	n-suk-idi	'I washed'	tu-nik-ini	'we ground'
	m-bud-idi	'I hit'	tu-sim-ini	'we prohibited'
b.	ku-sakid-il-a	'to congratulate for'	ku-nat-in-a	'to carry for'
	ku-toot-il-a	'to harvest for'	ku-dumuk-is-in-a	'to cause to jump for'

Intervening vowels and other consonants are transparent to the harmony. Yaka has a similar nasal harmony pattern (Hyman 1995). In other languages, the nasal harmony is restricted to apply only across an intervening vowel, as in Lamba (Odden 1994), Bemba (Hyman 1995), Ndonga (Viljoen 1973), and Herero (Booysen 1982), and may be restricted to roots only. The main distinctions between nasal consonant harmony and general nasal harmony are (i) vowels are not nasalized, (ii) the trigger is a nasal consonant that targets a similar consonant (voiced stop or approximant), and (iii) other consonants and vowels do not block harmony. See CHAPTER 78: NASAL HARMONY for more extensive discussion of the distinction between the two kinds of nasal harmony.

Nasal harmony operates progressively from root to suffix. However, it cannot be reduced in all cases to a stem-control effect. In Kikongo, roots such as /dumuk/ are possible, with a voiced stop preceding a nasal. The reverse order of nasal followed by voiced stop is not attested (Ao 1991; Piggott 1996), indicating that nasal harmony applied progressively within the root. The same pattern is attested in Yaka (Rose and Walker 2004).

2.4 *Liquid harmony*

Liquid harmony involves alternations between /r/ and /l/ (CHAPTER 31: LATERAL CONSONANTS). In Bukusu (Bantu), liquid harmony is attested in roots (Hansson 2001a). In addition, the benefactive suffix /-il-/ is realized as [-ir-] following a stem with [r] (Odden 1994). Vowel height harmony applies to the suffix.

(13) *Bukusu liquid harmony*

teex-el-a	'cook for'	reeb-er-a	'ask for'
lim-il-a	'cultivate for'	kar-ir-a	'twist'
iːl-il-a	'send thing'	resj-er-a	'retrieve for'

In Sundanese (Malayo-Polynesian), /l/ triggers harmony of /r/ to [l] (Cohn 1992), as illustrated with the plural infix /-ar-/ in the final form in (14).

(14) *Sundanese liquid harmony*

kusut	'messy'	k-ar-usut	'messy (PL)'
rahit	'wounded'	r-ar-ahit	'wounded (PL)'
ləga	'wide'	l-al-əga	'wide (PL)'

Liquid harmony is also attested in Pohnpeian (Rehg and Sohl 1981). There are cases in which liquids alternate with glides in Bantu languages, such as in Basaa (Lemb and de Gastines 1973) and Pare (Odden 1994), and in which /l/ alternates with a lateral tap in ChiMwiini (Kisseberth and Abasheikh 1975). All cases of liquid harmony are either root-restricted or involve suffiix alternations, so no directionality bias can be detected.

2.5 Dorsal harmony

Dorsal harmony is found in Malto, Gitksan (Tsimshianic), Aymara, and the Totonacan languages, and involves alternations between velar and uvular consonants. In Tlachichilco Tepehua (Watters 1988; Hansson 2001a), a uvular /q/ causes a preceding velar to become uvular, which in turn conditions lowering of the preceding high vowel (15b).

(15) *Tlachichilco Tepehua dorsal harmony*

a.	ʔuks-k'atsaː	[ʔuksk'atsaː]	'feel, experience sensation'
b.	ʔuks-laqts'-in	[ʔoqslaqts'in]	'look at Y across surface'

Hansson (2001a) notes that intervening vowels are not affected by the harmony even though uvulars lower adjacent vowels. In the word /lak-puːtiq'i-ni-j/ → [laqpuːteʔenij] 'X recounted it to them' (the /q'/ is realized as [ʔ]) the vowel /uː/ fails to lower to [oː], despite appearing between two uvulars. Compare this with (15b). In Gitksan (Brown 2008), the harmony effect is a static co-occurrence restriction that can operate at a distance. Dorsal harmony causes velars to become uvular. While most dorsal harmony cases are regressive and target roots, this could be either a directionality effect or due to the trigger consonant, the uvular, being in an affix.

2.6 Stricture and secondary articulation harmony

In addition to the main types reported in §2.2–§2.5, Hansson (2001a) also lists stricture and secondary articulation harmonies. Stricture involves alternations between stops and fricatives, as in Yabem, e.g. /se-dàgùʔ/ → [tédàgùʔ] 'they follow (REALIS)'. Secondary articulation refers to labialization, palatalization, velarization, or pharyngealization. There are a few reported cases discussed in Hansson (2007a): pharyngealization in Tsilhqot'in (also known as Chilcotin, Athapaskan) (Cook 1983, 1993), which interacts with sibilant harmony, velarization in Pohnpeian (Micronesian) (Rehg and Sohl 1981; Mester 1988), and palatalization in Karaim (Turkic) (Kowalski 1929; Hamp 1976; Nevins and Vaux 2004), as shown below:

(16) *Karaim palatal harmony*

d^{j}ort^{hj}-un^{j}tʃ^{j}u	'fourth'
alt^{h}ɯ-ntʃɯ	'sixth'

In sum, consonant harmony targets a range of segments: dorsals, liquids, and coronals, as well as segments differentiated by nasal and laryngeal features. Hansson (2001a) and Rose and Walker (2004) point out that a consistent characteristic of consonant harmony is the high degree of similarity between the interacting segments. Harmony is restricted to minor place or tongue features distinguishing among coronals and dorsals or to features that are also prone to local assimilation. Notably absent, however, is harmony for major place features such as [labial], [coronal], or [dorsal], as well as classificatory features that tend not to assimilate locally, such as [sonorant], [continuant], or [consonantal]. Rose and Walker (2004) relate the absence of place harmony to the inability of major place to change even in local assimilations, citing articulatory speech error research that shows that major place gesture errors tend to be additive rather than replacive (Goldstein *et al.* 2007; Pouplier 2007). Gafos (1999), on the other hand, argues that major place features cannot spread across vowels (*contra* Shaw 1991) without serious interruption of the vowel gestures. Only minor features such as tongue tip position can do this. See §3.1.1 for further discussion.

The lack of major place consonant harmony is intriguing in light of two related phenomena: child language and dissimilation. Consonant harmony for major place is attested in child language (Vihman 1978), and according to CHAPTER 72: CONSONANT HARMONY IN CHILD LANGUAGE, it is the most common type of consonant harmony in child language. Recent analyses and proposals are discussed in Goad (1997), Berg and Schade (2000), Rose (2000), Pater and Werle (2003), and Rose and dos Santos (2006). The same mechanisms that underlie child phonology and adult phonology may not be the same; some child productions may be due to developmental factors (Rose and dos Santos 2006; Inkelas and Rose 2008). See CHAPTER 72: CONSONANT HARMONY IN CHILD LANGUAGE for an overview.

Some authors have drawn a connection between long-distance consonant assimilation and long-distance consonant dissimilation (MacEachern 1999; Walker 2000c; Gallagher 2008), arguing that they are alternate responses to the same pressure. This does appear to be the case for laryngeal and liquid harmony. Yet there are key differences. A common dissimilation process occurs between labial consonants (Alderete and Frisch 2007), whereas labial consonant harmony is unattested. Nasal dissimilations involve prenasalized stops and nasals (Odden 1994), but these segments do not interact in nasal consonant harmony; prenasalized stops are transparent to nasal harmony and do not act as triggers (CHAPTER 78: NASAL HARMONY). A general theory of the relationship between long-distance consonant assimilation and consonant dissimilation currently appears elusive.

3 Analyses of long-distance assimilation

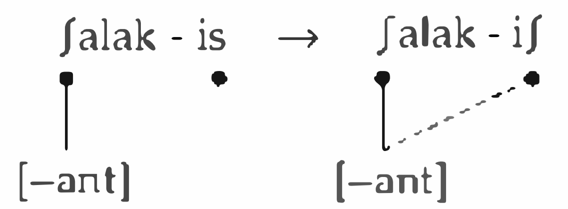
Two main theoretical analyses have been formulated for long-distance consonant harmony: *spreading* and *correspondence*. Spreading involves the extension of a gesture or feature across a string of segments, building upon early autosegmental analyses of vowel harmony and nasal harmony. Correspondence is proposed in Walker (2000b, 2000c), Hansson (2001a), and Rose and Walker (2004) and requires similar consonants to "correspond" and match each other for particular features, regardless of intervening consonants and vowels. The following sections outline these two approaches and identify the strengths of each proposal, as well as the challenges they encounter.

3.1 *Spreading*

Autosegmental phonology represented a major shift in the analysis of harmony systems. Although the distance effects of harmony systems had been explored within a Firthian prosodic analysis framework (Palmer 1970), Clements's (1980) groundbreaking analysis of vowel harmony and extension to nasal harmony launched the study of harmony systems using autosegmental spreading. Early autosegmental analyses of consonant harmony include Halle and Vergnaud (1981) on Navajo and Poser (1982) on Chumash.

In an autosegmental representation, the harmonizing feature (P-segment) is projected onto its own tier and linked to the segment (P-bearing segment) by means of an association line. Spreading involves extending the feature to other segments in the word via new association lines, as shown in (17) for the Sidaama word [ʃalakiʃ] 'cause to slip':

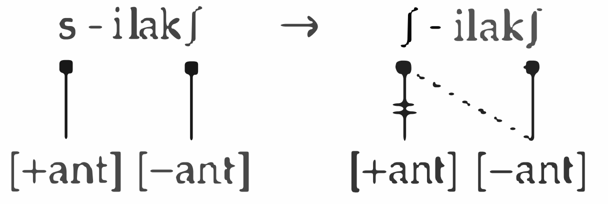
(17) *Long-distance spreading*



(17) illustrates a feature-filling rule, in which the feature [−anterior], characterizing the post-alveolar /ʃ/, spreads to the /s/, but the /s/ itself is unspecified for the feature [+anterior]. Within models of underspecification (Archangeli 1988; Paradis and Prunet 1989), the default feature [+anterior] is assumed to be filled in by a default rule at the end of the derivation if no specification is provided by a specific rule.

Consonant harmony may also be *feature-changing* where the target and trigger have opposite values for the spreading feature. Sibilant harmony in Ineseño Chumash has been analyzed as feature-changing (Poser 1982, 1993; Lieber 1987; Steriade 1987b; Shaw 1991), because harmony can be triggered by either [+anterior] /s/ or [−anterior] /ʃ/, altering the specification of the other sibilant. The target consonant acquires the specification of the trigger through spreading, and loses its own feature specification by delinking its original association line. This is illustrated in (18) for [ʃ-ilakʃ] 'it is soft':

(18) *Feature-changing harmony*



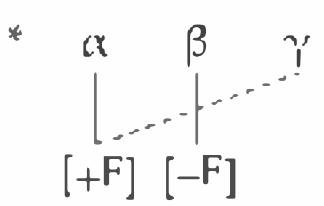
In (18), the feature [anterior] is shown linked to the rest of the segment. In more articulated models of feature geometry (Clements 1985, 1991; Sagey 1986; McCarthy 1988; Clements and Hume 1995), features link to organizing nodes, which in turn link to the root node, connected directly to prosodic structure.

3.1.1 Transparency and spreading

In both of the representations above, spreading takes place between two sibilants, despite the fact that other consonants and vowels intervene between the trigger and target. Intervening segments have the potential to block harmony, a phenomenon that is routinely observed in both vowel harmony and local vowel–consonant harmony. However, blocking is not generally observed with consonant harmony (Hansson 2001a; Rose and Walker 2004).[4]

Blocking segments, or "opaque" segments, can be characterized as those that are specified with the opposite value to the spreading feature. For example, in van der Hulst and Smith (1982a) [−nasal] segments block spreading of [+nasal] in nasal harmony. The [+nasal] feature cannot spread over the association line linking [−nasal] to a non-nasal segment, as this would violate the No-Crossing Constraint (Goldsmith 1979), a principle of autosegmental phonology (CHAPTER 14: AUTOSEGMENTS), which prevents association lines crossing. This is illustrated schematically in (19), where the symbols α, β, and γ represent segments and [+F] the spreading feature, and where the [−F] value of a feature blocks the [+F] value from spreading.

(19) *No-Crossing Constraint*

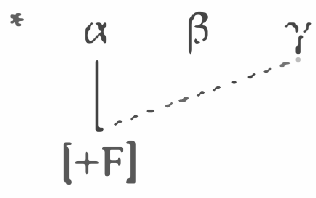


In more recent accounts, the blocking segment is assumed to be incompatible with the spreading feature due to an articulatorily grounded constraint (Archangeli and Pulleyblank 1994), preventing co-occurrence of the spreading feature and a specific feature or features of the blocking segment. For example, in nasal harmony, obstruents block nasal harmony in many languages (Walker 2000a), so [+nasal] may be restricted from associating to a [−sonorant] segment. Under this scenario, the No-Crossing Constraint does not apply; instead *locality* considerations prevent the spreading feature from skipping over the blocking segment and spreading to another segment. How locality should be defined has been a matter of debate.

[4] One identified case is a voicing alternation in Imdlawn Tashlhiyt Berber, which is parasitic on sibilant harmony. Voiceless obstruents block voicing from transferring from stem to prefix.

If locality is defined at the level of the root node or segment, such that adjacent segments are local, no segment can be skipped in spreading, a theory referred to as *strict locality*. This is illustrated schematically in (20).

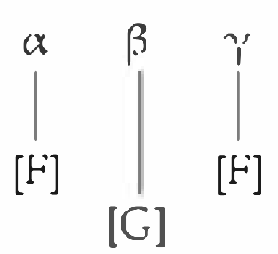
(20) *Strict locality*



Under this view, consonant harmony would be similar to local assimilation (CHAPTER 81: LOCAL ASSIMILATION), but extended over longer strings of segments.

Locality may also be defined at the level of vowel nuclei of adjacent syllables (Archangeli and Pulleyblank 1987, 1994), in which case intervening consonants would be considered transparent. This is referred to as *maximal scansion* in Archangeli and Pulleyblank (1987). A third possibility is to define locality with respect to autosegmental feature tiers, or feature node tiers in a feature-geometry model, referred to as *minimal scansion* (Archangeli and Pulleyblank 1987; Steriade 1987a). Segments that lack specification on such tiers would not be computed for locality. In the following schematized representation, the features [F] of the segments α and γ would be adjacent on the F tier, despite not belonging to adjacent segments:

(21) *Tier-based locality*



Tier-based locality lies at the heart of autosegmental spreading analyses of transparency in harmony.

Steriade (1987b) argued that intervening transparent consonants and vowels in Ineseño Chumash sibilant harmony lack specification for the feature [anterior] at the point when the harmonic spreading rule applies. Dorsal and labial consonants are excluded from participation in the harmony, as they have place feature specifications on other tiers – Dorsal and Labial. The feature [anterior] is assumed to be relevant only for coronal consonants. The same holds true for vowels in a system in which vowels are considered dorsal (Sagey 1986; Steriade 1987a).[5] Yet the coronal consonants /t l n/ are also transparent in Chumash. Steriade adopts a form of contrastive specification, wherein only segments that contrast for a given feature need to be specified for that feature. The feature [anterior] is needed to distinguish sibilants in Chumash, but /t l n/ do not have [–anterior] counterparts, and so are predictably [+anterior]. Predictable features are left unspecified and filled in as default later in the derivation. Harmonic

[5] This analysis might be problematic for feature systems in which coronals and front vowels share specification (Clements and Hume 1995), depending on how locality is defined (see Odden 1994 for discussion of vowel–consonant locality issues in this model).

spreading of the [anterior] feature operates unhindered between sibilants before a redundancy rule ([+coronal, -continuant] → [+anterior]) fills in predictable values on the other coronals. Locality is defined on the tier [anterior] as the spreading rule targets only consonants specified for [anterior], not those specified as Coronal, the organizing node on which [anterior] is dependent (Shaw 1991). Chumash is a feature-changing rule, but if spreading rules are feature-filling, targets would need to be defined with respect to other features or the node to which the feature attaches.

Harmony does not always operate at the level of individual features, however. Shaw (1991) argues that a more complex harmony system in Tahltan involves spreading the Coronal node (CHAPTER 12: CORONALS). Tahltan has a rich inventory of coronal consonants, contrasting dental stops, lateral continuants, interdental/pre-dorsal sibilants, alveolar sibilants, and palatal sibilants. The latter three sibilant classes participate in coronal harmony, but the stops and laterals do not (examples in (2)). Shaw argues that in order to distinguish among a series of three sibilants, at least two features dependent on Coronal are needed. Under the assumption that a single unified spreading rule should capture the harmony, Shaw proposes that harmony involves spreading of the Coronal node. The other two transparent classes must be underspecified for Coronal at the time the rule applies. Similar harmonic effects (e.g. /s/ → [ʃ]), therefore, involve different spreading rules, depending on the particular inventory of the language.

Gafos (1998, 1999) rejects tier-based locality, and presents a model of "Articulatory Locality," in which locality is defined in terms of articulatory gestures (Browman and Goldstein 1986, 1989, 1990). Vowel gestures are contiguous across a consonant, whereas consonant gestures are not contiguous across a vowel. Vowel harmony may appear to skip over consonants, but consonants are in fact unaffected audibly by the spreading gesture. This strict locality view is also adopted by Ní Chiosáin and Padgett (1997, 2001), Walker and Pullum (1999), and Walker (2000a). Under strict locality, only coronal harmony, which involves assimilation for a tongue tip-blade feature, is predicted to be possible, due to non-interference with vowels. The tongue tip-blade is independent of the tongue dorsum used in the production of vowels, and its exact posture has no significant acoustic effect on vowel quality. By the same reasoning, dorsal and labial consonants would be predicted to intervene as "transparent," since changes in the tongue tip-blade would not affect their production. Moreover, if the feature that distinguishes /s/ and /ʃ/ is apicality (tongue tip) *vs.* laminality (tongue blade), languages with no apical–laminal coronal stop contrast may allow stops to fluctuate between apical and laminal in different harmonic contexts, a suggestion made by Peter Ladefoged, as reported in Steriade (1995). Gafos (1999) formalizes this idea and proposes two new tongue tip-blade parameters: Tongue Tip Constriction Orientation (TTCO) and Tongue Tip Constriction Area (TTCA), gestures that do not skip over other segments, but are maintained through their production with little perceptible effect. Coronal segments /t n l/ in Chumash harmony are predicted to alter their production in accordance with the harmonic domain in which they occur, either apical [n̺] in words like /k-sunon-us/ → [ksun̺on̺us] 'I obey him' or laminal [n̻] in words like /k-sunon-ʃ/ → /[kʃun̻on̻ʃ] 'I am obedient'. TTCO is identified as tip-up (↑) for apical and tip-down (↓) for laminal. As non-sibilant coronals do not contrast on this dimension in Chumash, they are not perceived as distinct.

(22) *Gestural extension under strict locality*

	apical	*laminal*
	k - s u n̪ o n̪ - u s	k - ʃ u n̪ ● n̪ - ʃ
TONGUE TIP	TTCO ↑	TTCO ↓

Gafos argues that the strict locality view of consonant harmony explains the absence of other types of place harmonies. Major place gestures that define dorsal and labial consonants cannot spread across vowels (*contra* Shaw 1991) without serious interruption of the vowel gestures. Minor features such as tongue tip position can. Tier-based locality is unable to adequately explain why only Coronal, and not Labial and Dorsal, nodes can spread in feature geometry. In addition, the restriction of harmony to subclasses of coronals, such as sibilant fricatives and affricates, is explained, as these segments involve contrast along the tongue orientation dimension.

In conclusion, spreading approaches to harmony involve the spreading of a feature or the extension of a gestural parameter over other vowels and consonants. The non-participation of these consonants receives two explanations. In the autosegmental framework, it is due to a version of feature underspecification (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION) and tier-based locality, allowing for certain kinds of harmony interactions between specific features. In a gestural framework (Gafos 1999), transparency is illusory – articulators perform the harmonic gestures, but have little impact on consonants and vowels that do not involve those articulators, or for which changes in the articulation are non-contrastive and hence perceptually non-distinct.

3.1.2 *Challenges to the spreading approach*

While the analysis of consonant harmony as feature spreading or gestural extension seems appropriate for characterizing retroflex harmony and some sibilant harmony cases, it encounters several challenges when applied to a fuller typology of consonant harmony systems as outlined in §2. Gafos (1999) assumes that coronal harmony is the only type of consonant harmony. Shaw's (1991) typology of consonant harmony identifies only laryngeal harmony as another possibility.[6] Laryngeal features primarily distinguish among obstruents. As vowels and sonorants are inherently voiced and unspecified for laryngeal features (Itô and Mester 1986), laryngeal harmonies can operate between laryngeal tiers specified only on obstruents, thereby respecting tier locality. Gafos (1999) does not explicitly discuss laryngeal harmony within the Articulatory Locality model.

However, the larger typology outlined in §2 prompts Hansson (2001a) and Rose and Walker (2004) to conclude that autosegmental spreading is inadequate as a general model of consonant harmony. Their arguments rest on several key properties of consonant harmony not shared with vowel harmony and vowel–consonant harmony, as well as predictions that some spreading models make about the participation of intervening segments. I focus on two main properties here: (i) no blocking and transparency, and (ii) similarity of target and trigger. Hansson (2001a) also notes the lack of sensitivity to prosody and regressive

[6] Shaw does identify other harmonies, such as labial, but these are *dissimilatory* morpheme structure constraints or morphological affixation, rather than true consonant harmony as defined in this chapter.

directionality as defining properties of consonant harmony, but the prosody insensitivity may be due to other factors. Regressive directionality is a strong tendency, but progressive directionality is also observed for consonant harmony. Furthermore, regressive directionality is not an exclusive domain of consonant harmony; it has also been observed for vowel harmony (Hyman 2002) and some forms of vowel–consonant harmony – i.e. emphasis harmony exhibits more restrictions and blocking when progressive than when regressive (Watson 1999).

3.1.3 *Blocking and transparency*

Consonant harmony differs from other types of harmony with respect to blocking effects and transparency. If nasal consonant harmony is compared with nasal vowel–consonant harmony, there are two key differences with respect to the participation of segments. In nasal consonant harmony, nasal consonants harmonize with voiced stops or approximant consonants across other consonants and vowels, even obstruents. In contrast, nasal vowel-consonant harmony shows blocking effects, usually by the same segments that are skipped in nasal consonant harmony. Second, intervening vowels do not show nasalization in nasal consonant harmony, whereas they make the best targets in nasal vowel-consonant harmony and are generally not skipped. If nasal vowel–consonant harmony involves autosegmental spreading or gesture extension, how does one explain the differences with nasal consonant harmony? It does not appear to behave as if spreading of [nasal] is involved.

A similar argument can be applied to two other types of consonant harmony. Laryngeal harmony shows no extension of voicing/devoicing or glottalization over intervening segments. Dorsalization shows no effects on intervening vowels, despite the fact that uvulars routinely lower adjacent vowels. If harmony operates as advocated by Gafos (1999), with gestures extended across other segments, these facts are unexpected. If tier-based locality is the explanation, it is hard to give a reason for the neutrality of contrastive voiced and voiceless fricatives in a laryngeal harmony involving voice harmony between stops, as is the case in Chaha.[7]

3.1.4 *Similarity*

The concept of similarity is implicitly recognized in autosegmental spreading analyses of sibilant harmony, as spreading occurs only between segments specified for the spreading feature. However, it is not a formalized aspect of spreading theory. Hansson (2001a, 2001b) and Rose and Walker (2004) propose that similarity is the driving factor in consonant harmony, and has its functional roots in speech production. For example, sibilants are highly similar to one another and it is hypothesized that production is eased if they match for the position of the tongue tip-blade. Similar, but different, consonants present production difficulties that are manipulated in tongue-twisters and emerge as speech errors in both natural and experimentally induced situations (Fromkin 1971; Shattuck-Hufnagel and Klatt 1979; Frisch 1996; Rose and King 2007; Walker 2007; Kochetov and Radišić 2009). Nasal stops harmonize with oral sonorants or voiced stops, which differ

[7] One solution might be to use the feature [spread glottis] to characterize fricatives and the feature [voice] for stops (Vaux 1998). However, voiceless stops become voiced preceding voiced stops and voiced fricatives alike in Chaha (Rose and Walker 2004).

minimally from nasals. Voicing harmony occurs between obstruents, but is usually restricted to stops, excluding fricatives. Homorganicity further contributes to similarity; some laryngeal and nasal harmonies operate only between homorganic segments.[8] All cases of harmony involve strong similarity between the harmonizing segments, even in ways that local assimilations do not. For example, while local voicing assimilation operates between all obstruents, voicing harmony may be restricted to a sub-type of obstruents based on manner.

Rose and Walker (2004) determine similarity using the metric developed in Frisch *et al.* (2004), wherein similarity is assessed on the basis of shared natural classes of distinctive features in a given language. The numbers of shared and unshared natural classes of two consonants are compared. Both the size and contrastiveness of the segment inventory contribute to the similarity ratings. Natural classes, which incorporate the notion of contrastiveness, are better able to predict gradient phonotactics and capture major class subregularities than are models based simply on distinctive feature specification. However, see Mackenzie (2005, 2009) for some criticisms of this metric.

3.2 Correspondence

Given these observations about the typology of consonant harmony, Hansson (2001a) and Rose and Walker (2004), based on Walker (2000a, 2000c), developed an account of consonant harmony within Optimality Theory (OT), termed "agreement-by-correspondence." A correspondence relationship is created between similar segments, expressed as CORR-C↔C constraints (indicated in the diagram in (23) by co-indexation). This is reminiscent of Zuraw's (2002) aggressive reduplication model, although this model does not encode similarity directly. Crucially, there is no autosegmental feature spreading between the segments, so their feature specifications are distinct.

(23) C_x V C V C_x
| |
[F] [F]

The CORR-C↔C constraints are arranged in a fixed implicational hierarchy from most similar to least similar, for example CORR-T^h↔T^h >> CORR-T^h↔T >> CORR-K^h↔T (Rose and Walker 2004: 500). Separate IDENT-CC constraints require the corresponding consonants to agree for a given feature. Input–output faithfulness constraints are placed between the CORR-C↔C constraints to achieve harmony of different similarities, or below them to produce full harmony.

The following tableau illustrates an example of sibilant harmony in Sidaama for the word /ʃalak-is/ → [ʃalak-iʃ]. CORR-s↔ʃ refers to anterior and non-anterior fricative pairs, while CORR-t↔ʃ refers to anterior stop and non-anterior fricative combinations. Candidate (24a) has a CC-correspondence relationship (indicated by the subscript $_x$ on the output sibilant consonants) and sibilant agreement, thereby

[8] Hansson (2007a) has argued that secondary articulation consonant harmonies may have a diachronic explanation related to (re)interpretation of C-V co-articulation, but similarity at the level of the secondary articulation is still observed.

satisfying the two high-ranked constraints. This candidate violates IDENT-OI[ant], due to the change /s/ → [ʃ]. IDENT-OI[ant] is violated by segments that alter an input [anterior] specification in the output. Candidate (24b) has no correspondence relationship between /s/ and /ʃ/, indicated by the different subscripts $_x$ and $_y$. Due to the lack of CC-correspondence, this candidate does not violate IDENT-CC[ant]. Candidate (24c), on the other hand, does have sibilants in a CC-correspondence relationship. The sibilants do not agree for anteriority, thereby violating IDENT-CC[ant]. The [anterior] feature is used here although other features such as [distributed] or [Tongue Tip Constriction Orientation] are also possible.

(24)

/ʃalak-is/	IDENT-CC[ant]	CORR-s↔ʃ	IDENT-OI[ant]	CORR-t↔ʃ
☞ a. $ʃ_x$alak-i$ʃ_x$			*	
b. $ʃ_x$alak-is_y		*!		
c. $ʃ_x$alak-is_x	*!			

No correspondence relationship is established between the fricative and the voiceless stop /k/, or with the vowels, as these two sounds are not sufficiently similar. Other work analyzing coronal harmony systems as involving corresponding segments or feature copy includes Clements (2001) and McCarthy (2007).

The correspondence-based approach to consonant harmony allows similar consonants to agree at a distance; transparent segments are those that are not similar enough to participate in the harmony. No blocking is predicted, as lack of harmony is due to either the lack of or the low ranking of correspondence between intervening segments. This approach sets consonant harmony apart from vowel harmony and vowel–consonant harmony in using a different analytical mechanism.[9]

A more accurate typology of consonant harmony has led to alternate analytical devices, using correspondence-based relations rather than autosegmental spreading. The assumption that all harmony systems are alike and therefore subject to the same type of analysis has also been called into question, representing a significant departure in the analysis of consonant harmony *vs.* other harmony systems.

3.2.1 *Challenges to the correspondence approach*

Despite the advances of the correspondence approach in unifying the typology of consonant harmony and setting it apart from other types of harmonies, challenges to this model have arisen.

In the arena of coronal harmony, there is still debate over whether correspondence is the appropriate mechanism. McCarthy (2007) argues that Chumash harmony should be analyzed via correspondence, as it shows clear differences from local assimilations and dissimilations. Arsenault and Kochetov (forthcoming) also support the correspondence approach in their analysis of sibilant and retroflex harmony in Kalasha. They argue that since coronal harmony in Kalasha is restricted to apply only between consonants with the same manner of articulation, this

[9] Krämer (2001, 2003) develops a surface correspondence approach for vowel harmony, with adjacency defined at a moraic or syllabic level. Pulleyblank (2002) offers a different perspective that accounts for both vowel and consonant harmony using a "no-disagreement" harmony-driver (see also Archangeli and Pulleyblank 2007).

lends support to the correspondence approach, which formally encodes similarity. Spreading approaches would need to explain why harmony operates only between consonants of like manner.

Gallagher and Coon (2009) nevertheless argue that correspondence is appropriate for harmonies that require complete identity between consonants, but not for those that induce limited featural agreement, such as most sibilant harmonies. Gallagher and Coon focus on harmony data from Chol, a Mayan language of Mexico. The Chol pattern is an interaction between laryngeal harmony and coronal strident harmony. Total identity (25a) between consonants is required in two cases: (i) two ejectives in a root or (ii) two plain stridents. If the two consonants differ in terms of laryngeal features (ejective and plain), then only strident harmony is enforced (25b).

(25) a. *Total identity*

Plain stridents			Ejectives		
*ts-s	sus	'scrape'	*k'-p'	k'ok'	'healthy'
*s-ʧ	ʧiʧ	'older sister'	*t'-ts'	t'ot'	'snail'
				ʧ'iʧ'	'blood'

b. *Strident harmony*

*ts'-ʃ	ts'is	'sew'
*s-ʧ'	ʃuʰʧ'	'thief'

Strident harmony is always enforced, regardless of laryngeal specification, but laryngeal harmony requires complete identity. Gallagher and Coon's analysis requires that similar consonants (those that share certain features) are "linked" (i.e. correspond), and an identity constraint requires them to be completely identical. Ejectivity renders consonants more similar than stridency. Although the proposal accounts for the particular case of total identity seen in Mayan languages,[10] it does not extend to other cases of consonant harmony outlined in §2, which show partial identity effects, but cannot receive a spreading analysis due to the transparency of the intervening segments. Moreover, it is not clear why ejectivity in particular requires a total identity between segments. Laryngeal harmonies are often restricted to a subset of obstruents (stops) and homorganicity is also frequently involved. This signals that more research on how to define similarity is required.

In the arena of blocking, Hansson (2007b) argues that while lack of blocking is a descriptive characteristic of consonant harmony systems, it does not necessarily follow from the agreement-by-correspondence approach. Blocking could arise in scenarios in which three segments are in correspondence, but in different correspondence relationships, either different local relationships for the same feature or different featural relationships. More research is necessary to determine whether such scenarios are actually attested, or whether the correspondence analysis requires modification. If attested, this would undermine one of the strong arguments for why consonant harmony should not receive the same analysis as vowel harmony or vowel-consonant harmony.

[10] Indeed, the analysis predicts that combinations such as /ts ts'/ are acceptable as they disagree for ejectivity, when in fact they are not attested.

Despite various criticisms, the correspondence approach to harmony has stimulated new areas of research in the analysis of long-distance assimilations and has pushed researchers to examine languages in more detail, to conduct corpus studies of morpheme structure constraints, and to investigate harmony from an experimental angle.

3.3 *The role of contrast in consonant harmony*

The concept of *contrast* (CHAPTER 2: CONTRAST) has long played a role in autosegmental spreading analyses of consonant harmony, specifically in determining feature specification. Steriade (1987b) and Shaw (1991) rely on the fact that only sibilants contrast for [anterior] to explain the transparency of stops and sonorants to sibilant harmony. In the correspondence model of harmony, the role of contrast is not emphasized. Although Hansson (2001a) and Rose and Walker (2004) mention contrast, its role in determining harmony systems does not figure prominently in their model, except indirectly via the natural classes model of computing similarity (Frisch 1996; Frisch *et al.* 2004), which Rose and Walker (2004) adopt. Yet recent research has returned to the issue of contrast, both as a means of constraining harmony, and in promoting contrast as the driving force behind consonant harmony.

Mackenzie (2005, 2009) argues that similarity in consonant harmony should be formalized on the basis of contrastive featural specifications determined by a language's inventory. Segments that are similar to one another in their contrastive specifications, not necessarily segments that are most similar phonetically, interact in harmony. In the harmony system of Bumo Izon, an Ijoid language, voiced implosives and voiced stops may not co-occur, with two exceptions: /g/ and /ɡ͡ɓ/ (Efere 2001); these sounds freely co-occur with stops of the opposite pulmonic value: [ɗúgó] 'to pursue' or [ɡ͡ɓóɗaɡ͡ɓóda] 'rain (hard)'. Mackenzie points out that these sounds lack a contrastive counterpart: there is no /ɠ/ or /gb/. The natural classes model (Frisch *et al.* 2004) of calculating similarity does take phonemic inventory into account when computing natural classes and similarity, but it fares poorly with asymmetric inventories in which contrastive counterparts are missing. Hansson (2004) notes this problem with respect to laryngeal harmony in Ngizim, where implosives do not participate. Mackenzie's solution is to determine similarity via pairwise contrasts that are partially language-specific. If two sounds are not specified for a feature due to lack of contrast, they do not participate in the harmony, which references presence of features.

Hansson (2008) examines the role of contrast in the typology of vowel harmony and consonant harmony and notes that only consonant harmony has cases of "symmetric neutralization," in which a lexical [±feature] contrast emerges in affixes only with neutral roots. The regressive sibilant harmony systems of Chumash and Navajo involve neutralization of contrasts on both roots and affixes, as either /s/ or /ʃ/ can trigger harmony as long as it is the rightmost sibilant in the word. Hansson argues that systems of this type are attested in consonant harmony, but not vowel harmony, because they are *recoverable*. The loss of contrast between /s/ and /ʃ/ is minimal in the consonant inventory, and affects only a small subclass of consonants. The learner has a large number of contexts provided by neutral roots to compensate for neutralization. Hansson (2001a) notes that the specific combination of symmetric neutralization with absolute directionality of assimilation creates problems for some models of phonology, such as Declarative Phonology and standard Optimality Theory.

Finally, Gallagher (2008, 2010) argues for a notion of laryngeal contrast rooted in dispersion theory with a more global view of contrast within a language's lexicon. A constraint, Laryngeal Distance, penalizes contrasts between roots that have only one laryngeally marked stop *vs.* those with two. Plain stops are unmarked for laryngeal features. Gallagher argues that the distinction between a root with one laryngeal feature and a root with two laryngeal features is perceptually weak. Avoidance can play out as harmony (only two ejectives [k'ap'a] or two plain stops [kapa] are allowed) or as dissimilation (only one plain and one ejective are allowed [k'apa] or two plain [kapa]).

In conclusion, contrast remains a powerful and debated concept in the study of consonant harmony, one that is sure to resonate in future research.

4 Experimental approaches to consonant harmony

As debate about the most appropriate analysis of consonant harmony has come to center on hypotheses about its grounding in articulation or perception, experimental studies of consonant harmony have been conducted (CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY). The correspondence approach to consonant harmony proposes that harmony is grounded in production difficulties caused by phonological planning and the similarity of interacting consonants. Several experiments have been undertaken to test this hypothesis in the area of speech errors.

Walker (2007) conducted an experimental study inducing speech errors. Consonants that were more similar and known to participate in nasal harmony, such as nasals and voiced stops, were predicted to be more prone to speech errors than other combinations. Nonce words with combinations of nasals and voiced stops and nasals and voiceless stops were tested with English speakers, as English is not reported to have nasal consonant harmony. Indeed, more errors arose with nasal–voiced stop combinations than nasal–voiceless. Walker concluded that nasal harmony could indeed be grounded in difficulties with the production of similar sounds.

Kochetov and Radišić (2009) performed a similar experiment on combinations of four sibilant fricatives /s sʲ ʃ ʃʲ/ in Russian in a repetition task performed at a fast rate of speech. Errors (assessed by examining acoustic effects of production) were observed for both primary place of articulation and secondary articulation. The primary-place assimilation errors were generally regressive and involved /s/ changing to [ʃ], reflective of the "palatal bias" effect reported in other speech-error studies on English. Although Russian is not reported to have sibilant consonant harmony, the speech-error effect is similar to that found in harmony languages, supporting Hansson's (2001a, 2001b) observation of the correlation. However, Kochetov and Radišić (2009) also note that consonants differing only in secondary articulation did not participate in as many errors and that those errors were progressive. This seems to lend support to Hansson's (2007a) contention that speech production difficulties may not underlie secondary articulation harmonies. Kochetov and Radišić (2009) speculate that feature spreading or gestural extension may be a better analysis for these cases, paralleling vowel harmony.

Rose and King (2007) examined the impact of harmony constraints on speech errors in languages observed to have laryngeal harmony, namely Chaha and Amharic. They found higher speech-error rates for certain sequences that violated laryngeal harmony than for those that did not. In particular, the researchers compared the laryngeal pairs with consonant pairs that were also highly similar and infrequent in verb roots, but did not violate any constraints. These pairs did not show high error rates in comparison. Rose and King conclude that laryngeal harmony is not only based on production difficulties, but also, once encoded grammatically, triggers more errors when speakers encounter sequences that violate it.

Walker *et al.* (2008) investigated coronal harmony in Kinyarwanda by means of electromagnetic articulography. Kinyarwanda exhibits retroflex harmony, previously reported in the literature as an alveolar–post-alveolar sibilant harmony (Walker and Mpiranya 2006). Harmony is blocked by alveolar stops and affricates, retroflex stops, and palatal consonants. Intervening vowels and non-coronal consonants do not block the harmony and are not perceptibly affected. The blocking effect is suggestive of spreading, while the transparent vowels and non-coronals point to a correspondence analysis. This is important since retroflex harmony is recognized as both a type of consonant harmony (Arsenault and Kochetov, forthcoming) and a possible vowel–consonant harmony for which spreading may be a more appropriate analysis than correspondence (Gafos 1999; Hansson 2001a; Rose and Walker 2004). Walker *et al.* (2008) found evidence that the harmonizing retroflex posture persists during apparent transparent non-coronal consonants when they occur between harmonizing fricatives. Such a result is more supportive of a spreading or gestural analysis, in line with Gafos's (1999) Articulatory Locality model. Results were not conclusive for the intervening coronals. Research of this nature should be conducted on languages that have robust non-retroflex sibilant harmony to help address the question of whether spreading or correspondence is a more appropriate analysis. At the same time, this raises the issue of whether gradient phonetic articulatory results should be used to determine phonological representations.

Finally, Gallagher (2010) utilizes perceptual experiments to test the validity of her contrastive perceptual distance model of laryngeal harmony. American English subjects who listened to pairs of Bolivian Quechua words with combinations of ejectives and plain stops had the greatest difficulty perceiving contrasts between two words where an ejective and a plain stop contrasted with two ejectives ([k'apa] *vs.* [k'ap'a]). The two harmonic forms wherein two ejectives contrast with two plain stops ([k'ap'a] *vs.* [kapa]) were the easiest to perceive, with [k'apa] *vs.* [kapa] occupying an intermediary position. It is argued that these results provide support for a perceptual motivation for consonant harmony, with harmony viewed as a response to avoid difficult perceptual contrasts.

Experimental research in consonant harmony using a variety of techniques may help illuminate the causes of harmony (perceptual, articulatory, or both) and the best phonological analysis of this phenomenon. It may help sort out whether consonant harmony should be viewed as a unified phenomenon or as several disparate phenomena that share the common characteristic of assimilation at a distance.

5 Conclusion

Consonant harmony has intrigued researchers for many years, due to its tantalizing similarities to other types of harmony. Recognizing what distinguishes it from other harmony systems has nevertheless pushed analysis in new directions. Two competing approaches have been advanced: *spreading* of features or gestural extension and *correspondence* between segments requiring matching for features. Both analyses have positive attributes, but both are not without challenges. It is also possible that spreading is appropriate for some harmony systems but correspondence for others, as has been suggested by different researchers (Hansson 2001a; Gallagher and Coon 2009; Kochetov and Radišić 2009). Research in experimental directions may help shed light on which analysis is ultimately correct and whether altogether new analyses will eventually emerge.

ACKNOWLEDGMENTS

Thanks to Beth Hume, Marc van Oostendorp, and two anonymous reviewers for their instructive comments on various aspects of this chapter. All errors are my own.

REFERENCES

Alderete, John & Stefan A. Frisch. 2007. Dissimilation in grammar and the lexicon. In de Lacy (2007), 379–398.

Andersen, Torben. 1988. Consonant alternation in the verbal morphology of Päri. *Afrika und Übersee* 71. 63–113.

Andersen, Torben. 1999. Consonant alternation and verbal morphology in Mayak (Northern Burun). *Afrika und Übersee* 82. 65–97.

Ao, Benjamin. 1991. Kikongo nasal harmony and context-sensitive underspecification. *Linguistic Inquiry* 22. 193–196.

Applegate, Richard B. 1972. Ineseño Chumash grammar. Ph.D. dissertation, University of California, Berkeley.

Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5. 183–207.

Archangeli, Diana & Douglas Pulleyblank. 1987. Minimal and maximal rules: Effects of tier scansion. *Papers from the Annual Meeting of the North East Linguistic Society* 17. 16–35.

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Archangeli, Diana & Douglas Pulleyblank. 2007. Harmony. In de Lacy (2007), 353–378.

Arsenault, Paul & Alexei Kochetov. Forthcoming. Retroflex harmony in Kalasha: Agreement or spreading? *Papers from the Annual Meeting of the North East Linguistic Society* 39.

Berg, Thomas & Ulrich Schade. 2000. A local connectionist account of consonant harmony in child language. *Cognitive Science* 24. 123–149.

Booysen, Jacobus M. 1982. *Otjiherero: 'n Volledige grammatika met oefeninge en sleutels in Afrikaans*. Windhoek: Gamsberg.

Breeze, Mary. 1990. A sketch of the phonology and grammar of Gimira (Benchnon). In Richard Hayward (ed.) *Omotic language studies*, 1–67. London: School of Oriental and African Studies.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6. 201–251.

Browman, Catherine P. & Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341–376. Cambridge: Cambridge University Press.

Brown, Jason. 2008. Theoretical issues in Gitksan phonology. Ph.D. dissertation, University of British Columbia.

Burrow, Thomas & Sudhibhushan Bhattacharya. 1970. *The Pengo language: Grammar, texts, and vocabulary*. Oxford: Clarendon Press.

Clements, G. N. 1980. *Vowel harmony in nonlinear generative phonology: An autosegmental model*. Indiana University Linguistics Club.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. 1991. Place of articulation in consonants and vowels: A unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5. 77–123.

Clements, G. N. 2001. Representational economy in constraint-based phonology. In T. A. Hall (ed.) *Distinctive feature theory*, 71–146. Berlin & New York: Mouton de Gruyter.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

Cohn, Abigail C. 1992. The consequences of dissimilation in Sundanese. *Phonology* 9. 199–220.

Cook, Eung-Do. 1983. Chilcotin flattening. *Canadian Journal of Linguistics* 28. 123–132.

Cook, Eung-Do. 1993. Chilcotin flattening and autosegmental phonology. *Lingua* 91. 149–174.

Davidson, Joseph. 1977. A contrastive study of the grammatical structures of Aymara and Cuzco Kechua. Ph.D. dissertation, University of California, Berkeley.

Dayley, Jon P. 1985. *Tzutujil grammar*. Berkeley: University of California Press.

de Lacy, Paul (ed.) 2007. *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press.

Dell, Gary S., Lisa K. Burger & William R. Svec. 1997. Language production and serial order: A functional analysis and a model. *Psychological Review* 104. 123–147.

Dereau, Léon. 1955. *Cours de kikongo*. Namur: A. Wesmael-Charlier.

Ebert, Keren. 1979. *Sprache und Tradition der Kera (Tschad)*, vol. 3: *Grammatik*. Berlin: Reimer.

Efere, Emmanuel. 2001. The pitch system of the Bumo dialect of Izon. *UBC Working Papers in Linguistics* 4. 115–259.

Frisch, Stefan A. 1996. Frequency and similarity in phonology. Ph.D. dissertation, Northwestern University.

Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22. 179–228.

Fromkin, Victoria A. 1971. The non-anomalous nature of anomalous utterances. *Language* 47. 27–52.

Gafos, Adamantios I. 1998. Eliminating long-distance consonantal spreading. *Natural Language and Linguistic Theory* 16. 223–278.

Gafos, Adamantios I. 1999. *The articulatory basis of locality in phonology*. New York: Garland.

Gallagher, Gillian. 2008. The role of contrast in laryngeal cooccurrence restrictions. *Proceedings of the West Coast Conference on Formal Linguistics* 27. 177–184.

Gallagher, Gillian. 2010. The perceptual basis of long-distance laryngeal restrictions. Ph.D. dissertation, MIT.

Gallagher, Gillian & Jessica Coon. 2009. Distinguishing total and partial identity: Evidence from Chol. *Natural Language and Linguistic Theory* 27. 545–582.

Goad, Heather. 1997. Consonant harmony in child language: An optimality-theoretic account. In S. J. Hannahs & Martha Young-Scholten (eds.) *Focus on phonological acquisition*, 113–142. Amsterdam & Philadelphia: John Benjamins.

Goldsmith, John A. 1979. *Autosegmental phonology*. New York: Garland.
Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.
Goldstein, Louis, Marianne Pouplier, Larissa Chen, Elliot L. Saltzman & Dani Byrd. 2007. Dynamic action units slip in speech production errors. *Cognition* 103. 386–412.
Halle, Morris & Jean-Roger Vergnaud. 1981. Harmony processes. In Wolfgang Klein & Willem J. Levelt (eds.) *Crossing the boundaries in linguistics*, 1–22. Dordrecht: Reidel.
Hamp, Eric P. 1976. Palatalization and harmony in Gagauz and Karaite. In Walther Heissig, John R. Krueger, Felix J. Oinas & Edmond Schütz (eds.) *Tractata altaica: Denis Sinor, sexagenario optime de rebus altaicis merito dedicata*, 211–213. Wiesbaden: Harrassowitz.
Hansson, Gunnar Ólafur. 2001a. Theoretical and typological issues in consonant harmony. Ph.D. dissertation, University of California, Berkeley.
Hansson, Gunnar Ólafur. 2001b. The phonologization of production constraints: Evidence from consonant harmony. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 37. 187–200.
Hansson, Gunnar Ólafur. 2004. Tone and voicing agreement in Yabem. *Proceedings of the West Coast Conference on Formal Linguistics* 23. 318–331.
Hansson, Gunnar Ólafur. 2007a. On the evolution of consonant harmony: The case of secondary articulation agreement. *Phonology* 24. 77–120.
Hansson, Gunnar Ólafur. 2007b. Blocking effects in agreement by correspondence. *Linguistic Inquiry* 38. 395–409.
Hansson, Gunnar Ólafur. 2008. Effects of contrast recoverability on the typology of harmony systems. In Paul Avery, B. Elan Dresher & Keren Rice (eds.) *Contrast in phonology: Perception and acquisition*, 55–81. Berlin & New York: Mouton de Gruyter.
Hardman, Martha James, Juana Vásquez & Juan de Dios Yapita (eds.) 1974. *The Aymara language project*, vol. 3: *Outline of Aymara phonological and grammatical structure*. Gainesville: Department of Anthropology, University of Florida.
Henderson, John. 1998. Topics in Eastern and Central Arrernte grammar. Ph.D. dissertation, University of Western Australia.
Hulst, Harry van der & Norval Smith. 1982a. Prosodic domains and opaque segments in autosegmental theory. In van der Hulst & Smith (1982b: part II), 311–336.
Hulst, Harry van der & Norval Smith (eds.) 1982b. *The structure of phonological representations*. 2 parts. Dordrecht: Foris.
Hulst, Harry van der & Norval Smith (eds.) 1988. *Features, segmental structure and harmony processes*. 2 parts. Dordrecht: Foris.
Hyman, Larry M. 1995. Nasal consonant harmony at a distance: The case of Yaka. *Studies in African Linguistics* 24. 5–30.
Hyman, Larry M. 2002. Is there a right-to-left bias in vowel harmony? Paper presented at the 9th International Phonology Meeting, Vienna.
Inkelas, Sharon & Yvan Rose. 2008. Positional neutralization: A case study from child language. *Language* 83. 707–736.
Itô, Junko & Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17. 49–73.
Jenewari, Charles E. W. 1989. Ijoid. In John Bendor-Samuel (ed.) *The Niger-Congo languages*, 105–118. Lanham, MD: University Press of America.
Kawachi, Kazuhiro. 2007. A grammar of Sidaama (Sidamo), a Cushitic language of Ethiopia. Ph.D. dissertation, State University of New York, Buffalo.
Khumalo, James S. M. 1987. An autosegmental account of Zulu phonology. Ph.D. dissertation, University of Witwatersrand, Johannesburg.
Kisseberth, Charles W. & Mohammad Imam Abasheikh. 1975. The perfect stem in Chimwi:ni and global rules. *Studies in African Linguistics* 6. 249–266.
Kochetov, Alexei & Milica Radišić. 2009. Latent consonant harmony in Russian: Experimental evidence for agreement by correspondence. In Maria Babyonyshev, Darya Kavitskaya

& Jodi Reich (eds.) *Proceedings of the 17th Formal Approaches to Slavic Linguistics Meeting*, 111–130. Ann Arbor: Michigan Slavic Publications.

Kowalski, Tadeusz. 1929. *Karaimische texte im Dialekt von Troki*. Krakow: Polska Akademia Umiejętności.

Krämer, Martin. 2001. Yucatec Maya vowel alternations: Harmony as syntagmatic identity. *Zeitschrift für Sprachwissenschaft* 20. 175–217.

Krämer, Martin. 2003. *Vowel harmony and correspondence theory*. Berlin & New York: Mouton de Gruyter.

Lemb, P. & F. de Gastines. 1973. *Dictionnaire Basaa–Français*. Douala: College Libermann.

Lieber, Rochelle. 1987. *An integrated theory of autosegmental processes*. Albany: SUNY Press.

Lucca, Manuel de. 1987. *Diccionario práctico aymara–castellano castellano-aymara*. La Paz: Editorial Los Amigos del Libro.

MacEachern, Margaret R. 1996. Ordering restrictions in aspirated and ejective stops in Aymara. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 22. 244–254.

MacEachern, Margaret R. 1999. *Laryngeal cooccurrence restrictions*. New York: Garland.

Mackenzie, Sara. 2005. Similarity and contrast in consonant harmony systems. *Toronto Working Papers in Linguistics* 24. 169–182.

Mackenzie, Sara. 2009. Contrast and similarity in consonant harmony processes. Ph.D. dissertation, University of Toronto.

Mahapatra, B. P. 1979. *Malto: An ethnosemantic study*. Manasagangotri: Central Institute of Indian Languages.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John J. 2007. Consonant harmony via correspondence: Evidence from Chumash. In Leah Bateman, Michael O'Keefe, Ehren Reilly & Adam Werle (eds.) *Papers in Optimality Theory 3*, 223–237. Amherst: GLSA.

McDonough, Joyce. 1990. Topics in the phonology and morphology of Navajo verbs. Ph.D. dissertation, University of Massachusetts, Amherst.

McDonough, Joyce. 1991. On the representation of consonant harmony in Navajo. *Proceedings of the West Coast Conference on Formal Linguistics* 10. 319–335.

Mester, Armin. 1988. Dependent tier ordering and the OCP. In van der Hulst & Smith (1988: part II), 127–144.

Nater, Hank F. 1989. Some comments on the phonology of Tahltan. *International Journal of American Linguistics* 55. 24–42.

Nevins, Andrew & Bert Vaux. 2004. Consonant harmony in Karaim. In Aniko Csirmaz, Youngjoo Lee & Mary Ann Walter (eds.) *Proceedings of the 1st Workshop on Altaic Formal Linguistics*, 175–194. Cambridge, MA: Department of Linguistics, MIT.

Ní Chiosáin, Máire & Jaye Padgett. 1997. Markedness, segment realization, and locality in spreading. Technical Report LRC-97-01, Linguistics Research Center, University of California, Santa Cruz.

Ní Chiosáin, Máire & Jaye Padgett. 2001. Markedness, segment realization, and locality in spreading. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 118-156. Cambridge: Cambridge University Press.

Odden, David. 1994. Adjacency parameters in phonology. *Language* 70. 289–330.

Palmer, Frank (ed.) 1970. *Prosodic analysis*. Oxford: Oxford University Press.

Paradis, Carole & Jean-François Prunet. 1989. On coronal transparency. *Phonology* 6. 317–348.

Pater, Joe & Adam Werle. 2003. Direction of assimilation in child consonant harmony. *Canadian Journal of Linguistics* 48. 385–408.

Pearce, Mary. 2005. Kera tone and voicing. *UCL Working Papers in Linguistics* 17. 61–82.

Piggott, Glyne L. 1988. A parametric approach to nasal harmony. In van der Hulst & Smith (1988: part I), 131–167.

Piggott, Glyne L. 1992. Variability in feature dependency: The case of nasality. *Natural Language and Linguistic Theory* 10. 33–77.

Piggott, Glyne L. 1996. Implications of consonant nasalization for a theory of harmony. *Canadian Journal of Linguistics* 41. 141–174.

Piggott, Glyne L. 2003. Theoretical implications of segment neutrality in nasal harmony. *Phonology* 20. 375–424.

Piggott, Glyne L. & Harry van der Hulst. 1997. Locality and the nature of nasal harmony. *Lingua* 103. 85–112.

Poser, William J. 1982. Phonological representations and action-at-a-distance. In van der Hulst & Smith (1982: part II), 121–158.

Poser, William J. 1993. Are strict cycle effects derivable? In Sharon Hargus & Ellen M. Kaisse (eds.) *Studies in Lexical Phonology*, 315–321. San Diego: Academic Press.

Pouplier, Marianne. 2007. Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech* 50. 311–341.

Pulleyblank, Douglas. 2002. Harmony drivers: No disagreement allowed. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 28. 249–267.

Reh, Mechthild. 1996. *Anywa language: Description and internal reconstructions*. Cologne: Rüdiger Köppe Verlag.

Rehg, Kenneth L. & Damian G. Sohl. 1981. *Ponapean reference grammar*. Honolulu: University of Hawai'i Press.

Rose, Sharon & Lisa King. 2007. Speech error elicitation and co-occurrence restrictions in two Ethiopian Semitic languages. *Language and Speech* 50. 451–504.

Rose, Sharon & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80. 475–531.

Rose, Sharon & Rachel Walker. Forthcoming. Harmony systems. In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.) *The handbook of phonological theory*. 2nd edn. Malden, MA & Oxford: Wiley-Blackwell.

Rose, Yvan. 2000. Headedness and prosodic licensing in the L1 acquisition of phonology. Ph.D. dissertation, McGill University.

Rose, Yvan & Christophe dos Santos. 2006. Facteurs prosodiques et articulatoires dans l'harmonie consonantique et la métathèse en acquisition du français langue première. *Recherches linguistiques de Vincennes* 35. 77–102.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.

Schuh, Russell. 1997. Changes in obstruent voicing in Bade/Ngizim. Unpublished ms., University of California, Los Angeles.

Shahin, Kimary. 2002. *Postvelar harmony*. Amsterdam & Philadelphia: John Benjamins.

Shattuck-Hufnagel, Stefanie & Dennis H. Klatt. 1979. The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior* 18. 41–55.

Shaw, Patricia A. 1991. Consonant harmony systems: The special status of coronal harmony. In Carole Paradis & Jean-François Prunet (eds.) *The special status of coronals: Internal and external evidence*, 125–157. San Diego: Academic Press.

Stafford, R. L. 1967. *An elementary Luo grammar*. Oxford: Oxford University Press.

Steriade, Donca. 1987a. Locality conditions and feature geometry. *Papers from the Annual Meeting of the North East Linguistic Society* 17. 595–617.

Steriade, Donca. 1987b. Redundant values. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 339–362.

Steriade, Donca. 1995. Underspecification and markedness. In Goldsmith (1995), 114–174.

Straight, H. Stephen. 1976. *The acquisition of Maya phonology: Variation in Yucatec child language*. New York: Garland.

Tabain, Marija & Kristine Rickard. 2007. A preliminary EPG study of stop consonants in Arrernte. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 501–504. Saarbrücken: Saarland University.

Trail, Ronald L. & Gregory R. Cooper. 1999. *Kalasha dictionary – with English and Urdu.* Islamabad: National Institute of Pakistan Studies & Summer Institute of Linguistics.

Traill, Anthony, James S. M. Khumalo & Paul Fridjhon. 1987. Depressing facts about Zulu. *African Studies* 46. 255–274.

Tucker, Archibald N. 1994. *A grammar of Kenya Luo (Dholuo).* Edited by Chet Creider. Cologne: Rüdiger Köppe Verlag.

Vaux, Bert. 1998. The laryngeal specifications of fricatives. *Linguistic Inquiry* 29. 497–511.

Vihman, Marilyn M. 1978. Consonant harmony: Its scope and function in child language. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 2: *Phonology*, 281–334. Stanford: Stanford University Press.

Viljoen, Johannes J. 1973. *Manual for Ndonga*, part 1. Pretoria: University of South Africa.

Walker, Rachel. 2000a. *Nasalization, neutral segments, and opacity effects.* New York: Garland.

Walker, Rachel. 2000b. Yaka nasal harmony: Spreading or segmental correspondence? *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 26. 321–332.

Walker, Rachel. 2000c. Long-distance consonantal identity effects. *Proceedings of the West Coast Conference on Formal Linguistics* 19. 532–545.

Walker, Rachel. 2003. Reinterpreting transparency in nasal harmony. In Jeroen van de Weijer, Vincent J. van Heuven & Harry van der Hulst (eds.) *The phonological spectrum*, vol. 1: *Segmental structure*, 37–72. Amsterdam & Philadelphia: John Benjamins.

Walker, Rachel. 2007. Nasal and oral consonant similarity in speech errors: Exploring parallels with nasal consonant harmony. *Language and Cognitive Processes* 22. 1073–1113.

Walker, Rachel & Fidèle Mpiranya. 2006. On triggers and opacity in coronal harmony. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 31. 383–394.

Walker, Rachel & Geoffrey K. Pullum. 1999. Possible and impossible segments. *Language* 75. 764–780.

Walker, Rachel, Dani Byrd & Fidèle Mpiranya. 2008. An articulatory view of Kinyarwanda coronal harmony. *Phonology* 25. 499–535.

Watson, Janet C. E. 1999. The directionality of emphasis spread in Arabic. *Linguistic Inquiry* 30. 289–300.

Watters, James. 1988. Topics in Tepehua grammar. Ph.D. dissertation, University of California, Berkeley.

Yip, Moira. 1989. Feature geometry and cooccurrence restrictions. *Phonology* 6. 349–374.

Younes, Munther. 1993. Emphasis spread in two Arabic dialects. In Mushira Eid & Clive Holes (eds.) *Perspectives on Arabic linguistics V*, 119–145. Amsterdam & Philadelphia: John Benjamins.

Zawaydeh, Bushra. 1999. The phonetics and phonology of gutturals in Arabic. Ph.D. dissertation, Indiana University.

Zuraw, Kie. 2002. Aggressive reduplication. *Phonology* 19. 395–439.

78 Nasal Harmony

Rachel Walker

Nasal harmony refers to phonological patterns where nasalization is transmitted in long-distance fashion. The long-distance nature of nasal harmony can be met by the transmission of nasalization either to a series of segments or to a non-adjacent segment. Nasal harmony usually occurs within words or a smaller domain, either morphologically or prosodically defined. This chapter introduces the chief characteristics of nasal harmony patterns with exemplification, and highlights related theoretical themes. It focuses primarily on the different roles that segments can play in nasal harmony, and the typological properties to which they give rise. The following terminological conventions will be assumed. A *trigger* is a segment that initiates nasal harmony. A *target* is a segment that undergoes harmony. An *opaque segment* or *blocker* halts nasal harmony. A *transparent segment* is one that does not display nasalization within a span of nasal harmony, but does not halt harmony from transmitting beyond it.

Three broad categories of nasal harmony are considered in this chapter. They are (i) nasal vowel–consonant harmony with opaque segments, (ii) nasal vowel–consonant harmony with transparent segments, and (iii) nasal consonant harmony. Each of these groups of systems show characteristic hallmarks. Nasal vowel–consonant harmony refers to systems in which vowels and consonants participate in the pattern as either triggers or targets. Sundanese manifests one of the most limited forms of nasal vowel–consonant harmony, where vowels and laryngeal segments are targets of harmony. Other systems are more liberal in the scope of their targets, including some or all sonorant consonants, and sometimes even obstruents. In many patterns of nasal vowel–consonant harmony, Sundanese included, the segments that are not targets are blockers. In other systems, some or all obstruents are transparent, while vowels, laryngeals and sonorant consonants are targets. On the other hand, in nasal consonant harmony, only consonants participate, and vowels are transparent. Certain categories of consonants are also transparent in cases where nasal consonant harmony can occur at any distance within its domain. Nasal stops trigger harmony in these systems, and the targets are typically drawn from the set of stops and/or approximant consonants.

A core topic of theoretical debate revolves around whether systems of nasal vowel–consonant harmony with opaque segments and those with transparent segments should be analyzed as the same or different with respect to the driving force for harmony and the types of representations involved. Further, the distinct properties of nasal consonant harmony and nasal vowel–consonant harmony have led to proposals about their formal differences, such as the level at which the feature [nasal] spreads in the structure; more recently, an approach has been advocated in which consonant harmony is achieved through relations constructed between similar segments rather than feature spreading. Signposted throughout this chapter will be many other issues in phonology with which the topic of nasal harmony intersects. For instance, nasal harmony has played a significant role in discussions that concern the grounding of phonological patterns in phonetic principles, contrast and speech planning. The nature of locality in phonology is another primary theme. Directionality effects are also examined in this chapter.

1 Nasal vowel–consonant harmony with opaque segments

Nasal vowel–consonant harmony refers to nasal harmony in which both vowels and consonants participate as triggers and/or targets (see CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). A central typological observation that emerges from cross-linguistic studies is that nasal vowel–consonant harmony with opaque segments respects the implicational scale in (1). Categories further to the right are progressively disfavored as targets, and the inclusion of any category as a target in a harmony pattern implies that all categories to its left will be targets (Schourup 1972; Piggott 1992, 2003; Cohn 1993a, 1993b; Walker and Pullum 1999; Walker 2000a; see Hume and Odden 1996 for related discussion). The category "laryngeals" refers to [h ʔ].

(1) vowels > laryngeals > glides > liquids > fricatives > obstruent stops

The hierarchy in (1) summarizes a pervasive generalization that emerges from nasal vowel–consonant harmony patterns across languages. However, some fine-tuning may be needed, as discussed in connection with particular cases below. The hierarchy has been suggested to have a phonetic basis, whereby nasalization of segments that are lower on the scale is disfavored for reasons of articulation, aerodynamics, and/or perceptibility (see Walker 2000a and references therein). Maintenance of a system of contrasts has also been suggested to underlie the hierarchy in (1) (e.g. Flemming 2004). This scale shows a striking similarity to the sonority hierarchy, on which see CHAPTER 49: SONORITY. Nevertheless, they are distinct in the ranking of nasal stops, which are usually situated between liquids and obstruents in the sonority hierarchy, but could plausibly be located at the top left of the scale in (1). Furthermore, consensus is lacking on the placement of laryngeals in the sonority scale. I return to this issue later in this section, in the context of the impedance hierarchy proposed by Hume and Odden (1996); for further discussion, see Cohn (1993a), Gnanadesikan (1995), Boersma (1998, 2003), and Walker (2000a).

Following the exemplification of hierarchical effects in the typology, some ways of formalizing them are discussed. The exemplification of nasal vowel–consonant harmony with opaque segments begins with patterns that show a narrower set of blockers and progresses to ones with blocking by more categories.

The dialect of Scottish Gaelic spoken in Applecross (Ross-shire), henceforth "Applecross Gaelic," displays a nasal harmony where nasalization is reported to target all segment categories except obstruent stops (Ternes 2006). Nasalization spreads from a stressed nasal vowel. Stress is usually – but not always – assigned to the initial syllable. Progressive nasalization is halted by an obstruent stop. In addition, consonants in the onset of a syllable with a stressed nasal vowel are nasalized except obstruent stops.

(2)	a.	*Monosyllables*		
		l̃ˠɔ̃ũ·m	'bare, naked'	
		l̃ˠãːɟ̃	'hand'	
		mĩã·r̃	'finger, toe'	*lenited* ṽĩã·r̃
		kʰr̃ɛ̃ãk	'maggot'	
		str̃ãĩ·ɣ̃	'string'	
		tãɣ̃	'●x, stag'	
	b.	*Polysyllabic words*		
		'f̃r̃ĩãũn	'root (PL)'	
		'kãnã·x̃	'sand'	
		'l̃ˠã·h̃ũk	'axe, hatchet'	
		's̃ɔ̃h̃ĩ	'tame'	
		'tʰr̃ĩĩʃ̃ãr̃	'plate'	
		'ʃ̃ɛ̃nɛ̃·ṽãr̃	'grandmother'	
		'kʰɔ̃ĩs̃paxk	'wasp'	
		'mĩnĩʃtʲar	'minister (clergyman)'	
		's̃nˠã·nʲdʲan	'thread'	

Transcriptions of nasalized fricatives in Applecross Gaelic follow Ternes; however, the realization of such segments in general have been the subject of debate. Nasalized fricatives present an aerodynamic confound with consequences for perception: an increase in velopharyngeal opening tends to reduce frication and a decrease in velopharyngeal aperture can reduce perceptible nasalization (CHAPTER 28: THE REPRESENTATION OF FRICATIVES). Gerfen (1999, 2001) has brought instrumental research to this question for the nasal harmony of Coatzospan Mixtec. See Shosted (2006) and Solé (2007a) for recent reviews of the issues and experimental investigations. Whether and how the gradient trade-offs in realization should be represented in phonology remains an open question.

Color and height distinctions in nasal vowels of Applecross Gaelic are a proper subset of those in its oral vowels. The mid-high vowel series /e ə o/ is always oral, and, like the obstruent stops, they block nasal harmony. Examples with blocking by [ə] are given in (3). All other vowel qualities in the language ([i ɨ u ɛ ɔ a]) can be phonemically oral or nasal and can become nasalized through nasal harmony.

(3)	's̃ãũl̃ˠəxkənʲ	'to compare'
	'ãɟ̃əlˠ	'angel'
	'nʲĩən	'girl, daughter'
	kʰa'tʰr̃ĩãnə	'Catherine'

Blocking of nasal harmony by mid vowels is also attested in Mọ̀bà Yoruba, discussed in §2. The resistance of mid-high vowels to contrastive or contextual nasalization has plausible origins in effects of nasalization on the perception of vowel height. The perceived distance of a height distinction between two oral vowels is reduced when those same vowels are nasalized (e.g. Wright 1986), which could give rise to the restriction of nasalization to a subset of the oral vowel heights in Applecross Gaelic (Homer 1998; Walker 2000a). In general, the number of nasal vowels in a language never exceeds the number of oral vowels, and it is relatively common for one or more mid nasal vowels to be missing relative to the oral inventory in a language. Nasalization may interfere most with detection of height distinctions involving mid vowels; also perceptual integration of nasalization and height in vowels disfavors mid percepts (Kingston 2007). This suggests the possibility that the descriptive hierarchy in (1) could be moderated by or interact with the effects of contrast, an issue to which I return later in this section.

Given that Applecross Gaelic has nasal stops, the question arises whether they too can trigger nasal harmony. When they occur in the onset to a stressed syllable with an oral vowel, they do not.

(4) mur 'sea'
mara·v 'dead person'

These examples do not necessarily demonstrate that triggers of nasal harmony in Applecross Gaelic must be stipulated to exclude nasal stops. It is conceivable that stressed vowels not only trigger nasal harmony when they are nasal, but also block it when they are oral; that is, they can spread nasalization but do not alter their own phonemic oral/nasal quality. A situation of this kind in Guaraní is discussed in §2. I found no examples in Applecross Gaelic with a stressed oral vowel and a following nasal with which to test whether a nasal stop triggers harmony in a following unstressed syllable. The scarcity of such forms is likely because vowel nasalization in Applecross Gaelic in most cases arose historically from a nasal consonant in the vicinity that was either retained, lost, or lenited. In some cases, the nasal consonant is still reflected in the orthography: [tãɣ̃] *damh* 'ox, stag', ['s̃ã·h̃ũx̃] *sàmhach* 'quiet'.

A pattern of nasal harmony that includes a supralaryngeal fricative among its blockers is found in Epena Pedee (Saija), a Chocó language of Colombia (Harms 1985, 1994). Nasal vowels trigger progressive nasal harmony, as shown in (5a). Certain consonants in the onset of a syllable that contains a nasal vowel also become nasalized, as will be discussed presently. Of particular relevance is the set of consonants that block progressive nasal harmony, seen in (5b); this includes /s/, which is the only supralaryngeal fricative phoneme in the language, as well as other obstruent phonemes and the trill /r/.[1] Within the stem, non-continuant obstruents become prenasalized following a nasal vowel (Harms 1994). The phonemic analysis and phonetic description for these examples follow Harms. Some phonemic forms are constructed on the basis of his orthography, which is close to phonemic.

[1] The description of Harms (1985: 16) states that /s/ blocks progressive nasal harmony. The later description by Harms (1994: 8) seems to indicate that /s/ does not always block spreading, but includes the example ['mĩʔ̃ãsu] 'spear', where it is opaque (1994: 6). Another example, ['s̃ĩʔ̃ãs̃õ] 'sugarcane' (1994: 5), could be regarded as showing that /s/ does not block spreading, since Harms's phonemic transcription of this word posits only the first syllable as underlyingly nasal. However, it is also compatible with a treatment in which both the first and last syllables contain nasal vowels underlyingly.

Regressive nasalization occurs within syllables containing a nasal vowel but does not usually transmit beyond the onset.[2] Syllables in Epena Pedee are open and begin with a consonant; consonant clusters are infrequent. A consonant that is tautosyllabic with a nasal vowel becomes nasalized except for [pʰ tʰ kʰ ʧ] ([p t k] and [r] do not occur preceding a nasal vowel). In contrast to their behavior in the progressive harmony, /s/ is characterized as nasalized when preceding a nasal vowel and voiced stops become full nasals. Harms analyzes [m n] as allophones of voiced stops; they are not part of the phonemic inventory. Harms includes [ʔ] in the category of segments that do not become nasalized in the onset to a nasal vowel. While nasalization would not be audible during [ʔ], it is possible that the velum is lowered during this segment. The status of [ʔ] in nasal harmony in general is discussed later in this section.

(5)	a.	/ˈdãw̃e/	[ˈnãw̃ẽ][3]	'mother'
		/peˈɾõɾa/	[peˈɾ̃õɾ̃ã]	'guagua (a groundhog-like animal)'
		/kʰãˈjaɾa/	[kʰãˈɲãɾ̃ã]	'than'
		/hẽˈsã:/	[h̃ẽˈs̃ã:]	'stinging ant'
		/hebẽˈdẽ/	[hemẽˈnẽ]	'to play'
	b.	/kʰĩˈsia/	[kʰĩˈsiə][4]	'think'
		/ˈbĩãsu/	[ˈmĩə̃su]	'spear'
		/ˈãʧi/	[ˈʔãⁿʧi]	'they'
		/ˈhõpʰe/	[ˈh̃õᵐpʰe]	'fish (sp.)'
		/kʰũˈtra:/	[kʰũˈⁿtra:]	'young man'
		/ĩˈbɨsi/	[ʔĩˈᵐbɨsi]	'neck'
		/w̃ãhiˈda/	[w̃ãh̃ĩˈⁿda]	'go (PAST PL)'
		/ˈãĩgi/	[ˈʔãĩᵑgɨ]	'daughter-in-law'
		/ˈtũ:ra/	[ˈtũ:ra][5]	'pelican'

Blocking of nasal harmony by /r/ is likely due to the aerodynamic and perceptual difficulties that a nasal trill would present (Solé 2002, 2007b). Solé points out that "an open velopharyngeal port would bleed the intraoral pressure required to make a relaxed oscillator vibrate for trills" (2002: 677). In addition, a velopharyngeal opening that was small enough to not impair a trill would likely be of insufficient size to produce perceptible nasalization. However, Solé observes that a tap is compatible with nasalization. This is consistent with the pattern that Epena Pedee displays. The distinct behavior of taps and trills suggests that the category of liquids in the scale of targets for nasal harmony should be segregated into a category that includes taps, flaps, and lateral approximants and another lower-ranked category that contains trills. A case where taps and laterals are both targets to the exclusion of obstruents occurs in Ịjọ, a Niger-Congo language of Nigeria (Williamson 1965, 1969, 1987).

2 Harms (1985: 16) describes "a minor degree" of nasalization on a vowel that precedes a nasal syllable, but he characterizes it as "so slight" that he does not represent it in transcription.

3 A fricative variant [β] of /w/ does not block harmony: [ˈnãw̃ẽ] ~ [ˈnãβ̃ẽ].

4 This form is transcribed in Harms (1985) with a prenasalized [s], but later description in Harms (1994) indicates that [s] is not prenasalized following a nasal vowel.

5 Harms (1985: 16) transcribes this form without aspiration of [t]; however, given his description of voiceless stops before a nasal vowel (1985: 15), it is presumably aspirated.

The prenasalization of non-continuant obstruents following a nasal vowel raises questions about their representation. One issue is whether the feature [nasal] is specified for a portion of these segments, and if it is, how the nasal-oral sequence is represented.[6] Prenasalized segments have been represented by some researchers as a single root or slot with specifications for both [+nasal] and [−nasal] (e.g. Bivin 1986; Sagey 1990), but Padgett (1995) and Piggott (1997) have analyzed prenasalization as a combination of nasal and oral segments. Steriade (1993) proposes aperture-based representations in which the closure and release phases of a stop can each form a separate anchor for [nasal], in which case a prenasalized stop has [nasal] associated with the closure but not the release. Tied in with Steriade's representation is a claim that [nasal] is a privative feature (see also Trigo 1993; Steriade 1995), which supports a bipositional representation for prenasalized plosives. Beckman (1999) applies the aperture-based approach to prenasal stops in Guarani, which may occur at the boundary between a nasal span and an oral span (see §2). While much contemporary research concurs with the need for a sequence of segments or phases within a segment, the specifics of the representation remain at issue. On a related topic, Botma (2009) proposes that post-nasalized stops that trigger nasal harmony in Yuhup are underlyingly nasals that have undergone denasalization in a particular context.

Epera (also known as Ẽperã) has a nasal harmony that is roughly similar to Epena Pedee, to which it is related, but it differs in the respect that voiced stops do not block progressive nasal harmony within a morpheme (Morris 1978; Bivin 1986). Thus, nasal harmony transmits through vocoids, laryngeals, flaps and voiced stops, but is blocked by voiceless obstruents, as illustrated in (6). The availability of this pattern could be connected to the lack of contrast between voiced and nasal stops in the language. Like Epena Pedee, nasal stops [m n] have been analyzed as allophones of voiced stops. However, not all languages that lack this contrast target voiced stops, as Epena Pedee shows. Nasalization in the underlying form is shown on the first vowel in a segment sequence that shows nasal harmony.

(6)	/bẽa/	[mẽã]	'bush'
	/hĩ-/	[h̃ĩ-]	'to tie'
	/bẽɾa/	[mẽɾ̃ã]	'greetings'
	/põsoa/	[põ^{n}soa]	'termite'
	/sãki/	[sã^{ŋ}ki]	'which'
	/ĩbaba/	[ĩmãmã]	'tiger'[7]

Nasal harmony that affects voiced stops (in some cases ones that are prenasalized in oral contexts) and that is blocked by voiceless obstruents is also reported for Orejon (Pulleyblank 1989) and Parintintin (Pease and Betts 1971; Bivin 1986). These languages, too, seem to lack a contrast between oral and nasal voiced stops, although more detailed descriptions are needed. These patterns suggest that when voiced

[6] I will use [nasal] to cover a privative nasal feature or a [+nasal] specification. Where an equipollent *vs.* privative distinction is relevant, I will disambiguate with ± notation or in surrounding discussion.

[7] For this example, the consonants' phonetic realization is based on the description provided by Morris and Bivin.

stops and nasals are not in contrast, voiced stops are more prone to undergo nasal harmony. Whether the voiced stops in such patterns should be analyzed as obstruents is open to debate. For particular cases where oral stops that alternate with nasals in nasal harmony have been analyzed as sonorants, see Piggott (1992), Rice (1993), Piggott and van der Hulst (1997), Botma (2004, 2009), and Botma and Smith (2007), although most often voiceless obstruents are transparent in such patterns (see also CHAPTER 8: SONORANTS).

Arabela, a Zaparoan language of Peru, shows a nasal harmony that targets only vowels and glides, as illustrated in (7) (Rich 1963). Nasalization spreads progressively from nasal stops and /h̃/. The glottal fricative is nasal in all contexts in this language and lacks an oral counterpart. [ʔ] occurs to close a phonological phrase and is non-phonemic.

(7)	'nẽẽnũʔ	'to turn over'
	'h̃ãnũʔ	'to fly'
	'mj̃æ̃nũʔ	'swallow'
	'nũw̃ãʔ	'partridge'
	'h̃ĩj̃æ̃nĩʔ	'old woman'
	'nj̃æ̃æ̃'riʔ	'he laid it down'[8]
	'kıronĩʔ	'deep'
	'komãh̃ĩʔ	'over there'
	'mɯ̃w̃ẽẽgur'h̃ũnũʔ	'wiggling'
	'h̃ẽẽgıʔ	'termites'
	'nẽj̃æ̃'tuʔ	'daughter'
	'nĩ'tjænũʔ	'to carry on the back'

Sundanese, an Austronesian language spoken in Western Java, has a progressive nasal harmony that targets only vowels and glottals (Robins 1957; Cohn 1990, 1993a). Nasal stops are the triggers.

(8)	ɲãĩãn	'wet (ACTIVE)'
	ɲãũr	'say (ACTIVE)'
	mĩh̃ãk	'take sides (ACTIVE)'
	kumãh̃ã	'how?'
	bɨŋh̃ãr	'to be rich'
	nũʔ̃ũs	'dry (ACTIVE)'
	ŋãjak	'sift (ACTIVE)'
	ŋãwih	'sing (ACTIVE)'
	ŋũliat	'stretch (ACTIVE)'
	mãrios	'examine (ACTIVE)'
	ŋõbah	'change (ACTIVE)'
	ŋĩsar	'displace (ACTIVE)'
	ŋãtur	'arrange (ACTIVE)'

The status of the laryngeals, [h ʔ], in nasal harmony systems deserves comment. Laryngeals rarely – perhaps never – block nasal harmony (Walker and Pullum

[8] Rich does not distinguish degrees of stress in her transcription.

1999; Walker 2000a). Blocking by a glottal stop has been reported for Rejang (Austronesian) (McGinn 1979: 187), but field research by Robert Blust suggests otherwise (Walker and Pullum 1999: 776, n. 17). In Kaiwá (Tupí-Guaraní), nasal harmony transmits through [ʔ] at a normal speech rate, but [ʔ] is reported to block nasal harmony in slow speech (Harrison and Taylor 1971: 17). It would be valuable to verify these descriptions with modern investigative techniques. Across languages, the overwhelming tendency is for nasal harmony to transmit through laryngeals. This has prompted researchers to situate laryngeals above the category of glides in the hierarchy that characterizes cross-language variation in targets of nasal vowel–consonant harmony, as in (1) (Schourup 1972; Piggott 1992; Walker and Pullum 1999). Levi (2005) has proposed a refinement in which laryngeals are situated higher than phonemic glides in particular, that is, higher than glides that are not derived from vowels rather than glides that are the non-syllabic realization of vowels.

Laryngeals have sparked discussion about the representation of nasal segments and the definition of the feature [nasal] (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION; CHAPTER 17: DISTINCTIVE FEATURES; CHAPTER 27: THE ORGANIZATION OF FEATURES). If [nasal] reports to the supralaryngeal node in the feature geometry, then laryngeal segments could not be phonologically nasal, whereas if [nasal] were a dependent of the root node, any segment could potentially be specified for [nasal] in the phonology (Cohn 1990, 1993a). Cohn assumes the former representation, but leaves the issue open for further research. Under the assumption that [nasal] is dependent on the supralaryngeal node, laryngeals are phonetically nasalized in the context of nasal segments but they do not participate in nasal processes in the phonology or have the capacity to show nasal contrasts. Piggott (1992) takes a different perspective, in which laryngeals can be phonologically [nasal], and may therefore undergo nasal harmony. In the geometry that he assumes, [nasal] can be dependent on a soft palate node, which reports to the root. He supposes, however, that a nasalized glottal stop is not phonetically possible, because of its lack of egressive nasal airflow. Accordingly, Piggott postulates a feature co-occurrence restriction over [nasal] and [constricted glottis] that applies at the later level of phonetic implementation.

Walker and Pullum (1999) also contend that laryngeals can be phonologically specified for [nasal]. Support they cite for this claim includes patterns in which [h̃] triggers nasal harmony (e.g. Arabela) or contextual vowel nasalization. For two such languages (Kwangali, Seimat), they note evidence for a phonemic contrast between [h] and [h̃]. In addition, Walker and Pullum observe that the scarcity of blocking of nasal harmony by laryngeals points to their being highly compatible with acquired nasalization, a tendency that can be straightforwardly captured if laryngeals can be phonologically nasal. To allow the possibility of nasalized laryngeals, they conclude that [nasal] should be defined as corresponding to an open velopharyngeal port rather than requiring nasal airflow (see also Cohn 1993a; Padgett 1995; Hume and Odden 1996). Therefore, a glottal stop can be nasalized by virtue of the lowered velum posture even though there is no airflow through the nasal cavity. In accordance with this perspective, laryngeals are transcribed as nasalized in this chapter when they occur within a nasal harmony span. It is plausible that [ʔ] at the periphery of a nasal harmony span, e.g. in Epena Pedee and Arabela, should likewise be treated as specified for [nasal]. Because of the lack of airflow, nasalization will not be perceptible during [ʔ̃]. This makes it unlikely that glottal stops will show a phonemic contrast in nasality (but see Walker and Pullum 1999 for a hypothesized scenario).

The language data presented above illustrate the scale of targets in patterns of nasal harmony that show blocking. All of the particular cases considered show progressive harmony, with regressive harmony in the syllable in some instances. However, some patterns show more robust regressive harmony, as discussed in §4. The scalar effects for targets have been analyzed by Walker (2000a) as the result of a hierarchy of feature co-occurrence constraints abbreviated as in (9) (excluding laryngeals, discussed below). The constraint *Nas-ObstruentStop prohibits a lowered velum during a segment that also has the features that characterize an obstruent stop, a segment that is highly difficult if not impossible. Nasalization during an articulation with stoppage in the oral cavity usually results in a sonorant stop (e.g. [m n ɲ ŋ], etc.).

(9) *Nasalized segment constraint hierarchy*

*Nas-ObstruentStop >> *Nas-Fricative >> *Nas-Liquid >> *Nas-Glide >> *Nas-Vowel

In addition to nasal harmony surveys, this markedness scaling gains support from other facts about nasal patterns, such as segment inventories and nasal place assimilation (e.g. Pulleyblank 1989; Cohn 1993a; Padgett 1995). The constraint hierarchy can be used to obtain cross-language differences in the sets of targets and opaque segments by ranking a harmony-driving constraint at different breaks in the hierarchy. Feature co-occurrence constraints that dominate the harmony driver will correspond to blocking segments and ones that are dominated will correspond to targets.

Walker suggests that the hierarchy of constraints is grounded in factors of articulatory compatibility, aerodynamic difficulty, and ease of perceptibility. However, some researchers have observed that conflating these factors is problematic with regard to inventory contrasts and laryngeals. Ní Chiosáin and Padgett (1997) point out that with respect to articulatory compatibility, a constraint against [ʔ̃] is expected to be low-ranked, likely at the same level or even below *Nas-Vowel. However, because of the lack of perceptibility of nasalization during a glottal stop, [ʔ̃] *vs.* [ʔ] makes a poor phonemic contrast. If the hierarchy in (9) included *[ʔ̃] at or near the bottom, it would correctly predict the lack of blocking by [ʔ] in nasal harmony, but it would not account for the disfavored status of a contrast between [ʔ̃] and [ʔ]. Ní Chiosáin and Padgett propose that *[ʔ̃] is low-ranked in the articulatory markedness hierarchy, but attribute the contrastive distribution to the activity of a separate constraint Contrast[nas], which penalizes a [ʔ̃]/[ʔ] distinction. Flemming (2004) goes a step further, suggesting that blocking effects in nasal harmony are a consequence of constraints governing the maintenance of contrasts, a possibility also noted by Ní Chiosáin and Padgett (chapter 2: contrast). Under Flemming's formalization, the constraint hierarchy in question scales nasalized segments according to their proximity to a nasal stop. Thus nasalized fricatives are conceived as highly indistinct from nasal stops, but nasalized vowels and laryngeals are at the upper end of the scale of distinctness from a nasal stop. Boersma (2003) proposes an account of the nasal glottal stop that distinguishes its articulation and perception but with different implementation.

Other work that emphasizes the role of segmental distinctions includes Homer (1998), who employs contrast-sensitive constraints to obtain blocking in nasal harmony. Piggott (2003) proposes faithfulness constraints specific to categories of blocking segments (stop, fricative, liquid, glide) that simultaneously prohibit segment deletion and feature change (CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). He argues that this prevents the possibility that a less favored segment category could undergo nasal harmony to the exclusion of preferred targets, by replacement of the less favored category with a nasal stop. For example, replacement of a liquid with a nasal stop would bypass the violation of a feature co-occurrence constraint on nasalized liquids, in which circumstance a nasal harmony system could be expected to exist that allows liquids to undergo harmony, but not glides. A faithfulness-based approach to hierarchical nasal harmony effects is also proposed by Boersma (1998, 2003), using constraints that penalize adding nasalization in consonants. Constraints for consonants with greater oral constriction are generally ranked higher, as nasalization is posited to have a greater perceptual effect in these segments.

Another perspective on the basis of an implicational scale of targets makes a connection with its similarity to the sonority hierarchy, mentioned earlier. Hume and Odden propose that the effects of both hierarchies reduce to an impedance hierarchy, where impedance is defined as "the resistance offered by a sound to the flow of air through the vocal tract above the glottis" (1996: 358), a concept reminiscent of Boersma's reference to degree of oral constriction. Among supralaryngeal segments, obstruent stops have the greatest impedance, and vowels and glides the least. Segments with low impedance values are favored as syllable peaks, a characteristic traditionally diagnostic of high-sonority segments, and they show greater susceptibility to nasalization. Laryngeals have an impedance value of 0. This renders them highly susceptible to nasalization, but their inability to constitute a syllable peak follows from an assumption that a syllable peak has some impedance value, i.e. a non-zero value. This addresses the earlier-mentioned discrepancy concerning laryngeals in scales governing nasal harmony and what are traditionally sonority-based phenomena (i.e. syllabification), yet it posits a common underlying basis from which scalar effects across these phenomena are derived. This approach is extended by Clements and Osu (2003), who interpret resistance to nasalization in terms of a scale of obstruence, a near-synonym of impedance. Their study revolves around Ikwere, an Igboid language of Nigeria, in which nasal harmony transmits through vowels, approximants, and non-explosive stops, but is blocked by fricatives and obstruent stops. This leads them to add a category consisting of implosives and other non-explosive stops between liquids and fricatives on the nasalizability scale.

Despite the differences in formal perspectives on patterns of nasal harmony with opaque segments, there is broad consensus that groups of targets *vs.* blockers essentially conform to the descriptive hierarchy in (1). In addition to new case studies, like that of Ikwere, future research bearing on these approaches may rest largely on the scope of coverage and emphasis, for example the treatment of sonority or contrast in the theory, which situate the account of nasal harmony in a wider context. Where explanatory overlap exists, general issues of theoretical implementation will also be relevant. For instance, future work on the division of labor between contrast, segmental markedness, and faithfulness in the theory could inform the types of constraints that are expected to be possible.

2 Nasal vowel–consonant harmony with transparent segments

Patterns of nasal vowel–consonant harmony with transparent segments are also attested. Whether the hierarchy in (1) is relevant for these patterns is a matter of debate, intersecting with fundamental questions about the kinds of representations that are involved and whether these systems are of the same basic "type" as ones with opaque segments.

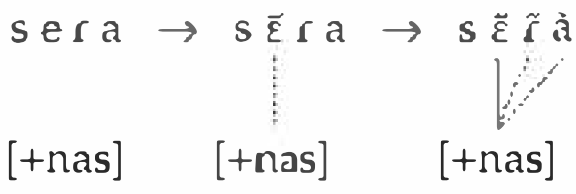
A well-known pattern of nasal vowel–consonant harmony with transparent segments is widely attested in the Tucanoan family. Typically, all voiced segments in a morpheme are either nasal or oral. Voiceless obstruents are consistently oral. They may occur in nasal morphemes, and do not prevent nasal harmony from operating among flanking voiced segments. Examples of nasal harmony in morphemes and words of Tucano, spoken in Colombia, are given in (10a), and oral items are provided in (10b) (West and Welch 1967; Noske 1995).[9] Although not marked as nasal in the sources, I show laryngeals as nasalized in nasal morphemes (see discussion in §1). Noske notes that [h̃] occurs in nasal contexts in other Eastern Tucanoan languages and she tentatively postulates that /h/ is likewise realized as nasal in Tucano.[10]

(10)	a.	w̃ẽõ	'panpipe flutes'
		w̃ĩʔ̃mã	'child'
		ɨ̃mɨ̃	'man'
		w̃ãtĩ	'devil'
		mãh̃ã	'macaw'
		mãʔ̃ã	'trail'
		nĩtĩ	'charcoal'
		j̃õh̃kã	'a drink made from bitter manioc'
		sẽɾ̃ã	'pineapple'
		mãsã	'people'
		ʔ̃ɛ̃ʔ̃kẽã	'nose'
		sũkũã	'small of back'
	b.	jai	'jaguar'
		jɛsɛ	'pig'
		oho	'banana'
		kahpea	'eye'
		oso	'bat'
		ake	'monkey'
		patu	'coca'
		paga	'stomach'
		sɛɾɛ	'a skin disease'
		ᵐbɛʔro	'later'
		ⁿdiɛri	'eggs'
		etagi	'the one who is arriving'

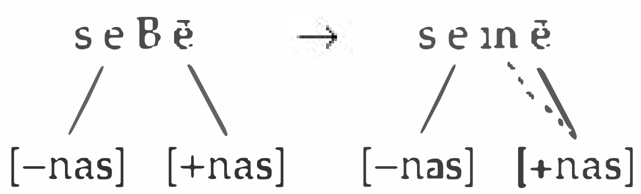
[9] On the operation of nasal harmony from roots to certain suffixes in Tucano, see Trigo (1988) and Noske (1995).

[10] An off-glide [h], realized predictably in word-final position, is not shown in these transcriptions.

In Tucano, a complementary distribution exists between nasal stops and voiced stops (realized as oral or prenasalized, depending on context), with the former occurring in nasal morphemes and the latter in oral morphemes. In nasal morphemes, Noske (1995) postulates that [+nasal] is a feature of the entire morpheme, and it is floating, i.e. unassociated, in the underlying representation. She assumes that [+nasal] links to the first vowel in the word and spreads within the morpheme, as illustrated in (11).[11]

(11) 

Tucano has a small number of disharmonic roots that contain both nasal and oral vowels, e.g. [kõ'^{m}pɛ] 'left', [semẽ] 'paca'. Noske treats these with [±nasal] features that are specified for individual segments underlyingly. She argues that [−nasal] specifications are needed in addition to [+nasal], to prevent [+nasal] from spreading to all voiced segments. This is shown in (12). /B/ represents the phoneme that is variously realized as a nasal or voiced labial stop.[12]

(12) 

In systems where voiceless obstruents do not impede harmony, descriptions largely converge on the realization of these consonants as voiceless obstruents in nasal harmony contexts. Consider the case of nasal vowel–consonant harmony in Guaraní, where voiceless consonants are reported to be transparent.[13] An acoustic study of voiceless stops in nasal harmony contexts in unsuffixed Guaraní words found no evidence of nasal airflow energy during the stop closure, nor was the closure fully voiced (Walker 1999). On the ongoing debate about instrumental evidence for nasalized fricatives, see the aforementioned references on that topic. Examples of nasal harmony in Guaraní are given in (13). The data are from Gregores and Suárez (1967), Rivas (1975), Piggott and Humbert (1997), and Kaiser (2008). Nasal harmony that targets voiced segments and laryngeals is triggered by a stressed nasal vowel (a, b), and stressed syllables that contain an oral vowel block harmony (b). Harmony is robust in the regressive direction and is also triggered by a prenasalized stop (c).[14] Progressive harmony might be more restricted.[15] There is

11 [ɛ] is an allophone of [e] in Tucano.

12 Noske assumes an intermediate stage in this derivation that is not relevant to the issues under focus.

13 However, the locative suffix shows an alternation between [-pe] and [-mẽ]. See Piggott and Humbert (1997) for discussion.

14 Kaiser (2008) finds that word-initial vowels often do not undergo nasal harmony in words over two syllables long. She speculates that a morpheme boundary might be blocking harmony in these forms.

15 Kaiser (2008) observes that the data available do not confirm whether progressive harmony can affect more than one syllable. See also Piggott and Humbert (1997) on observed asymmetries in progressive *vs.* regressive harmony in Guaraní. If it were determined that progressive harmony does not advance beyond one syllable, then the second two examples in (13b) would be most relevant to establish blocking by a stressed syllable, for regressive harmony from the following nasal vowel or prenasalized stop.

a long history of discussion of the Guaraní pattern in the theoretical literature. For overviews, see Piggott and Humbert (1997), Beckman (1999) and Walker (2000a).

(13)	a.	/ⁿdo-ɾoi-ⁿdu'pã-i/	[nõɾõĩnũ'pãĩ]	'I don't beat you'
		/ɾo-ᵐbo-po'ɾã/	[ɾ̃õmõpõ'ɾ̃ã]	'I embellished you'
	b.	/iᵈja,kãɾa'ku/	[ĩɲã,kãɾ̃ã'ku]	'is hot-headed'
		/u'peiʃa,ɾõ/	[u'peiʃã,ɾ̃õ][16]	'then, because of that'
			[ɲãsã,ĩ'ⁿduɪ]	'moonlight'
	c.	/ɾo-ᵐbo-he'ⁿdu/	[ɾ̃õmõh̃ẽ'ⁿdu]	'I made you hear'
		/ɾo-ᵐbo-uɰʷa'ta/	[ɾ̃õᵐbouɰʷa'ta]	'I made you walk'

Mọ̀bà Yoruba has a regressive nasal harmony in which both voiced and voiceless stops are transparent, as well as fricatives, as shown in (14). The harmony targets vowels, glides, and liquids (Ajíbóyè 2001; Piggott 2003; Archangeli and Pulleyblank 2007). However, as mentioned in §1, mid vowels block harmony. Mid vowels are always oral, with the exception of [ɔ̃] when it is an allophone of /ã/. In addition, despite the occurrence of phonemic /ã/ in the language, /a/ is opaque to harmony: [ìsasũ̀] 'kind of pot', [agũ̀ɪtã̀] 'sheep'.

(14)	/uwã́/	[ũw̃ɔ̃̀]	'lie'
	/jṹ/	[ɲṹ]	'go'
	/ùrĩ̀/	[ũ̀r̃ĩ̀]	'walk'
	/rũ̀-orĩ/	[ror̃ĩ]	'to chew chewing stick'
	/lã́/	[nã́]	'to spend'
	/idã́/	[ĩdã́]	'magic'
	/udũ̀/	[ũdũ̀ɪ]	'lover of sweet things'
	/egigũɪ/	[egĩgũ]	'bone'
	/ùgbĩ́/	[ũ̀gbĩ́]	'snail'
	/ì-sĩ̀/	[ĩ̀sĩ]	'worship'
	/itã́/	[ĩ̀tã́]	'story'
	/ikĩ/	[ĩkĩ]	'mucus'

Piggott suggests that voiced stops are underlyingly obstruents in Mọ̀bà Yoruba, unlike languages where voiced stops alternate with nasals in nasal harmony, for which he analyzes the voiced stops as sonorants (see also Botma 2009).

The patterns of nasal harmony in Guaraní and Mọ̀bà Yoruba are also revealing with respect to possible domains of nasal vowel-consonant harmony. In Guaraní, harmony can be bounded at an edge by a stressed syllable, as seen in (13b). This has led some researchers to analyze some or all of Guaraní nasal harmony as operating within metrically defined constituents or via them (e.g. Sportiche 1977; Halle and Vergnaud 1978; van der Hulst and Smith 1982; Flemming 1993; Piggott and Humbert 1997; see also CHAPTER 40: THE FOOT; CHAPTER 41: THE REPRESENTATION OF WORD STRESS). Beckman (1999) offers a different perspective, in which the role of stressed syllables in Guaraní is attributed to the preservation of the underlying oral/nasal quality of segments in these positions. That approach also accounts for the limitation of phonemic nasality in vowels to stressed syllables.

Another level of prosodic structure has been suggested to be relevant for harmony in Mọ̀bà Yoruba. In this language, nasal harmony can span a word boundary.

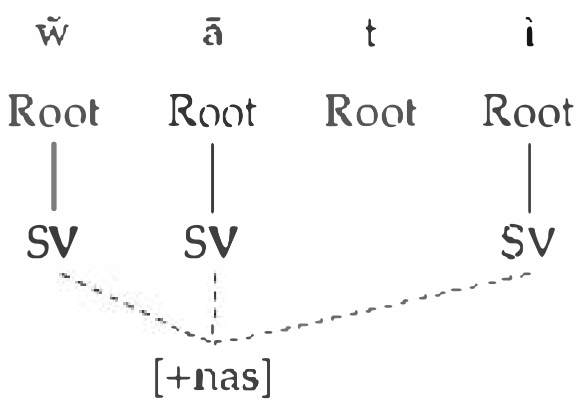
[16] The vowel sequence [ei] in this form is tautosyllabic (Kaiser 2008).

Examples are given in (15). Ajíbóyè (2001) describes /lĩ/ as a particle and analyzes the domain of harmony as the prosodic word, a constituent that can contain more than a morphological word.

(15)	/ki lĩ/	[kĩ́ nĩ]	'what is it?'	/kí à/	[kí à]	'what is it?'
	/ìsí lĩ/	[ĩsĩ́ nĩ]	'who is s/he?'	/isí à/	[ìsí à]	'who is s/he?'

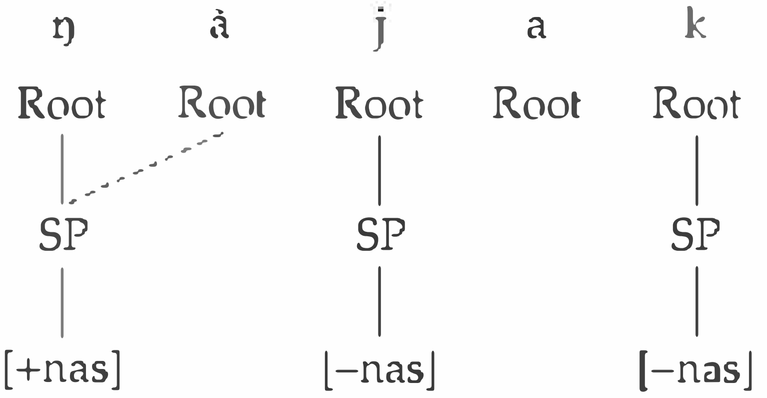
Returning to issues surrounding transparent segments and targets, for the Tucano-type patterns in particular, debate has surrounded their analysis and the conception of where they fit in the typology of nasal harmony. One primary approach to these systems posits that they involve different segmental representations from systems like those described in §1, with opaque segments (Piggott 1992). Specifically, they differ in the dependency of the feature [±nasal] in the feature geometry, and in the node that spreads in nasal harmony. In systems with transparent voiceless obstruents, [±nasal] is dependent on a spontaneous voicing node (SV), which is present in sonorant segments. Harmony results from the spreading of [+nasal] among adjacent SV nodes, as depicted in (16). Voiceless stops are transparent to harmony because they lack an SV node. Voiced stops are treated as sonorants in these systems (see discussion in §1).[17] Piggott suggests that sonorancy is the source of prenasalization of these consonants in certain oral contexts. The realization is attributed to an articulatory configuration needed to produce spontaneous voicing. Prenasalization in this circumstance thus does not involve a specification for [+nasal] but rather is an epiphenomenon of the sonorant stops' phonetic implementation.

(16)



In nasal harmony with opaque segments, [±nasal] is dependent on a soft palate (SP) node. An SP node is underlyingly specified in some consonants. Nasal harmony ensues from spreading of the SP node to segments that lack it, as shown in (17) for a Sundanese form. Under this approach, differences in the set of opaque segments arise from differences in the segments that are underlyingly specified for an SP node (governed by Piggott's Contrastive Nasality Principle).

(17)



[17] For other assumptions about the representation of sonorant stops in the context of nasal harmony systems, see Botma (2004, 2009) and Botma and Smith (2007).

In other work, nasal vowel–consonant harmony systems with transparency *vs.* blocking have been divided along the lines of relations between syllables and segments (Piggott 1996, 2003; Piggott and van der Hulst 1997). In nasal harmony with transparent consonants, [nasal] is considered to be licensed as a property of the syllable, whereas in harmony with blocking segments, the host for [nasal] is the segment. In the case of syllable licensing, [nasal] is associated with the syllable head – the nucleus (see CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE) – and becomes associated with all other sonorant segments in the syllable (Piggott 2003; see Botma 2004, 2009 for a related claim). Locality is respected in harmony with transparent consonants, because no syllable heads are skipped in the propagation of nasal harmony. Noske (1995) also assumes a licensing relation between the syllable and [±nasal] for the Tucano pattern, but with some different specifics in her assumptions.

In another approach that posits a basic difference between nasal vowel–consonant systems, those with opaque segments are caused by articulatory spreading, whereas those with transparent consonants involve spreading that is perceptually based (Boersma 1998, 2003). Patterns with blocking are claimed to be driven by an articulatory constraint that penalizes shifts in the position of the velum. This constraint can favor persistence or early onset of a lowered velum in a word that contains a nasal segment. For patterns with transparent obstruents, such as the Tucano type of system, Boersma proposes that a perceptually based constraint drives harmony, causing nasalized segments that are interrupted by an oral segment to be perceived with a single value of [+nasal]. This perceptual representation is distinct from the articulation in which the nasalized segments are interrupted by a velum-raising gesture. Boersma suggests that the reason all sonorants become nasalized in patterns with transparency is connected to the lexical-level specification of nasality in these languages (see (11)). Boersma reasons that if [±nasal] is a suprasegmental feature, it is less likely to be specified for individual segments. Segments are thus less likely to have a [–nasal] specification to which to be faithful, and segments that do not become nasalized will be the ones that are inherently problematic in combination with nasalization, i.e. fricatives and plosives.

The notion that nasal vowel–consonant harmony patterns with transparency are tied to perception is also pursued in work by Sanders (2003). He proposes that nasal harmony in Tucano-type languages is driven by dispersion constraints on the perceptual distance of systemic contrasts. These constraints favor words that differ to the greatest extent possible in the perception of a nasal/oral contrast, while obeying higher-ranked constraints that prohibit nasalized voiceless obstruents, i.e. they favor the morphemes in which all segments besides voiceless obstruents are the same in nasality.

In contrast to analyses where patterns with transparent segments are analyzed as involving representations or harmony imperatives that are different from those with opaque segments, another approach analyzes these systems as having a common source (Walker 2000a, 2003). This account emphasizes a complementarity in the patterns: there is no nasal vowel-consonant harmony in which all of the segments become nasalized, yet there are systems in which obstruents are transparent and the remaining segments are targets. Obstruents form the focus of the complementarity. All segments except (some) obstruents have the potential to be targets in nasal vowel–consonant harmony and only obstruents are transparent. Walker proposes a treatment of the patterns that analyzes systems with transparent obstruents as cases that correspond to the right endpoint of the hierarchy in (1), where nasalization transmits through all segment categories. Walker adduces typological evidence in support of conceptualizing transparent obstruents as on a par with targets. She observes that when obstruents are transparent, all other segment categories are targets, a generalization that would be expected if obstruents were targets in these systems, because they are lowest-ranked on the target scale. More generally, a survey of over 75 languages with nasal vowel–consonant harmony reveals that if a segment is "permeated" by nasal harmony, that is, if it is targeted or behaves as transparent, then all segments belonging to categories that are higher-ranked in the target hierarchy of (1) are also permeated.

A pattern involving voiced stops is brought to bear on the claim that obstruent stops can be targets in nasal vowel–consonant harmony. The nasal harmony of Tuyuca, another Tucanoan language, has been characterized as showing a difference in blocking and transparency effects when it occurs across morphemes *vs.* within them. Like Tucano, in harmony within a morpheme, voiced stops alternate with nasals and voiceless obstruents are transparent to harmony. However, harmony from stem to suffix is blocked by fricatives and voiced and voiceless stops (Barnes 1996; Walker 2000a).[18] Opaque voiced stops are realized as oral or nasal, depending on the nasality of the suffix to which they belong. (See Trigo 1988 and Walker 2000a on the separate phonological treatment of voiced/nasal velar stops.) Walker interprets the blocking of harmony by voiced stops across a morpheme boundary as evidence of their underlying obstruent status in Tuyuca (cf. Botma 2004); all suffixes that alternate in nasal harmony issuing from the stem therefore begin with a continuant sonorant or laryngeal.[19] When voiced stops undergo harmony within a morpheme, they would then be an instance of voiced obstruent stops that are targets in nasal harmony.

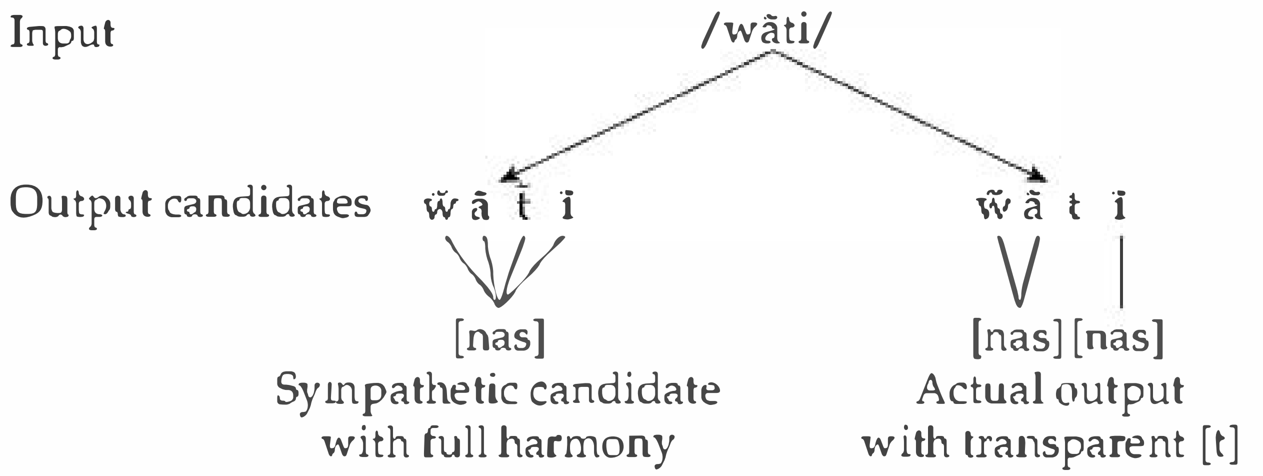
In this approach, feature spreading is analyzed as strictly local at the level of the segment (CHAPTER 81: LOCAL ASSIMILATION). This implies that segments cannot be skipped in harmony; they must either participate in harmony or block it. As a consequence, it is assumed that a phonological representation is available in which a "transparent" obstruent is nasalized (see CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS). The model that Walker proposes is illustrated in (18), implemented using the concept of a sympathetic candidate (McCarthy 1999). A sympathetic candidate is a designated form to which the actual output is encouraged to be similar via the activity of a candidate-to-candidate correspondence relation between this form and the actual output. Arrows represent the existence of correspondence relations among representations, which mediate the enforcement of identity between related forms. Among the candidate outputs generated for an input with a [nasal] specification is one where the feature spreads to all segments in the morpheme, satisfying the constraint that drives harmony. This is the candidate that becomes designated as "sympathetic." However, as [nasal] is not compatible with an obstruent stop, this candidate is not selected. Instead, a form is selected that is identical to the full harmony candidate except that the stop is

[18] Tucano has also been analyzed in this way by Trigo (1988), but it has come to light that a suffix beginning with a labial voiced stop alternates in nasal harmony in this language (Piggott and van der Hulst 1997; Botma 2004), which indicates that voiced stops do not systematically block harmony from the stem to suffix in this language.

[19] Whether laryngeals should be treated as sonorants is an open question.

oral. Because feature associations may not skip a segment, this candidate must have separate [nasal] specifications flanking the oral stop. The actual output is thus chosen not directly by the harmony-driving constraint but rather because of its similarity to a candidate that fares best with respect to that constraint.

(18)



This approach makes use of an abstract phonological representation in which obstruent stops are nasalized. Such representations do not occur within the actual output but they influence its selection. Another analysis that is similar in spirit posits the full harmony form as an intermediate level of representation that is generated as the product of a nasal spreading rule (Piggott 1988). A clean-up rule that prevents nasalized obstruents then causes the stop to be denasalized, to generate the derivational series: /w̃ati/ → w̃ãt̃ĩ → [w̃ãtĩ].

In summary, systems of nasal vowel-consonant harmony with transparent consonants have formed the center of discussion on several theoretical themes. One basic question is whether they should be considered the same "type" of nasal harmony as systems with opaque segments. Also at issue are the segmental representations involved, including the organization of [nasal] in the feature geometry, the level of structure at which [nasal] spreads, the locality of feature associations, and what kinds of abstract representations are involved. Questions about levels of representation have been touched upon, including whether there are distinct articulatory and perceptual representations and the possibility of intermediate or sympathetic forms. Finally, blocking and transparency effects in nasal vowel–consonant harmony have given rise to different perspectives on the harmony imperative and the sources that cause segments to block harmony or behave as transparent.

3 Nasal consonant harmony

In nasal consonant harmony systems, nasal harmony involves consonants only. Moreover, the participant consonants have been characterized as ones that are phonologically similar. These systems have been considered to differ from nasal vowel–consonant harmony in locality and the nature of participant segments, giving rise to proposals in which nasal consonant harmony involves a different harmony-driving imperative and/or different representations.

A prototypical case of nasal consonant harmony is found in Kikongo, a Bantu language spoken in the Democratic Republic of the Congo (Bentley 1887; Dereau 1955; Ao 1991; Odden 1994; Piggott 1996; Rose and Walker 2004; see also CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS). The nasal stop phonemes of

Kikongo are [m n]. Nasal consonant harmony causes voiced stops and /l/ to become nasal when following a prevocalic nasal stop at any distance in the stem. The stem constituent in Kikongo consists of the root and suffixes. Examples of alternations in the perfective active and applicative suffixes induced by nasal consonant harmony are shown in (19). The consonant in these suffixes is analyzed as /l/ underlyingly. In words where the conditions for nasal harmony are not met, /l/ is realized as [d] before [i]. Vowel quality alternations are due to vowel height harmony. Vowels and voiceless consonants are transparent to the nasal harmony; they remain oral when occurring between harmonizing consonants. The forms in (19) consist of stems, as indicated by the initial hyphen according to convention.

(19) a. *Perfective active forms*

-suk-idi	'wash'	-nik-ini	'grind'
-bud-idi	'hit'	-sim-ini	'prohibit'
-bak-idi	'catch'	-futumuk-ini	'revive, rise'
-sos-ele	'search for'	-le:m-ene	'shine'

b. *Applicative forms*

-sakid-il-a	'congratulate for'	-nat-in-a	'carry for'
-to:t-il-a	'harvest for'	-dumuk-is-in-a	'cause to jump for'

In addition to inducing alternations in suffixes, nasal consonant harmony is considered to operate within roots, which do not show a voiced stop or [l] after a prevocalic nasal stop.

Nasal stops that occur in an NC cluster do not trigger nasal consonant harmony (20a), nor do they prevent it from operating across them (20b). In addition, a voiced oral stop in an NC sequence does not undergo nasal harmony from a preceding prevocalic nasal.

(20) *Perfective active*

a.	-bantik-idi	'begin'
	-kemb-ele	'sweep'
	-biŋg-idi	'hunt'
	-tond-ele	'love'
b.	-mant-ini	'climb'
	-meŋg-ini	'hate'

Setting aside NC clusters, the targets of nasal consonant harmony are frequently voiced stops and approximant consonants (/l/ is the only approximant consonant in Kikongo). In some cases, nasal harmony is restricted to consonants separated by no more than a vowel. The Bantu language Ndonga shows this pattern (Viljoen 1973; Rose and Walker 2004). In Ngbaka, a Niger-Congo language spoken in the Democratic Republic of the Congo, the lack of co-occurrence of certain nasals and prenasalized stops within a morpheme has been analyzed as the product of nasal consonant harmony (Hansson 2001; Rose and Walker 2004). Ngbaka contrasts nasal, prenasalized, voiced, and voiceless stops (Thomas 1963, 1970; Wescott 1965). Nasals may occur together with voiced and voiceless stops in a morpheme but not with a prenasalized stop that has the same place of articulation as the nasal (Mester 1988; Sagey 1990; van de Weijer 1994; CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION).

Certain root co-occurrence restrictions (see CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS) on consonants in Ganda (Bantu), discussed by Katamba and Hyman (1991), have also been analyzed as the outcome of nasal consonant harmony (Hansson 2001; Rose and Walker 2004). The patterns in question involve nasals, voiced stops, and voiceless stops. Some of the voiced stops display approximant variants: [b/β], [d/l], [ɟ/j]. Within a root, nasals do not usually occur with a voiced stop (or its approximant variant) that has the same place of articulation. This restriction is observed regardless of the order of the nasal and voiced oral consonant.[20] In addition, the combination of a nasal and a voiceless stop with the same place of articulation is systematically absent if the nasal precedes the stop. In attested roots, identical nasals co-occur, as do oral voiced stops/approximants with the same place of articulation, as shown in (21a). Also attested are roots that combine a nasal and voiced consonant with different places of articulation (21b), and roots in which a voiceless stop precedes a nasal (21c).

(21)	a.	-mémèká	'accuse, denounce'
		-nónà	'fetch, go for'
		-bábùlá	'smoke over fire to make supple'
		-gùgá	'curry favor with'
	b.	-bónèká	'become visible'
		-màlà	'finish'
	c.	-táná	'grow septic, fester'

In the harmony-based analysis of the Ganda pattern, nasal consonant harmony operates within a root among oral stops and nasals with the same place of articulation. For voiced stops the harmony is bidirectional, whereas for voiceless stops it is progressive only. Hansson's (2001) treatment also takes into consideration restrictions on the co-occurrence of nasal stops and voiced prenasalized stops within a root in Ganda.

With respect to non-contour stops, surveys of nasal consonant harmony in Hansson (2001) and Rose and Walker (2004) reveal the following implications: (i) patterns that target voiceless stops with the same place of articulation as the nasal trigger also target voiced stops with the same place of articulation, and (ii) patterns that target voiced stops with a different place of articulation from the nasal trigger also target voiced stops with the same place of articulation as the nasal. An interpretation that has been brought to these generalizations is that nasal consonant harmony favors targets that are similar to nasals (Walker 2000b; Hansson 2001; Rose and Walker 2004). These patterns are suggested to have a basis in speech planning (i.e. the organization and sequencing of abstract units) and its physical execution (i.e. the motor controls that carry out the "plan"). The similarity hypothesis finds support from speech error research. It is well established that the likelihood of a speech error between two phonemes increases with their phonological similarity. A series of speech error elicitation tasks conducted by Walker (2007) found that consonants that are more likely to interact in nasal consonant harmony are also more likely to participate in speech errors with speakers of English.

[20] Ganda also shows a dispreference for particular pairs of voiced stops and nasals in a root when the voiced stop and nasal have a different place of articulation and the voiced stop follows the nasal (Katamba and Hyman 1991; Hansson 2001).

Prior to the extensive typological studies of consonant harmony by Hansson (2001) and Rose and Walker (2004), analyses of nasal consonant harmony were developed that involved nasal feature spreading at a node within the feature geometry (Ao 1991; Odden 1994; Hyman 1995) or at a suprasegmental level (Piggott 1996). The typological studies in question sharpened the characterization of the differences between nasal consonant harmony and nasal vowel–consonant harmony, leading the authors of those studies to analyze nasal consonant harmony as the product of a different harmony-driving mechanism and as involving different representations from those usually assumed for nasal vowel-consonant harmony.

Two of the chief differences between nasal consonant harmony and nasal vowel–consonant harmony involve locality and types of triggers and targets. Nasal consonant harmony targets segments that are phonologically similar to the nasal stop trigger, i.e. stops and approximant consonants. The harmonizing segments are usually non-adjacent, with at least a vowel intervening and sometimes longer transparent sequences. In contrast, in nasal vowel–consonant harmony, harmony affects a (near-)continuous sequence of segments, and vowels are never skipped. In the latter systems, favored targets follow the scale in (1), with vowels ranked at the top, a scaling suggested to have a basis in the segments' phonetic compatibility with nasalization or in maintaining distinct contrasts. Consonants that do not become nasalized in nasal vowel–consonant harmony most often block harmony, although in some systems, (some of) the obstruents behave as transparent.[21]

The role of phonological similarity and the capacity for action-at-a-distance are emphasized in the correspondence-driven approach to nasal consonant harmony (Walker 2000b; Hansson 2001; Rose and Walker 2004). In this account, the occurrence of high phonological similarity between consonants can spur a formal correspondence relation to be established between them. Corresponding segments are co-indexed with one another, as illustrated in (22). Nasal harmony is effected via the correspondence relation. Constraints for individual features, such as [nasal], are postulated that enforce identical specifications in corresponding segments, thus producing nasal consonant harmony, as in (22b). Because nasal assimilation is accomplished through the correspondence relation in this structure, the harmonizing segments are not required to share a single [nasal] specification, unlike the outcome of [nasal] spreading, which is usually assumed for nasal vowel–consonant harmony. A representation like that in (22b) is suggested to accommodate the potential for nasal consonant harmony to occur among non-adjacent segments.

(22) a. n_α i k-i l_α i — n_α linked to [nas]

b. n_α i k i n_α i — each n_α linked to its own [nas]: [nas] [nas]

In the correspondence-based approach, patterns that show harmony only between consonants in adjacent syllables are analyzed using proximity-sensitive constraints governing corresponding segments. The neutrality of preconsonantal nasals

[21] Some additional differences are that nasal consonant harmony never has opaque segments (Hansson 2001; Rose and Walker 2004), whereas many nasal vowel-consonant harmony systems do. Also nasal consonant harmony does not appear to show sensitivity to metrical structure, such as stress and foot boundaries, and it does not extend across word boundaries (Hansson 2001), although these characteristics are attested in some patterns of nasal vowel–consonant harmony.

in a Kikongo-type system has been attributed to their dissimilarity from the potential target oral consonants, either in terms of their role in syllable structure (the nasals in NC clusters are codas, whereas the oral consonants are onsets) or in terms of their release status (the nasals in NC clusters are unreleased, whereas the oral consonants are released). It is suggested that a voiced stop in an NC cluster does not undergo harmony because of avoidance of geminate nasals, which do not occur in Kikongo. In Ngbaka, prenasalized stops are considered to be singleton consonants – not NC clusters, as in Kikongo – so these issues do not arise. Whether the representation and patterning of nasal contours in Ganda fall in line with these treatments has yet to be closely considered.

The correspondence approach to nasal consonant harmony has been applied to consonant harmony systems in general. The basis for that proposal is that other systems of consonant harmony also show effects of similarity and action-at-a-distance. See Hansson (2001), Rose and Walker (2004), and CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS.

4 Directionality

This section turns to directionality in nasal harmony. Some systems can be considered bidirectional, with no apparent difference in the pattern of harmony in either direction. Nasal consonant harmony involving voiced stops in roots of Ganda is an example of this kind. Such patterns do not necessitate an overt statement of directionality. Also, root- or stem-controlled harmony where affixes in the domain of harmony occur only following or only preceding the root or stem can give the appearance of directional harmony but without requiring formal reference to a direction for harmony. However, other systems of nasal harmony show evidence of asymmetrical directionality, where harmony operates in only one direction or shows different patterns in its progressive *vs.* regressive operation.

A contrast in the direction of nasal vowel-consonant harmony with opaque segments is seen in the patterns of the Johore dialect of Malay, an Austronesian language of Malaysia, and Capanahua, a Panoan language spoken in Peru. Johore Malay shows a harmony from nasal stops that targets vowels, laryngeals and glides (Onn 1980). Liquids and obstruents block harmony. Examples in (23) show that the harmony is progressive only.

(23)	pənãŋãh̃ãn	'central focus'
	pəŋãw̃ãsan	'supervision'
	pəmãndaŋãn	'scenery'
	mãkan	'to eat'
	baŋõn	'to rise'
	mə̃nãw̃ãn	'to capture (ACTIVE)'
	mãj̃ãŋ	'stalk (palm)'
	mə̃ratappi[22]	'to cause to cry'
	mĩnõm	'to drink'
	mãʔãp	'pardon'

22 Vowel nasalization in this word is assigned according to Onn's description and harmony rule. Because vowel nasalization is predictable in Johore Malay, Onn only marks it when demonstrating rule applications.

Like Johore Malay, Capanahua displays a nasal harmony that targets vocoids and laryngeals and is blocked by other segments (Loos 1969; Piggott 1992); however, the direction is regressive. The regressive direction cannot be predicted from the position of triggers in the syllable structure, as they occur in both syllable onsets and codas. Word-final nasals are enclosed in parentheses because they are deleted but still trigger harmony.[23]

(24) h̃āmawi 'step on it'
h̃āmã̄ʔōna 'coming stepping'
kajatānaiʔ[24] 'I went and jumped'
wɯrãnai 'I pushed it'
bĩmi 'fruit'
ʧipōŋki 'downriver'
kɯ̃ɯ̃nʧap 'bowl'
bãw̃ĩ(n) 'catfish'
warã(n) 'squash'
pōj̃ã(n) 'arm'
ʔ̃ōnāmpā(n) 'I will learn'

An instance of directionality in nasal vowel-consonant harmony with transparent segments is found in Siriano, a Tucanoan language spoken in Colombia and Brazil (Bivin 1986). Suffixes in Siriano become nasalized following a nasal stem (excluding certain suffixes that are invariant in nasality). Examples of suffix alternations are given in (25). Underlying forms are as provided by Bivin.

(25) a. /wehe-gɨ/ [wɛhɛgi] 'when he is fishing'
to fish-3SG MASC
/w̃èhẽ-gɨ/ [w̃ɛ̃h̃ɛ̃ŋĩ] 'when he is killing'
to kill-3SG MASC
b. /igo-ɾe/ [ʔigoɾe] 'she (COMPLEMENT)'
/ĩgɨ̃-ɾɛ/ [ʔĩŋɨ̃nɛ̃][25] 'he (COMPLEMENT)'

The data in (25) are compatible with nasal harmony where directionality is an epiphenomenon of root or stem control. However, a small group of suffixes harmonize with a following suffix rather than the root, as shown in (26). The suffixes that exhibit this behavior are /-ju/ (SECOND HAND INFORMATION), /-de:/ (PAST NOMINALIZER), /-bu/ (INCEPTIVE), and /-ku/ (PROBABILITY).

(26) a. /wa:-ju-pi/ [waʔjupi] 'they say he left'
to go+EVID:SECOND HAND+3SG MASC
/wa:-ju-ɾã/ [waʔɲũɾ̃ã] 'they say they left'
to go+EVID:SECOND HAND+3PL ANIMATE
/w̃ēhē-ju-pɨ/ [w̃ɛ̃h̃ɛ̃jupi] 'they say he killed'
to kill+EVID:SECOND HAND+3SG MASC

[23] Capanahua also manifests a bidirectional nasal harmony triggered by a nasal stop that is deleted preceding an oral continuant consonant. For discussion, see Loos (1969), Safir (1982), and Trigo (1988).
[24] Vowel nasalization in this word and the next one is assigned according to Loos's description and rules.
[25] Bivin (1986: 71) transcribes the suffix consonant here as [n], but in other transcriptions that he provides for flaps that have undergone nasal harmony in Siriano the consonant is a nasalized flap.

b. /waː-bu-gɨ/ [waˀbugɨ] 'about to go (SG MASC)'
to go+INCEPTIVE+SG MASC
/waː-bu-ɾã/ [waˀɪnũɾ̃ã] 'about to go (PL)'
to go+INCEPTIVE+3PL ANIMATE
/w̃ẽhẽ-bu-gɨ/ [w̃ɛ̃h̃ɛ̃bugɨ] 'about to kill (SG MASC)'
to kill+INCEPTIVE+SG MASC

c. /waː-deː-ɾo/ [waˀaⁿdɛˀɾo] 'where he went'
to go+NOMINALIZER+LOC
/waː-deː-ɾã/ [waˀanɛ̃ˀɾ̃ã] 'the ones who went'
to go+NOMINALIZER+3PL ANIMATE
/w̃ẽhẽ-deː-ɾo/ [w̃ɛ̃h̃ɛ̃ⁿdɛɾo] 'where he killed'
to kill+NOMINALIZER+LOC

d. /waː-ku-a/ [waˀakoa] 'it left'
to go+PROBABILITY+INANIMATE
/waː-ku-bĩ/ [waˀakũmĩ] 'he left'
to go+PROBABILITY+EVID:3SG MASC
/w̃ẽhẽ-ku-a/ [w̃ɛ̃h̃ɛ̃kua] 'it killed'
to kill+PROBABILITY+INANIMATE

Bivin notes that /-ju/ and /-ku/ are evidentials, which must appear with a person-number suffix. Although he does not have data to verify the facts for /-deː/ and /-bu/, he speculates that they too must be used with an additional suffix. We may wonder whether a stipulation for Siriano is needed that harmony with these suffixes is regressive or whether this directionality could be made to follow from morphological structure. Both possibilities have been considered. Bivin suggests that the suffixes in question form a separate lexical class. He treats regressive nasal harmony using a left spreading rule for [nasal] that applies to that lexical class. He also considers an approach that posits an internal word boundary at the left of these particular suffixes. This would block them from harmonizing with the root, as harmony occurs only within words. Bivin disprefers this account because he finds no evidence from other Tucanoan languages to support the presence of an internal word boundary, nor does he find evidence for an analogous occurrence of internal word boundaries elsewhere in Siriano. On the other hand, for Desano, a language closely related to Siriano, Kaye (1971) treats a similar directionality phenomenon as an epiphenomenon of morphological constituency. Like Siriano, Desano has a limited number of suffixes that derive their nasality from the following suffix rather than the preceding morpheme. Three of these suffixes are the same as those in Siriano (Bivin 1986). Kaye suggests that the suffixes targeted by regressive nasal harmony form a morphological constituent with the following suffix that is separate from the preceding stem.[26] He pairs this assumption with a bidirectional nasal assimilation rule that applies cyclically to obtain differences in the direction of harmony. Siriano and Desano, then, are cases where directionality in nasal harmony can perhaps be reduced to the organization of morphological structure, but further study on this issue is needed.

There is possible evidence of directionality in nasal consonant harmony. In the nasal consonant harmony of Kikongo, introduced in §3, harmony is progressive in the stem. The examples in (27) show apparent directionality; a voiced stop or

[26] Miller (1999) treats the regressive nasal harmony in Desano as lexicalized.

consonantal approximant precedes a nasal in the stem but remains oral. (See also the discussion in CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS.)

(27) -dumukina 'jump for'
-bilumuka 'assemble in crowd'

Hansson (2001) speculates that progressive directionality in canonical Bantu nasal consonant harmony, like that of Kikongo, might be reducible to a system of harmony that is stem-controlled and that preserves the underlying oral/nasal quality of root-initial segments (e.g. using a faithfulness constraint specific to this position). However, this hypothesis has been questioned. Although the canonical Bantu root structure is CVC, Rose and Walker (2004) cite evidence that [bilum] is lexically stored as a whole. For the similar pattern of nasal consonant harmony in Yaka, another Bantu language, they point to evidence of stored forms with the sequences /CVlVN-/ and /CVbVN-/, which likewise do not show nasal harmony. Because the /l/ or /b/ is not root-initial in these cases, nor does it belong to a different cycle from the nasal, it would be expected to undergo nasal harmony that was not strictly progressive. Within a correspondence-based approach, Rose and Walker propose to analyze directional harmony in these patterns using a precedence-sensitive identity constraint for the feature nasal in corresponding segments.

In sum, there are systems of nasal vowel–consonant harmony and nasal consonant harmony that seem to display directionality effects that cannot be attributed to independent aspects of the system or structure. Differences in directionality in certain patterns of nasal vowel–consonant harmony with opaque segments present the strongest evidence for these effects. In some cases, certain researchers have suggested that morphological structure and/or prosodic position could obtain the effect of directional harmony, but there is not consensus on this explanation for the various patterns discussed above.

5 Conclusion

To conclude, at the heart of research on nasal harmony are patterns that fall into three descriptive categories: nasal vowel–consonant harmony with opaque segments, nasal vowel–consonant harmony with transparent segments, and nasal consonant harmony. Across languages, patterns belonging to the first category respect an implicational scale that governs favored targets. Whether nasal vowel–consonant harmony with transparent segments and systems with opaque segments share a common source remains in question. Studies bearing on this issue have generated diverse perspectives on the harmony imperatives, the levels of representation that are involved, and the nature of locality. Nasal consonant harmony presents differences from nasal vowel–consonant harmony in showing action-at-a-distance and in favoring harmony between segments that are phonologically similar. This has given rise to a correspondence-driven approach to nasal consonant harmony, situated in a general typology of consonant harmony. This approach is distinct from the treatment of nasal vowel–consonant harmony, which is most often assumed to involve spreading.

The study of nasal harmony can illuminate not only the nature of long-distance phonological assimilation but also themes in phonology that are more general in nature. Whereas in the last couple of decades the broad strokes of the typological characteristics of nasal harmony patterns have been reasonably well delineated, the details of many specific systems remain unknown. Future research could be fruitfully applied to developing more case studies. The resulting findings will doubtless in turn shed new light on the theoretical debates and the cross-linguistic characterization of nasal harmony, both alone and in the larger picture.

ACKNOWLEDGMENTS

For comments on this chapter, I am grateful to Beth Hume, Marc van Oostendorp, and two anonymous reviewers.

REFERENCES

Ajíbóyè, Oládiípò. 2001. Nasalization in Mòbà. *University of British Columbia Working Papers in Linguistics* 8. 1–18.

Ao, Benjamin. 1991. Kikongo nasal harmony and context-sensitive underspecification. *Linguistic Inquiry* 22. 193–196.

Archangeli, Diana & Douglas Pulleyblank. 2007. Harmony. In de Lacy (2007), 353–378.

Barnes, Janet. 1996. Autosegments with three-way contrasts in Tuyuca. *International Journal of American Linguistics* 62. 31–58.

Beckman, Jill N. 1999. *Positional faithfulness: An optimality theoretic treatment of phonological asymmetries*. New York: Garland.

Bendor-Samuel, David (ed.) 1971. *Tupi studies I*. Norman, OK: Summer Institute of Linguistics.

Bentley, W. Holman. 1887. *Dictionary and grammar of the Kongo language*. London: Baptist Missionary Society and Trübner.

Bivin, William E. 1986. The nasal harmonies of twelve South American languages. M.A. thesis, University of Texas at Arlington.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Boersma, Paul. 2003. Nasal harmony in functional phonology. In van de Weijer *et al.* (2003), 3–35.

Botma, Bert. 2004. Phonological aspects of nasality: An element-based dependency approach. Ph.D. dissertation, University of Amsterdam.

Botma, Bert. 2009. Transparency in nasal harmony and the limits of reductionism. In Kuniya Nasukawa & Philip Backley (eds.) *Strength relations in phonology*, 79–111. Berlin & New York: Mouton de Gruyter.

Botma, Bert & Norval Smith. 2007. A dependency-based typology of nasalization and voicing phenomena. In Bettelou Los & Marjo van Koppen (eds.) *Linguistics in the Netherlands 2007*, 36–48. Amsterdam & Philadelphia: John Benjamins.

Clements, G. N. & Sylvester Osu. 2003. Ikwere nasal harmony in a typological perspective. In Patrick Sauzet & Anne Zribi-Hertz (eds.) *Typologie des langues d'Afrique et universaux de la grammaire*, vol. 2, 69–95. Paris: L'Harmattan.

Cohn, Abigail C. 1990. Phonetic and phonological rules of nasalization. Ph.D. dissertation, University of California, Los Angeles.

Cohn, Abigail C. 1993a. The status of nasalized continuants. In Huffman & Krakow (1993), 329–367.

Cohn, Abigail C. 1993b. A survey of the phonology of the feature [±nasal]. *Working Papers of the Cornell Phonetics Laboratory* 8. 141–203.

de Lacy, Paul (ed.) 2007. *The Cambridge handbook of phonology.* Cambridge: Cambridge University Press.

Dereau, Léon. 1955. *Cours de kikongo.* Namur: A. Wesmael-Charlier.

Flemming, Edward. 1993. The role of metrical structure in segmental rules. M.A. thesis, University of California, Los Angeles.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 232–276. Cambridge: Cambridge University Press.

Gerfen, Chip. 1999. *Phonology and phonetics in Coatzospan Mixtec.* Dordrecht: Kluwer.

Gerfen, Chip. 2001. Nasalized fricatives in Coatzospan Mixtec. *International Journal of American Linguistics* 67. 449–466.

Gnanadesikan, Amalia E. 1995. Markedness and faithfulness constraints in child phonology. Unpublished ms., University of Massachuetts, Amherst (ROA-67).

Gregores, Emma & Jorge A. Suárez. 1967. *A description of colloquial Guaraní.* The Hague: Mouton.

Halle, Morris & Jean-Roger Vergnaud. 1978. Metrical structures in phonology. Unpublished ms., MIT.

Hansson, Gunnar Ólafur. 2001. Theoretical and typological issues in consonant harmony. Ph.D. dissertation, University of California, Berkeley.

Harms, Phillip L. 1985. Epena Pedee (Saija): Nasalization. In Ruth M. Brend (ed.) *From phonology to discourse: Studies in six Colombian languages*, 13–18. Dallas: Summer Institute of Linguistics.

Harms, Phillip L. 1994. *Epena Pedee syntax.* Arlington: Summer Institute of Linguistics and University of Texas at Arlington.

Harrison, Carl H. & John M. Taylor. 1971. Nasalization in Kaiwá. In Bendor-Samuel (1971), 15–20.

Homer, Molly. 1998. The role of contrast in nasal harmony. Ph.D. dissertation, University of Illinois, Urbana-Champaign.

Huffman, Marie K. & Rena A. Krakow (eds.) 1993. *Nasals, nasalization, and the velum.* Orlando: Academic Press.

Hulst, Harry van der & Norval Smith. 1982. Prosodic domains and opaque segments in autosegmental theory. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part II, 311–336. Dordrecht: Foris.

Hume, Elizabeth & David Odden. 1996. Reconsidering [consonantal]. *Phonology* 13. 345–376.

Hyman, Larry M. 1995. Nasal consonant harmony at a distance: The case of Yaka. *Studies in African Linguistics* 24. 5–30.

Kaiser, Eden. 2008. Nasal spreading in Paraguayan Guaraní: Introducing long-distance continuous spreading. *Amerindia: Revue d'ethnolinguistique amérindienne* 32. 283–300.

Katamba, Francis & Larry M. Hyman. 1991. Nasality and morpheme structure constraints in Luganda. In Francis Katamba (ed.) *Lacustrine Bantu phonology*, 175–211. Cologne: Institut für Afrikanistik, University of Cologne.

Kaye, Jonathan. 1971. Nasal harmony in Desano. *Linguistic Inquiry* 2. 37–56.

Kingston, John. 2007. The phonetics–phonology interface. In de Lacy (2007), 401–434.

Levi, Susannah V. 2005. Reconsidering the variable status of glottals in nasal harmony. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 41. 299–312.

Loos, Eugene Emil. 1969. *The phonology of Capanahua and its grammatical basis.* Norman, OK: Summer Institute of Linguistics.

McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16. 331–399.

McGinn, Richard. 1979. Outline of Rejang syntax. Ph.D. dissertation, University of Hawaii.

Mester, Armin. 1988. *Studies in tier structure.* New York: Garland.

Miller, Marion. 1999. *Desano grammar.* Arlington: Summer Institute of Linguistics & University of Texas at Arlington.

Morris, Judy. 1978. The syllable in Ẽperã. In Donald A. Burquest (ed.) *Topics in natural generative phonology*, 61–69. Dallas: Summer Institute of Linguistics.
Ní Chiosáin, Máire & Jaye Padgett. 1997. Markedness, segment realization, and locality in spreading. Technical Report LRC-97-01, Linguistics Research Center, University of California, Santa Cruz.
Noske, Manuela. 1995. The ternary use of distinctive features: In defence of the number three. Ph.D. dissertation, University of Chicago.
Odden, David. 1994. Adjacency parameters in phonology. *Language* 70. 289–330.
Onn, Farid M. 1980. *Aspects of Malay phonology and morphology*. Bangi: Universiti Kebangsaan Malaysia.
Padgett, Jaye. 1995. *Stricture in feature geometry*. Stanford: CSLI.
Pease, Helen & LaVera Betts. 1971. Parintintin phonology. In Bendor-Samuel (1971), 1–14.
Piggott, Glyne L. 1988. A parametric approach to nasal harmony. In Harry van der Hulst & Norval Smith (eds.) *Features, segmental structure and harmony processes*, part I, 131–167. Dordrecht: Foris.
Piggott, Glyne L. 1992. Variability in feature dependency: The case of nasality. *Natural Language and Linguistic Theory* 10. 33–77.
Piggott, Glyne L. 1996. Implications of consonant nasalization for a theory of harmony. *Canadian Journal of Linguistics* 41. 141–174.
Piggott, Glyne L. 1997. Licensing and alignment: A conspiracy in harmony. *Phonology* 14. 437–477.
Piggott, Glyne L. 2003. Theoretical implications of segment neutrality in nasal harmony. *Phonology* 20. 375–424.
Piggott, Glyne L. & Harry van der Hulst. 1997. Locality and the nature of nasal harmony. *Lingua* 103. 85–112.
Piggott, Glyne L. & Helga Humbert. 1997. Representations and constraints: The case of Guaraní nasal harmony. In Geert Booij & Jeroen van de Weijer (eds.) *Phonology in Progress – Progress in Phonology: HIL Phonology Papers III*, 219–256. The Hague: Holland Academic Graphics.
Pulleyblank, Douglas. 1989. Patterns of feature co-occurrence: The case of nasality. In S. Lee Fulmer, Masahide Ishihara & Wendy Wiswall (eds.) *Proceedings of the Arizona Phonology Conference* 2, 98–115. Tucson: University of Arizona.
Rice, Keren. 1993. A re-examination of the feature [sonorant]: The status of "sonorant obstruents." *Language* 69. 308–344.
Rich, Furne. 1963. Arabela phonemes and high-level phonology. In Benjamin F. Elson (ed.) *Studies in Peruvian Indian languages*, vol. 1, 193–206. Norman, OK: Summer Institute of Linguistics.
Rivas, Alberto M. 1975. Nasalization in Guaraní. *Papers from the Annual Meeting of the North East Linguistic Society* 5. 134–143.
Robins, R. H. 1957. Vowel nasality in Sundanese: A phonological and grammatical study. *Studies in Linguistic Analysis*, 87–103. Oxford: Blackwell.
Rose, Sharon & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80. 475–531.
Safir, Ken. 1982. Nasal spreading in Capanahua. *Linguistic Inquiry* 13. 689–694.
Sagey, Elizabeth. 1990. *The representation of features in non-linear phonology: The articulator node hierarchy*. New York: Garland.
Sanders, Nathan. 2003. Opacity and sound change in the Polish lexicon. Ph.D. dissertation, University of California, Santa Cruz.
Schourup, Lawrence, 1972. Characteristics of vowel nasalization. *Papers in Linguistics* 5. 530–548.
Shosted, Ryan. 2006. The aeroacoustics of nasalized fricatives. Ph.D. dissertation, University of California, Berkeley.

Solé, Maria-Josep. 2002. Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics* 30. 655–688.

Solé, Maria-Josep. 2007a. The stability of phonological features within and across segments: The effect of nasalization on frication. In Pilar Prieto, Joan Mascaró & Maria-Josep Solé (eds.) *Segmental and prosodic issues in Romance phonology*, 41–65. Amsterdam & Philadelphia: John Benjamins.

Solé, Maria-Josep. 2007b. Compatibility of features and phonetic content: The case of nasalization. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 261–266. Saarbrücken: Saarland University.

Sportiche, Dominique. 1977. Un fragment de phonologie du guaraní. Unpublished ms., MIT.

Steriade, Donca. 1993. Closure, release, and nasal contours. In Huffman & Krakow (1993), 401–470.

Steriade, Donca. 1995. Underspecification and markedness. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 114–174. Cambridge, MA & Oxford: Blackwell.

Ternes, Elmar. 2006. *The phonemic analysis of Scottish Gaelic*. 3rd edn. Dublin: Dublin Institute for Advanced Studies.

Thomas, Jacqueline. 1963. *Le parler ngbaka de Bokanga*. The Hague: Mouton.

Thomas, Jacqueline. 1970. *Contes ngbaka-ma'bo*. Paris: Éditions Klincksieck.

Trigo, Loren. 1988. On the phonological derivation and behavior of nasal glides. Ph.D. dissertation, MIT.

Trigo, Loren. 1993. The inherent structure of nasal segments. In Huffman & Krakow (1993), 369–400.

Viljoen, Johannes, J. 1973. *Manual for Ndonga*. Part 1. Pretoria: University of South Africa.

Walker, Rachel. 1999. Guarani voiceless stops in oral *vs.* nasal contexts: An acoustical study. *Journal of the International Phonetic Association* 29. 63–94.

Walker, Rachel. 2000a. *Nasalization, neutral segments, and opacity effects*. New York: Garland.

Walker, Rachel. 2000b. Yaka nasal harmony: Spreading or segmental correspondence? *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 26. 321–332.

Walker, Rachel. 2003. Reinterpreting transparency in nasal harmony. In van de Weijer *et al.* (2003), 37–72.

Walker, Rachel. 2007. Nasal and oral consonant similarity in speech errors: Exploring parallels with long-distance nasal agreement. *Language and Cognitive Processes* 22. 1073–1113.

Walker, Rachel & Geoffrey K. Pullum. 1999. Possible and impossible segments. *Language* 75. 764–780.

Weijer, Jeroen van de. 1994. Stem consonant co-occurrence restrictions in Ngabaka. In Reineke Bok-Bennema & Crit Cremers (eds.) *Linguistics in the Netherlands 1994*, 259–266. Amsterdam & Philadelphia: John Benjamins.

Weijer, Jeroen van de, Vincent van Heuven & Harry van der Hulst (eds.) 2003. *The phonological spectrum*, vol. 1: *Segmental structure*. Amsterdam & Philadelphia: John Benjamins.

Wescott, Roger W. 1965. Review of Thomas (1963). *Language* 41. 346–347.

West, Birdie & Betty Welch. 1967. The phonemic system of Tucano. In Benjamin F. Elson (ed.) *Phonemic systems of Colombian languages*, 11–24. Norman, OK: Summer Institute of Linguistics.

Williamson, Kay. 1965. *A grammar of the Kolokuma dialect of Ịjọ*. Cambridge: Cambridge University Press.

Williamson, Kay. 1969. Ịjọ. In Elizabeth Dunstan (ed.) *Twelve Nigerian languages*, 97–114. New York: Africana Publishing.

Williamson, Kay. 1987. Nasality in Ịjọ. In David Odden (ed.) *Current approaches to African linguistics*, vol. 4, 397–415. Dordrecht: Foris.

Wright, James T. 1986. The behavior of nasalized vowels in the perceptual vowel space. In John J. Ohala & Jeri J. Jaeger (eds.) *Experimental phonology*, 45–67. New York: Academic Press.

79 Reduction

Natasha Warner

1 Introduction

[b̥ɹʌʒləɪ̯] is a four-word phrase of American English (Figure 79.1). What might it be? How could a listener map this phonotactically improbable string back to the speaker's intended four-word phrase? Why would a native speaker produce this realization? The intended phrase was *but I was like*, using the quotative *like* in the utterance *but I was like, "why wouldn't you just go home?"*. The speaker was a young female native speaker holding a casual telephone conversation with a friend. The fact that /bʌt aɪ wʌz laɪk/ can be produced as [b̥ɹʌʒləɪ̯] is surprising, but

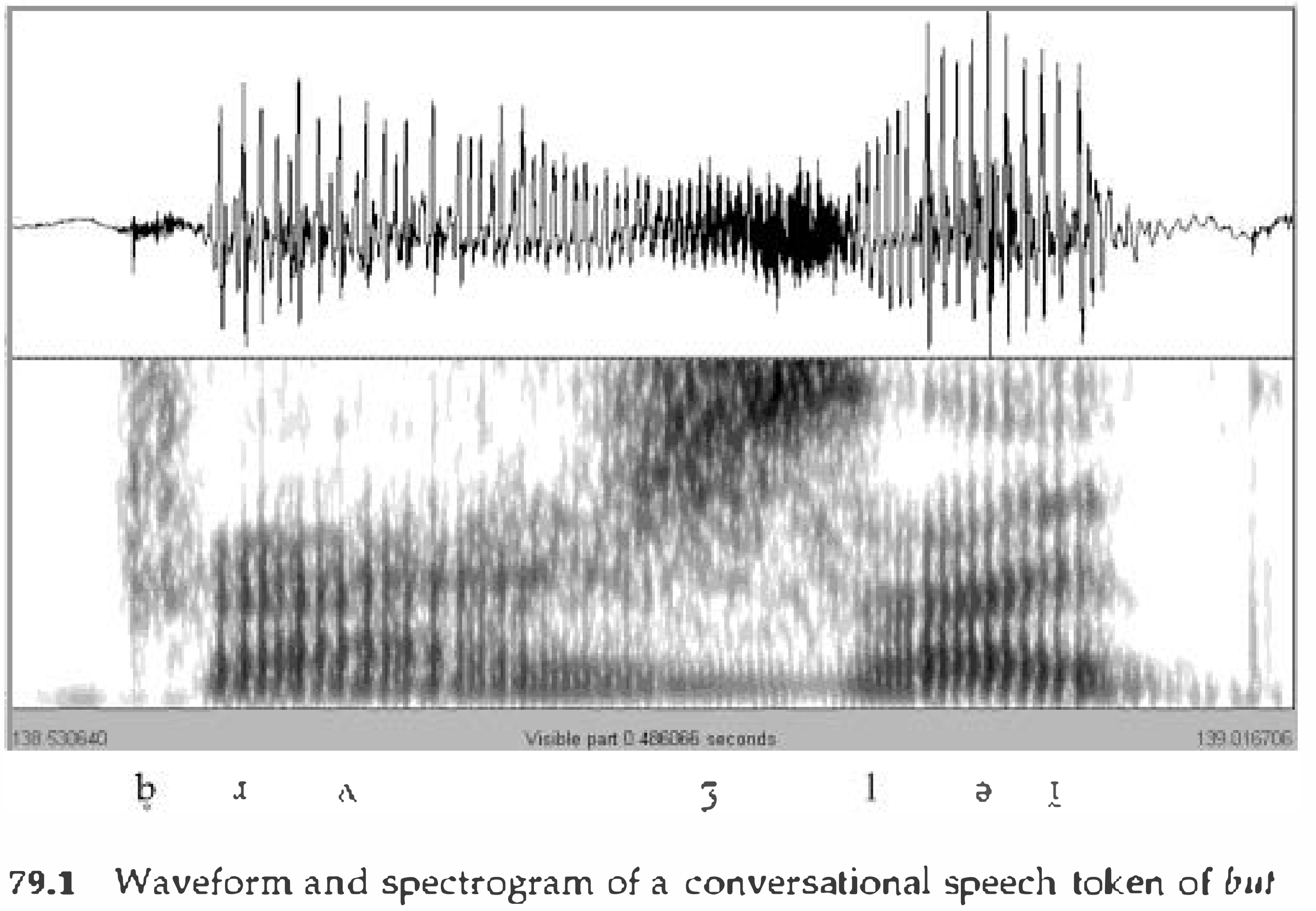


Figure 79.1 Waveform and spectrogram of a conversational speech token of *but I was like*, demonstrating numerous deletions, mergers, and changes to segments, as well as insertion of unexpected /r/-coloring, perhaps as a reflection of the flap of *but*. All spectrograms show a 0–5000 Hz range

perhaps even more surprising is that when a native listener hears the utterance, it sounds like a normal pronunciation of the phrase. While one might not expect this realization, when one examines spontaneous conversation, such surprising reductions are rather common. The fact that a surface realization can depart so drastically and in variable ways from the underlying representation poses problems for phonological theory. The fact that listeners find it unproblematic brings up further questions regarding what a phonological underlying representation is, and how speakers and listeners use it. There are three directions from which one might approach the topic of reduction: what sounds the speaker produces (acoustics of reduction), how the speaker produces them (reductions and changes in articulations), and how the speech is perceived.

Phoneticians and phonologists have long known that speakers sometimes produce reduced speech, at least in casual speech situations. [ʤiʔjɛʔ] for *did you eat yet?* is often used as a plausible example. However, such examples are sometimes presented in phonetics and phonology classes as something that occurs when one is at home talking with one's spouse, perhaps when one is not fully awake, not as a normal phenomenon. Such speech has often been removed from the realm of phonology, and left for phonetic implementation. Yet phoneticians have often paid hardly more attention to reduced speech, considering it too uncontrolled to give good results. Ladefoged (2003) suggests avoiding connected speech such as storytelling when documenting a language, and advises sticking to word lists in controlled frame sentences. Many of the chapters in the *Companion* summarize a debate about how phonological theories should handle a particular phenomenon. This one, instead, summarizes a debate about whether a phenomenon is even relevant.

The default assumption may be that reduction is not linguistically relevant. Speech style (e.g. casual, formal) and speech rate seem like non-linguistic factors, outside the boundaries of the abstract system that makes up the grammar. Or perhaps speech reduction really does occur primarily when conversing with one's spouse, at home, when tired – perhaps it is the exception. However, there is also reason to think that reduced speech is anything but peripheral to the linguistic system. It may be in fact the normal, typical way for humans to communicate information. Furthermore, neither phonetic nor phonological theories were built to handle reduced speech, so reduced speech raises large theoretical questions.

First, I will discuss terminology. There are many overlapping terms falling along more than one dimension, such as reduced, conversational, connected, spontaneous, fast, casual, and natural speech. One could separate a speech rate dimension (fast–slow) out from the dimension of formality (casual–formal) (see also CHAPTER 92: VARIABILITY). Neither of these is exactly a dimension of "reduction": some speakers talk very quickly in both formal and informal settings, yet seem to maintain almost all of their consonantal articulations. However, reduction is probably more common in casual conversation and fast speech.

Another dimension might be spontaneity. In my own understanding of the term, "spontaneous speech" refers to any speech in which the words are not chosen ahead of time. This excludes read speech, speech in a talk one has given repeatedly (e.g. politicians' campaign speeches), and speech that is explicitly prompted (e.g. shadowing tasks). When the Oregon Graduate Institute (OGI) collected its Japanese corpus (Muthusamy *et al.* 1992), for example, speakers were instructed to talk about themselves (e.g. hobbies, work) for a minute, which leads to spontaneous monologue speech. I take "connected" speech to be broader than "spontaneous speech," including anything where words form part of a longer utterance, perhaps even target words in a frame sentence.

"Conversational" speech sets a stricter requirement, meaning speech that occurs while two or more speakers are conversing, whether by telephone or in person, and whether they know each other or not. Speech can be spontaneous and connected but not conversational, as in the OGI monologues. "Conversational" speech excludes any scripted speech, even reading a scripted conversation. "Casual" speech goes one step further, requiring that the speakers be comfortable in the setting, comfortable with the topics and comfortable with their interlocutors. A job interview or a discussion of governmental policy is usually conversational and spontaneous, but not casual. However, monologues about hobbies or friends can be casual without being conversational if the speakers are comfortable with the recording setting.

Three terms differ from the rest: fast, natural, and reduced speech. Speech rate, measurable in intended syllables per second, for example, does seem to be a separate but usually correlated dimension. For example, some parts of spontaneous conversational speech are very fast, while others may be very slow (perhaps if the speaker is tired or uninterested in the topic). However, one may expect that casual conversation would usually have a faster rate in intended syllables per second than most careful read speech. "Naturalness" defies a simple definition. One can refer to natural speech as the opposite of synthesized speech, so that even nonsense CV syllables can be natural if produced by a human. Alternatively, linguistic anthropologists may put very strict requirements on the setting of "natural speech," such as the speakers not being seated in a sound booth, or not having head-mounted microphones on, as these things might make the speakers self-conscious, and affect their speech. Thus I will not attempt to fit "natural speech" into the other dimensions.

I take "reduced speech" to refer to changes in the segments or suprasegmentals relative to what would be expected in a careful pronunciation of the same word or phrase. "Reduction" thus includes changes to sounds (CHAPTER 66: LENITION; e.g. [ɣ̞] or [ɰ] for [g] in *gonna*; Figure 79.2), deletions of expected segments (CHAPTER 68: DELETION; e.g. [wɪə] for *we were*; Figure 79.2), and shrinkage of contrast space (e.g. a smaller overall vowel space, or a specific centering like [wʌ̃] for *when*; CHAPTER 2: CONTRAST). Types will be discussed below. Clearly, however, not all alterations of segments are reduction: speech errors, dialectal differences, historical sound change (CHAPTER 93: SOUND CHANGE), and obligatory (morpho-)phonological changes (e.g. /k/ becoming /s/ in *electric/electricity*; Chomsky and Halle 1968) are not reduced speech. "Reduction" refers to changes relative to a clearly pronounced surface form, not relative to the underlying form. Thus, spontaneous or conversational speech can be defined based on the setting in which it is produced, but I define reduced speech based on its acoustic or articulatory form. Notably, reduction does occur in relatively careful speech, although not as often as in spontaneous casual speech, so setting or style alone does not define it. I do not define reduction based on it being easier to articulate, or requiring less movement of the articulators, because it is difficult to define objectively for each situation what "easier to articulate" would mean.

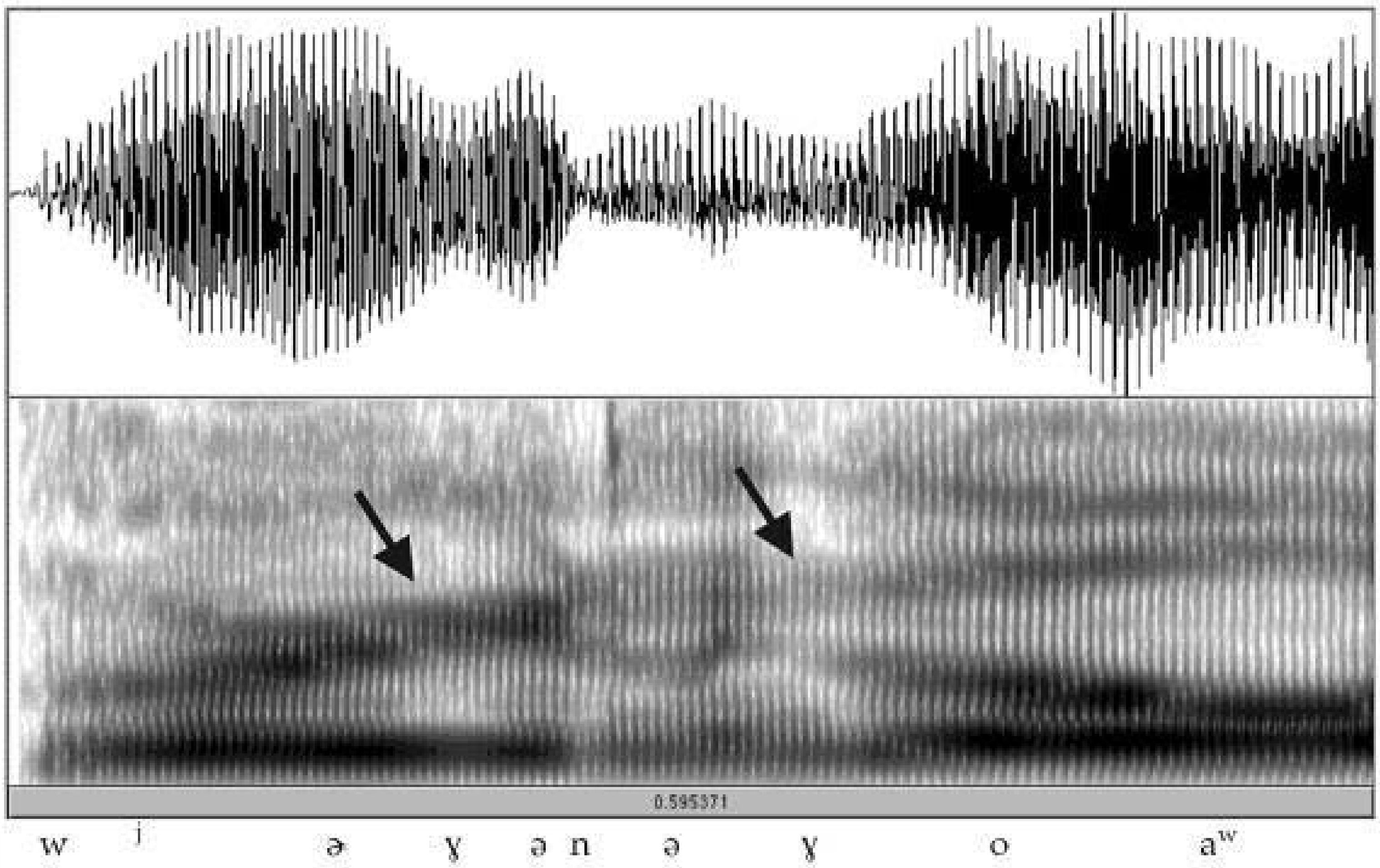


Figure 79.2 Waveform and spectrogram of a conversational token of *We were gonna go out*, demonstrating approximant realizations of the /g/ phonemes (marked with arrows)

2 Historical perspective

Scholars have long noticed that expected speech sounds are not always produced. Richter (1930) gave numerous examples of reduction in French, although from relatively careful speech. She notes severe reductions like the deletion of the entire syllable *-ble* in *impossible*, as well as single-segment deletions, assimilations, etc. Early scholars like Richter had to develop creative methods to detect reductions, since in context reductions sound notoriously unremarkable. Richter recorded speech on phonograph records, then played the records backwards while transcribing the sounds she heard. While this may have biased her results in some ways, it was a creative way to determine what kinds of reductions were present, without the benefit of spectrograms.

For several subsequent decades in phonetics and phonology, many scholars acknowledged that speech reduction exists and gave a few examples of reductions that might occur in conversation or in "sloppy" speech. However, they usually considered reduction phenomena as outside the area of interest, stating that one should analyze the full form of a word rather than the "elliptic," "slovenly," or "slurred" forms one finds in "rapid speech" or "familiar talk" (all terms from quotations in Johnson 2004). Johnson (2004) cites Hockett (1955) as giving examples [dɪʤə] for *did you?* and [wuʤə] for *would you?*. Johnson also discusses how Chomsky and Halle's focus on competence rather than performance removed speech reduction from the field of data to be analyzed. Johnson (2004) provides an excellent historical summary. Johnson does point out that European phonology (he emphasizes Stampe's Natural Phonology) gave more attention to reduction. For example, Dressler (1975) discusses implications of reduction for analysis of historical sound change, and advocates Natural Phonology as the best formal phonological approach to reduction. Shockey (2003) presents numerous transcriptions of reductions, and discusses how various phonological theories, including Natural Phonology, would handle reduction. Zwicky (1972) and Bolozky (1977) also address how to generate variable reduced surface forms in phonology. However, the publications that do address reduction represent a minuscule proportion of the work in the field, and papers not explicitly on reduction generally only mention careful speech forms. The training which new members of the field receive can also give an impression of the field's stance on reduction: one graduate introductory phonology course in an American linguistics department in the 1990s included a few examples of minor reductions such as [mɪʃʲu] *miss you* and a statement that what we should be studying is the clear speech form of words, not reductions. Overall, the field of phonology is primarily characterized by work that stops at the careful speech surface level. Many of the phonological works that do address reduction, furthermore, discuss reductions of a single segment, rather than the more drastic deletions of multiple syllables found in conversational corpora.

Turning to phonetics, Johnson (2004) notes that phoneticians have discussed reductions. The earliest example he cites is Dalby (1986), who examined how often schwas (CHAPTER 26: SCHWA) were deleted in English television talk-show speech and in a fast speech task in the lab. Richter (1930), mentioned above, gives an early example. Ladefoged *et al.* (1976) recorded speech styles ranging from conversational interviews to isolated word-list reading, and failed to find reduction of the vowel space. Koopmans-van Beinum (1980), however, found exactly that effect (in Dutch), in a wide variety of speech styles. A search of the *Journal of the Acoustical Society of America*'s database, going back to 1929, does show early studies of conversational speech, although often for engineering purposes, e.g. earplug effectiveness (Kryter 1946).

Still, the vast majority of phonetic research for many decades has been on carefully read speech. To take just one example, Löfqvist and Gracco's (1999) work using read sentences such as *Say "ipa" again* is more representative of the field than Dalby's (1986) work using television talk-show recordings. This is not a fault of the field. In part, technological limitations hindered analysis of large amounts of spontaneous conversational speech. Furthermore, there are many interesting questions that are more appropriate to answer with carefully controlled speech than with open, uncontrolled, variable conversation.

To confirm objectively that research using spontaneous or non-careful speech is still the exception in phonetics, I examined every article in a recent volume of the *Journal of Phonetics* (vol. 36, 2008). This volume contains 36 research articles, including a special issue on phonetics of North American indigenous languages (six articles). Of the 36, a total of four articles use speech material occurring in a relatively natural setting. Notably, two of those are on infants' speech productions during play (isolated words and babbling). The other two are from the special issue on indigenous languages and use field recordings (interviews, monologues, and elicitation) as corpora. One additional indigenous languages article uses a sentence translation task. Thus, in this volume of the journal, relatively natural speech settings seem to be used when it would be difficult or impossible to obtain data otherwise: with infants and with highly endangered or non-written languages. (One additional article on an African language uses a sentence repetition task rather than reading.) The remaining 32 articles use careful speech. The majority use target words or non-words in repetitive frame sentences of the *Now I say X again* type. Thus, although the sources discussed below demonstrate that there is now a substantial body of research on spontaneous speech and reduction, such research represents a small percentage of work in the field.

Until recently, speech perception and spoken word recognition research has also largely avoided speech that might be reduced. Speech perception stimuli tend to be even more careful and less natural than the speech for acoustic phonetics: a common type of stimulus might be synthesized /ba da ga/ nonsense syllables, or perhaps isolated words with one segment gated out. One would be unlikely to use stimuli collected from natural conversation unless perception of reduction were the topic. Turning to psycholinguistics, much of Cutler and colleagues' work is about how listeners segment words out of connected speech. Still, Cutler points out (1998; see also Mehta and Cutler 1988) that one usually uses carefully pronounced stimuli with little context (e.g. *apple* embedded in *vufapple* for a word-spotting task) to study the segmentation of connected speech, reducing experimental variability. Thus, even work that sets out to examine how one deals with connected speech uses rather careful speech as the testing ground.

Recently, phonetic investigation of reduced speech has been expanding rapidly. Three studies quantify how much reduction takes place overall by comparing pronunciations in corpora of spontaneous speech to expected careful pronunciations. Shattuck-Hufnagel and Veilleux (2007) surprisingly emphasize how *few* of the expected acoustic landmarks are deleted in spontaneous speech (14 percent in their study), while Greenberg (1999) emphasizes how *many* expected segments are deleted (12.5 percent in his study). Johnson (2004) finds 20 percent of words having at least one segment deleted and 5–6 percent of words having at least a syllable deleted. Although one paper concludes that deletions are rather rare and the other two conclude that they are common, the difference in deletion rate is not large, and the methods differ (deletion of Stevens-style landmarks *vs.* deletion of segments, in a map task – where speakers viewing a map give directions to listeners – as opposed to conversation). Thus, whether deletion is frequent may be a question of whether the glass is half full or half empty: is 12.5–14 percent deletion a lot? Is it enough to make listening challenging? Work such as that by Raymond *et al.* (2006) examines deletion of particular segments, also showing substantial rates of deletion, even for word-medial consonants. Reduction is not solely about deletion: Greenberg finds 117 pronunciations for the word *that*, for example, an astounding demonstration of the variability that listeners encounter in normal conversation. Strik *et al.* (2010) show a similar result for Dutch. Shockey (2003) contributes transcriptions of reductions from many dialects of English.

Ernestus and colleagues have conducted extensive research on reduction in Dutch (beginning with Ernestus 2000). Some of their studies locate all tokens of particular high-frequency words or suffixes in a corpus, such as *eigenlijk* 'actually' or its suffix *-lijk*, and analyze what affects their duration (Pluymaekers *et al.* 2005a, 2005b). Research on reduction has favored English and Dutch thus far, but does exist for other languages. Some examples are Tseng (2005) and Cheng and Xu (2008) on Mandarin, Furui and colleagues (e.g. Nakamura *et al.* 2007) and Maekawa and Kikuchi (2005) on Japanese, Kohler (2001) and colleagues on German, Engstrand and Krull (2001) for Swedish, Lennes *et al.* (2001) on Finnish, and Nicolaidis (2001) on Greek. The Nijmegen Speech Reduction Workshop (June 2008) included several talks on French, and Barry and Andreeva (2001) compare reduction phenomena in six languages. Furthermore, Keune *et al.* (2005)

find that degree of reduction can differ even for dialects, with more reduction in Belgian Dutch than in Netherlands Dutch. It is promising that there are data in an array of languages, although Europe still dominates. Another consideration about the past literature is that much of the spontaneous speech has been from the Map Task (Bard *et al.* 2001; Shattuck-Hufnagel and Veilleux 2007), monologues (Arai 1999), or other relatively careful speech, which may lead us to underestimate reduction. There has also been phonetic research on speech style in the opposite direction from reduced speech, specifically on the "clear" speech that speakers use to address hard-of-hearing or second-language listeners (Bradlow and Bent 2002; Smiljanić and Bradlow 2005, 2008).

Most phonetic research on reduction is on what is produced (articulatory or acoustic phonetics), but perception research is now increasing rapidly. Mehta and Cutler (1988) test perception of spontaneous *vs.* read speech, using a phoneme monitoring task, and find that the coherent intonation of planned speech facilitates processing, relative to the pauses and self-corrections of spontaneous speech. Koopmans-van Beinum (1980) finds that identification accuracy for Dutch vowels presented in isolation, cut out of the speech stream, is extremely low for casual speech. Arai (1999), on Japanese, and Ernestus *et al.* (2002), on Dutch, both investigate perception of severe reductions (e.g. syllable deletions, such as [ɛik] for Dutch [ɛixələk] 'actually'). Both Arai and Ernestus *et al.* find that listeners cannot retrieve the intended words well without context, but can with context. Shockey (2003), using a single conversational utterance, indicates that listeners usually misperceive reduced speech unless extensive conversational context is present.

Turning to perception of specific types of reduction, Mitterer and Ernestus (2006) find that listeners take patterns of where reduction is produced into account when recognizing reduced final /t/. Ranbom and Connine (2007) investigate recognition of English words with /nt/ reduced to nasal flap, as in *gentle* rhyming with *kennel*. They find that words which listeners hear more often with nasal flap are not as hard to recognize with the nasal flap as other words. However, the reduced form is still more difficult overall. Tucker (2007) finds that having heard fast, reduced speech in the preceding frame sentence can, to some extent, mitigate the reduced-speech difficulty. This suggests that listeners use information about reduction in the context to adjust their acoustic criteria for recognition of upcoming sounds. Isabelle Racine presented on recognition of French words with schwa deletion at the Nijmegen Speech Reduction Workshop. Oliver Niebuhr, also at that workshop, discussed listeners' use of lengthening in neighboring segments to recognize otherwise deleted segments. Warner *et al.* (2009b) test the contribution of various acoustic aspects of flap reduction to whether listeners hear a medial consonant in pairs such as *needle–kneel*.

Past research has also examined processing of flapped *vs.* extremely careful /t d/ pronunciations, e.g. recognition of *pretty* with flap *vs.* [t] in American English (McLennan *et al.* 2003, 2005; Connine 2004; CHAPTER 113: FLAPPING IN AMERICAN ENGLISH). However, since flap is the normal careful pronunciation, that work is more about the processing of phonological alternations than about reduction.

Speech reduction borders on other fields besides linguistics and psychology. It is particularly relevant for engineering, for purposes of automatic speech recognition (ASR), because speakers may expect ASR systems to respond correctly to their requests even if they use reduced speech. Some of the works cited above relate particularly to ASR (Greenberg 1999; Strik *et al.* 2010; Nakamura *et al.* 2007).

Within the field of linguistics itself, but beyond phonetics and phonology, speech reduction is of particular interest for sociolinguistics.

To sum up the historical perspective on reduction, formal phonology has long excluded reduction from its domain. Phonetics has long included the topic, but as a very small proportion of all phonetic research. Psycholinguistics has addressed questions of connected speech, although usually by means of careful speech stimuli. The last few years, however, have shown an explosion of research on reduction, speech style, and large connected speech corpora.

Perhaps it is not surprising that conversational speech and reduced speech make up a small proportion of research, even in phonetics. The field uses an astounding range of speech carefulness, from nonsense sequences such as /ipa, ipu/, possibly with articulograph pellets glued to the speaker's tongue, to spontaneous conversation between close friends, or even natural interactions outside the lab (e.g. Hicks Kennard's 2006 work on male and female marine corps drill instructors' phonetics). The former will yield extremely stable data with little random variability, with obvious statistical advantages (increased power) when the question is not about more natural speech. However, once there is a background literature on a given topic, we may be able to move to more spontaneous speech. We are then likely to encounter more reductions, even if reduction is not the topic of the work.

3 Reduction phenomena

3.1 *Duration*

Shorter duration of segments is perhaps an obvious meaning of reduction (duration is literally reduced), but if the segments are acoustically otherwise unaltered, overall shorter durations might stem from fast speech rather than reduction. However, shorter duration is usually correlated with reduction in manner of articulation, so duration may provide a convenient way to measure reduction. For example, we find a good correlation between intervocalic stop duration and how approximant-like the "stop" is (Warner and Tucker 2007). Ernestus and colleagues have made extensive use of duration as a simple, one-dimensional indicator for reduction (e.g. Pluymaekers *et al.* 2010). Figure 79.3 shows a phrase taken from casual conversation, with reduction of several syllables to very short durations, and the same phrase in careful speech. Ernestus and colleagues often use a partially automated system (automatic speech recognition, given phonetic transcription as input) for duration measurement (Pluymaekers *et al.* 2010).

3.2 *Alterations to segments*

Changes in the manner of articulation or voicing of a segment are extremely common in reduced speech. Warner and Tucker (2007), studying realizations of expected intervocalic stops and flaps, find that a great many tokens in all speech styles are actually produced as approximants (Figure 79.4a, even in read speech). Figure 79.4b shows an expected flap that is barely visible, and a second visible but extremely short flap. Figure 79.5 shows a Japanese example in which reduction causes devoicing rather than voicing (in an environment where vowel

(a) First pair with transcription

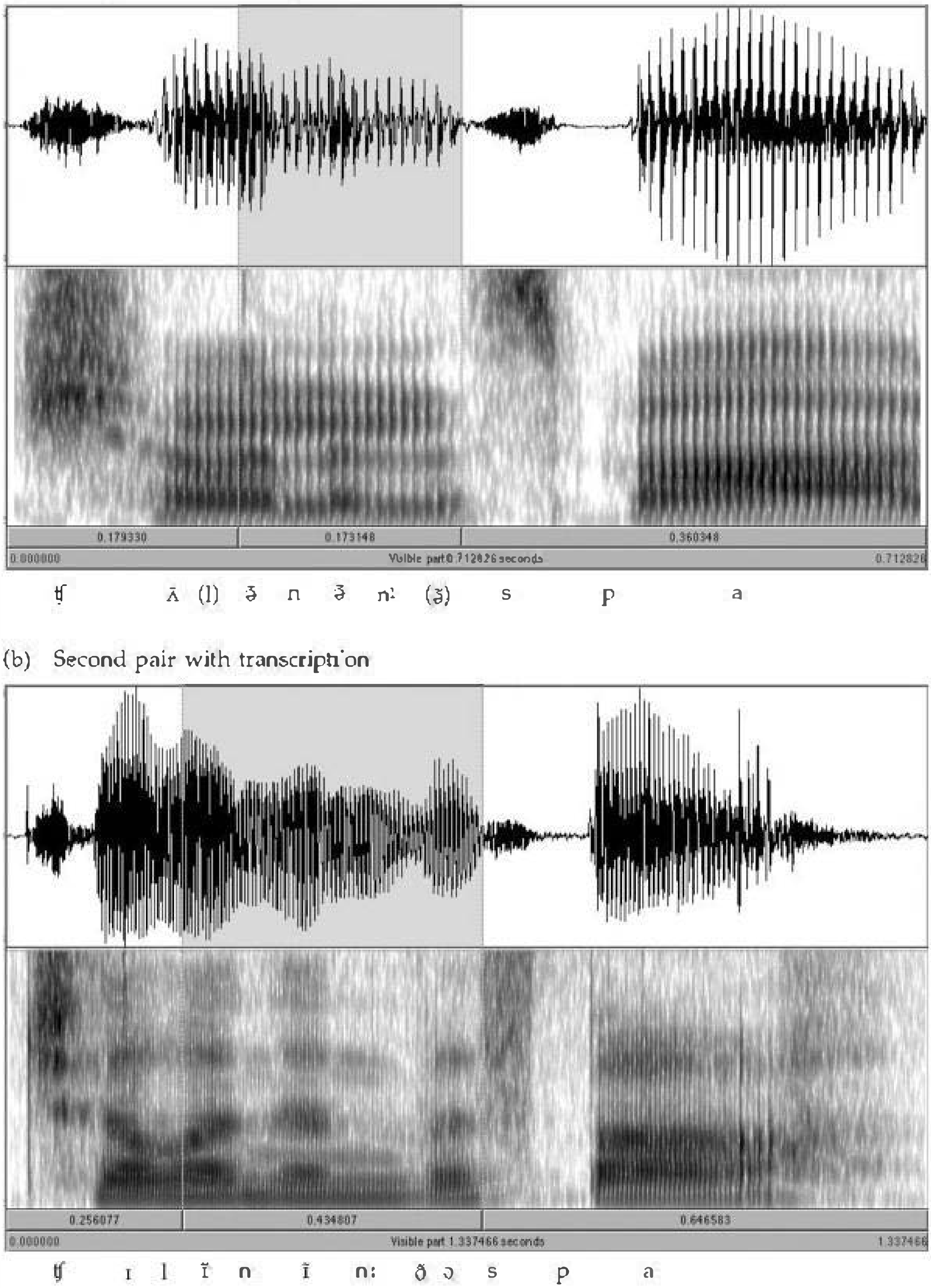
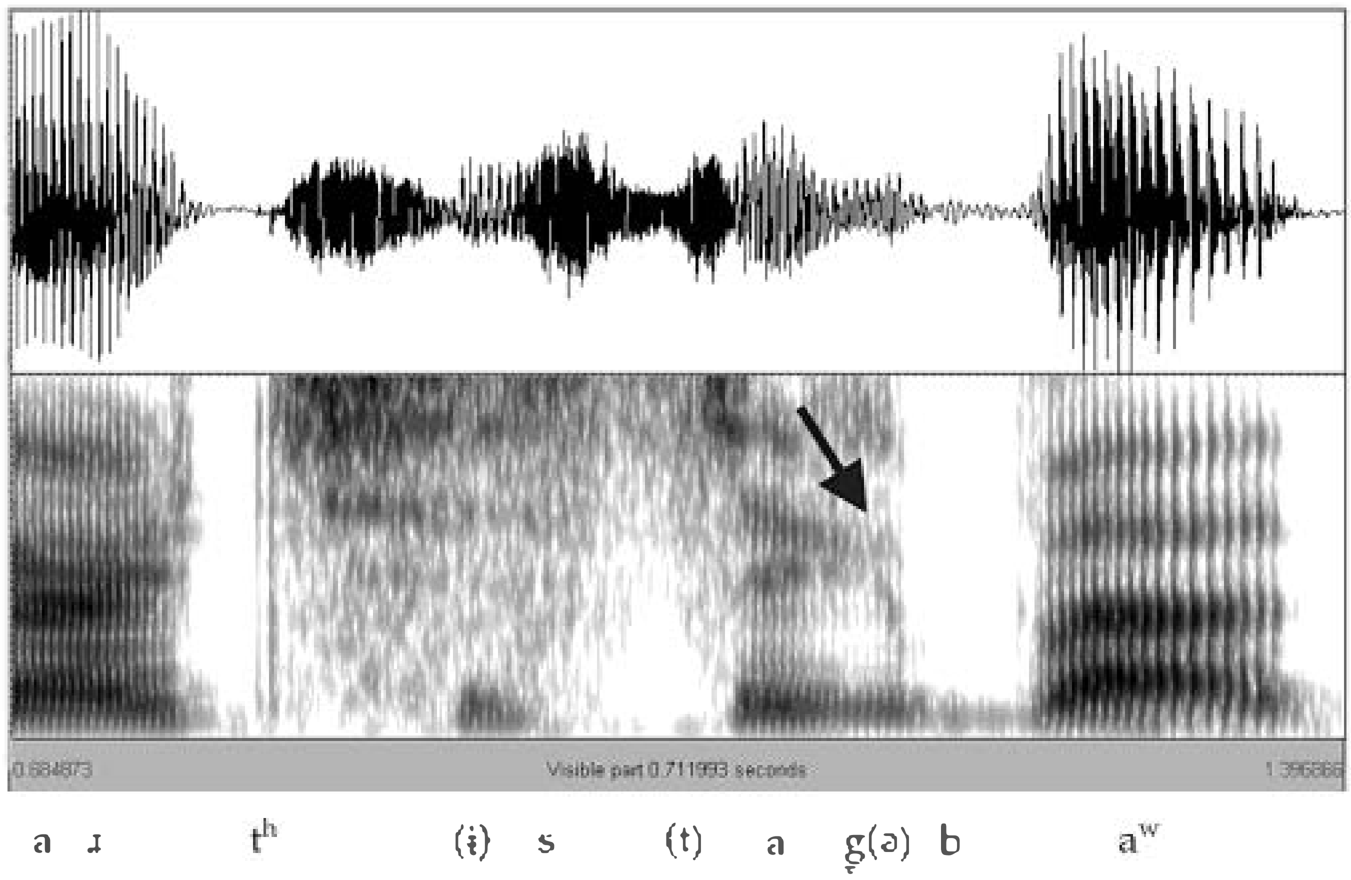


Figure 79.3 Waveforms and spectrograms of two tokens of *chillin' in the spa.* The highlighted portion corresponds to *-in' in the*. In all figures, phonetic symbols in parentheses indicate sound for which there is some evidence, but either they have extremely low amplitude or their presence is in doubt. (a) Utterance taken from conversation (highlighted portion: 173 msecs). (b) Carefully produced utterance of the same words (highlighted portion: 435 msecs)

(a) First pair with transcription



(b) Second pair with transcription

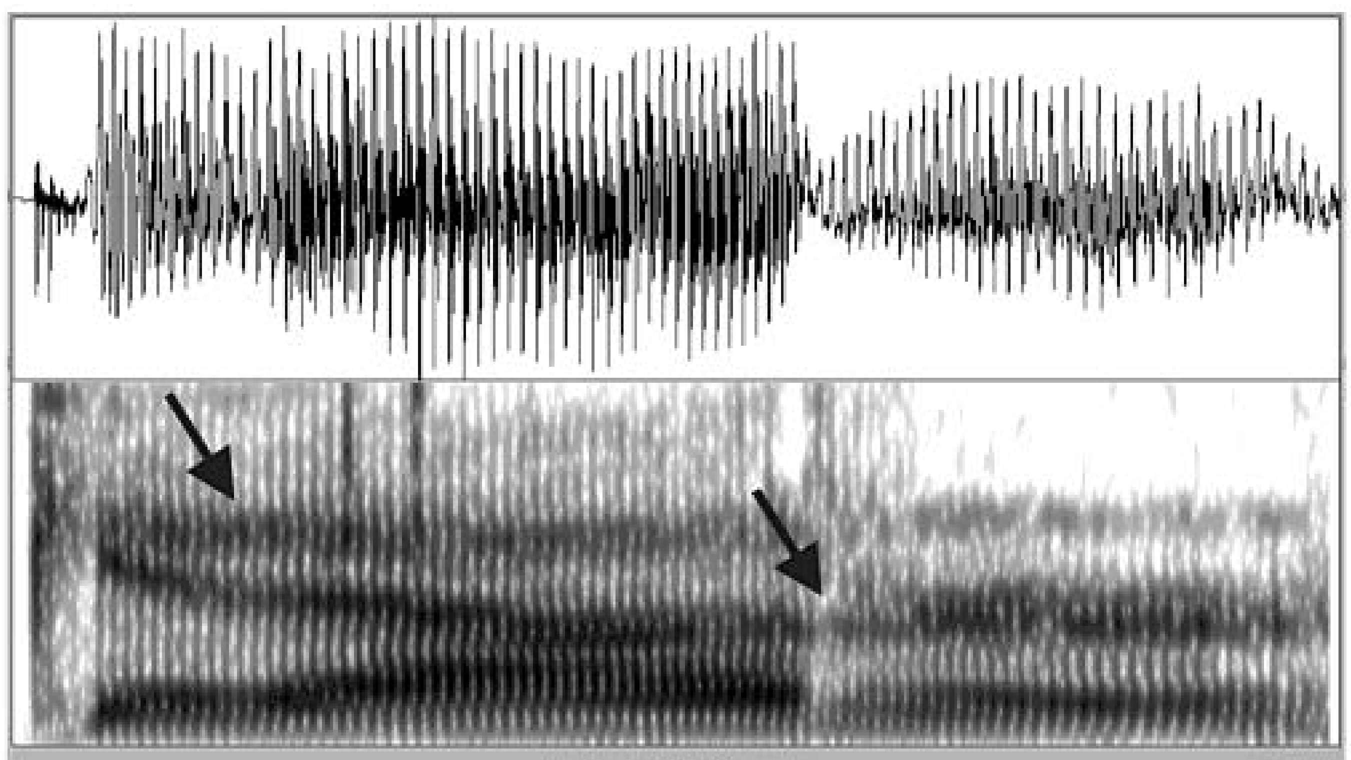


Figure 79.4 Waveforms and spectrograms showing reduced stops and flaps.
(a) . . . *artistic about* (read speech from a list of phrases), with /k/ marked by an arrow.
(b) . . . *get out, er* . . . (conversation), with arrows at expected flaps

devoicing should be phonologically impossible). The entire final syllable of the utterance, /ru/, is effectively devoiced, although the /u/ could be argued to have 3–4 extremely low-amplitude pulses.

If manner of articulation (CHAPTER 13: THE STRICTURE FEATURES) and voicing (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION) can change as part of reduction, one wonders whether place of articulation (CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION) can as well. One might hesitate to consider most

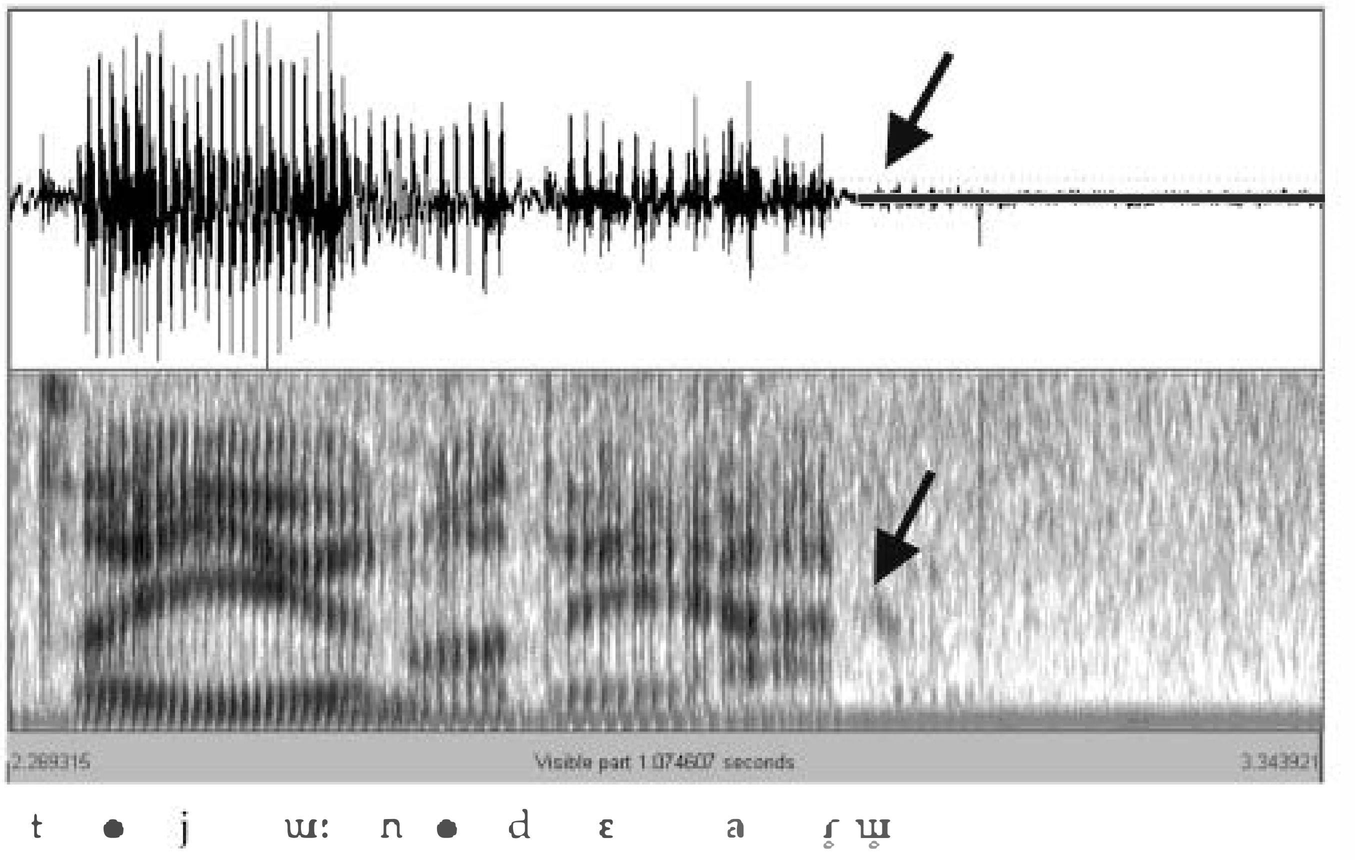


Figure 79.5 Waveform and spectrogram of a Japanese partial utterance . . . *to iu no de aru* 'it's that it's the case that' (from a cassette tape accompanying a third-year language textbook, read speech by a professional speaker). The arrow marks a consonant and vowel that are devoiced where not phonologically expected

assimilations (e.g. English /ɪn-/ in *impossible*; CHAPTER 81: LOCAL ASSIMILATION) to be reductions. However, variable phonetic assimilations may be part of reduction: *seve*[m] *plus* for *seven plus* would happen less often in careful speech, even though the tongue tip still makes an unperceived alveolar gesture (Browman and Goldstein 1989). This type of example has influenced the field through Articulatory Phonology (CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATIONS), although it is not generally presented as being "reduced" speech.

3.3 *Deletion and syllable count*

In reduced speech, one almost always finds deletions of segments (Figures 79.1–79.3), as quantified by Greenberg (1999), Johnson (2004), Raymond *et al.* (2006) and others. Some scholars might propose that deletion is not a type of reduction, because reducing something does not mean removing it. However, it is impossible to separate deletions from other reduced speech phenomena. Even for a given segment, it can be difficult to say whether there is some trace of the segment present. Figure 79.6 shows several examples ranging from reduced to probably deleted (even in Figure 79.6c, one perceives some trace of a consonant for the expected flap, although it is not visible in the spectrogram).

Furthermore, Browman and Goldstein (1989) review findings that, even when a segment appears to be fully deleted acoustically, there may still be a reduced-size articulatory gesture for it, as in the apparent deletion of the final /t/ in *perfect memory*. This is very similar to the *seve*[m] *boys* type of assimilation example above, and in Articulatory Phonology, the difference between deletion (*perfect memory*) and this assimilation is solely a matter of whether the following labial gesture overlaps a voiceless stop, rendering it inaudible, or a nasal, making it audible with

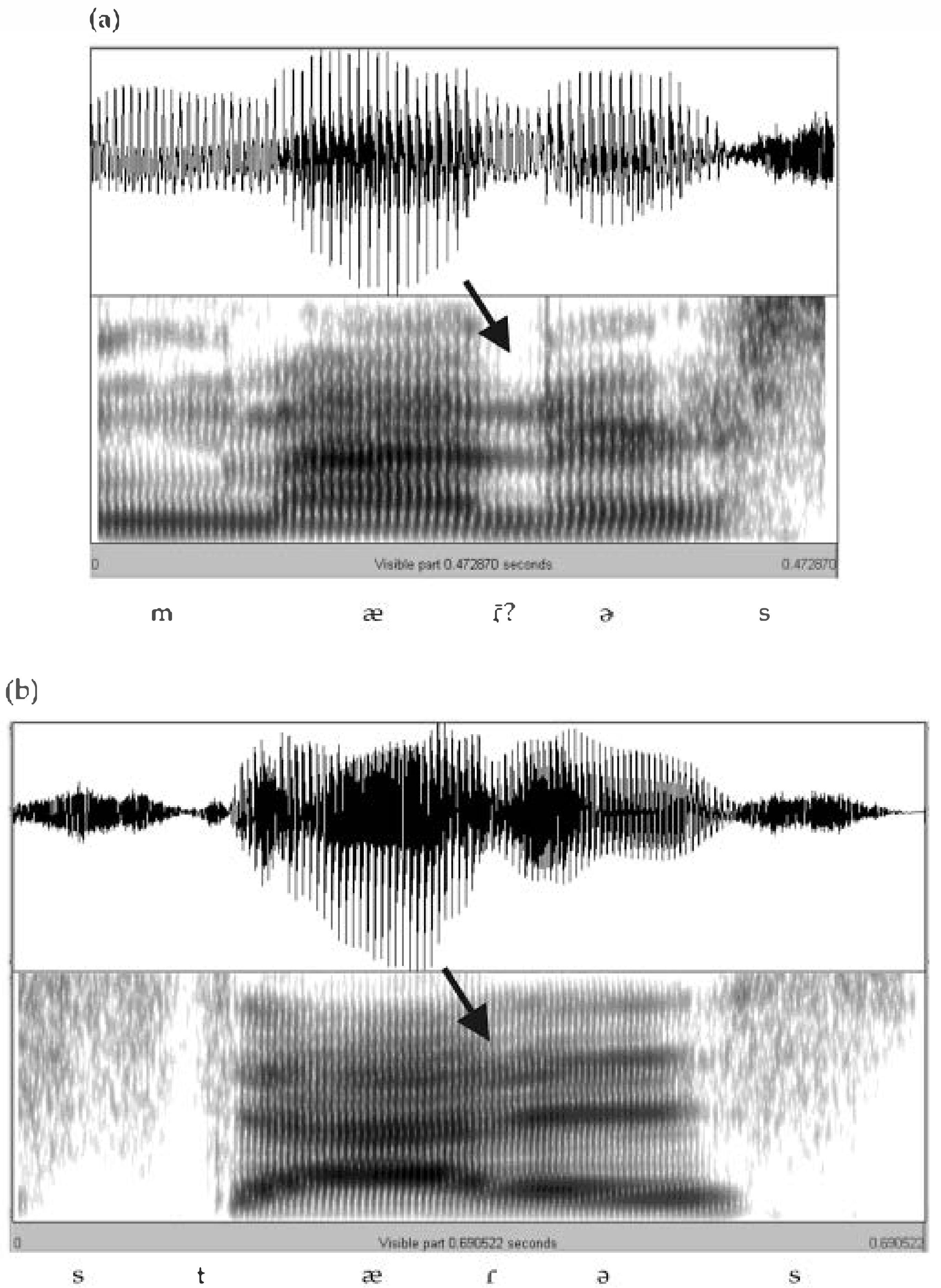


Figure 79.6 Waveforms and spectrograms of a variety of reductions of flap. Arrows mark the location, visible or expected, of the flap. All are from read speech ((a) from the reading of a story; (b) and (c) from word-list reading). (a) *matters*, with the flap slightly nasalized. (b) *status*, with the flap approximated. (c) *capitalist*, with the flap so reduced that no trace is visible, although there is some suggestion of a consonant perceptually

(Cont'd)

(c)

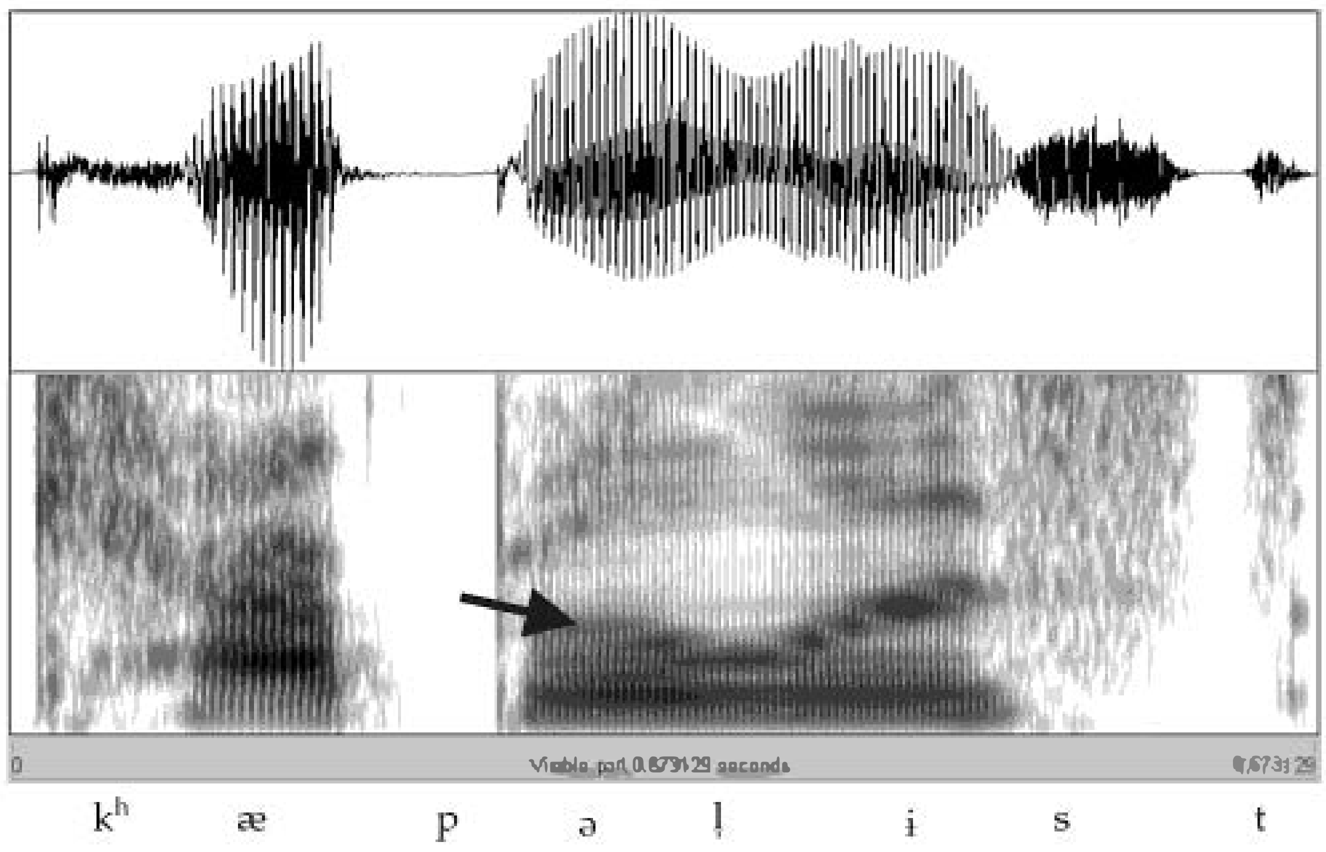


Figure 79.6 (*Cont'd*)

a different place. Thus there are several reasons to consider apparent segmental deletions to be quite literally a type of reduction: reduction *vs.* deletion is a continuum (Figure 79.6), acoustic deletions are known to often have small residual gestures (Browman and Goldstein 1989), and gesturally there is no difference between some deletions and some assimilations. Furthermore, deletion and reduced segments go together: in conversational or spontaneous speech, if we are finding many changes of manner or voicing in a recording, we will also be failing to find some expected segments at all.

Often, reduced speech overlaps gestures and perceptual cues so much that one cannot say which segments have been deleted and which are present. One hears some trace of expected segments, but on close examination of the spectrogram, and despite close listening, one cannot be sure what sounds are present at all (Figure 79.7). In *weekend were you* in Figure 79.7, there is a voiceless palatal fricative for the /k/, and later a high F2 for the /j/ of *you*. Between those, there is a vocalic stretch, then a low F2 that suggests a /w/, and one hears substantial nasalization and some r-quality. This unclear stretch must be a reduction of /ɛndwɚ/, and one can identify features such as nasalization, but it is very difficult to identify segments or transcribe it in order to determine what has been deleted. Any transcription of this stretch forces unjustified and artificial clarity onto the speech. Still, this utterance sounds completely normal and intelligible. Such examples make the issue of whether deletion and assimilation are part of reduction moot. Gestures and perceptual cues are highly overlapped, yet something remains of many of them. Both assimilation and deletion must be happening, but the speech is simply not clear enough to point to specific assimilations or deletions.

As a result of deletions, syllable count often drops relative to careful speech. As discussed above, Johnson (2004) finds that 5–6 percent of words in a spontaneous

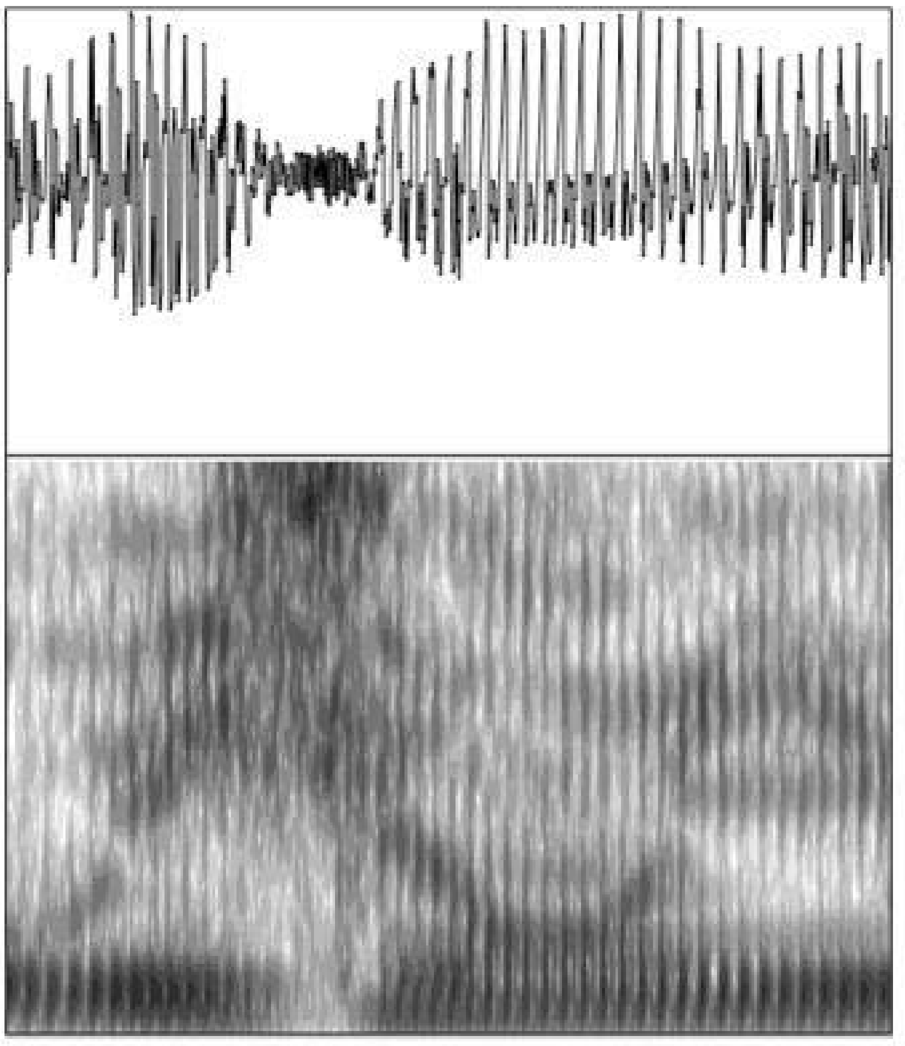


Figure 79.7 Waveform and spectrogram of . . . *weekend were you*, from conversational speech

English corpus have at least one syllable deleted, as determined from phonetic transcriptions. Arai (1999) studies perception of spontaneous Japanese speech, using the fact that the Japanese mora writing system allows one to evaluate perceived mora (or syllable) count. Arai (1999) presented a stretch of conversational speech, with varying amounts of context. Listeners were simply asked to write down what they had heard. Out of context, a stimulus which would have five moras in careful speech was perceived as containing an average of slightly more than two, but with even a little context, the same stretch was reported as constituting an average of nearly five moras. This study concurs with others that listeners cannot recognize reduced words or sounds well out of context, yet do so quite successfully in context. It also provides direct evidence that listeners perceive fewer moras or syllables for the same reduced acoustic material out of context.

3.4 *Shrinkage of acoustic spaces*

Literature on reduced speech often focuses on changes from one segment to another or deletions (e.g. Greenberg 1999; Johnson 2004; Shattuck-Hufnagel and Veilleux 2007). This is partly because of the method of transcribing a corpus, then comparing the transcription to dictionary listings for the words (Warner, forthcoming). However, some types of reduction do not result in a different transcription, as when the overall vowel space shrinks (Koopmans-van Beinum 1980). Many of the vowel tokens in Koopmans-van Beinum's work would probably be perceived as the full vowel phonemes (e.g. /i u a/, etc.) and not as /ə/, yet the vowel space shrinks with each step away from careful speech. Ladefoged *et al.* (1976) are unable to find this effect in Southern California English, but theirs is a relatively small study.

Other acoustic spaces may also compress, without causing changes in the transcribed segments. Berry (2009) shows reduction of the tonal space in spontaneous Mandarin speech (also demonstrating how reduction applies to suprasegmentals). Furui and colleagues (e.g. Nakamura *et al.* 2007) quantify reduction in spontaneous Japanese through a measure of spectral difference among phonemes, showing lesser spectral difference in reduced speech. This method is particularly useful for its ability to quantify reduction across all segment types, regardless of either underlying or surface manner of articulation.

4 The driving force behind reduction

What motivates reduction? Several motivations have appeared in the past literature, although some have primarily been proposed as explanations of other phenomena, and simply extend easily to reduction. The most obvious potential explanation may be "ease of articulation." In the 1990s, a trend developed that proposed speakers' desire for ease of articulation as an opposing principle to their wish for listeners to understand them, with these two wishes sometimes described as ranked constraints in Optimality Theory (reviewed and criticized by Hale and Reiss (2000); see also CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). While these constraints were not developed to describe spontaneous speech, their applicability is obvious. This would, however, be tantamount to saying that reduced speech is sloppy speech. Furthermore, what constitutes "ease" in articulation is neither clearly defined nor readily measurable.

Target undershoot is perhaps a more measurable version. Reduction in the size of gestures provides a convenient way to describe many reductions, such as expected stops being realized as approximants. Combining target undershoot and gestural overlap (discussed above), one could describe lowered syllable counts, as in *beret* realized as *bray* (Browman and Goldstein 1989). One would have to use gestural overlap and target undershoot considerably more heavily to model extreme reductions as in Figures 79.6c and 79.7, but this would be readily possible. However, the idea of reduced size and greater overlap of gestures does not provide a *motivation* for reduction, just a description. Task dynamic modeling, with its interaction among stiffness, targets and time, might do so, but gestures may be a better description of reduced speech than they are an explanation of what causes it. As far as the author is aware, the exact relationship of speech rate, stiffness, speech reduction, and speaker's choice of speech style remains for future research.

"Ease of articulation" and gestures focus on the speaker. One can alternatively focus on the listener. The OT account mentioned above suggests that listeners want clearer speech, and the perceptual literature does confirm that listeners find clearer speech easier to process (Ranbom and Connine 2007; Tucker 2007). However, listeners may not uniformly "want" unreduced speech. Speakers might accommodate the listener's specific informational needs in choosing when to reduce. This could be viewed as an overall motivation: overlap the segments (or gestures, or perceptual cues) during low-information portions of the signal, then use clearer speech for high-information portions, maximizing the speed and efficiency with which information is conveyed. Some results, such as Ernestus and colleagues' work on high-frequency words and suffixes (e.g. Dutch *eigenlijk* 'actually', mentioned above), support this view. In some items, they find greater reduction for higher-frequency words (CHAPTER 90: FREQUENCY EFFECTS), for words that have already occurred in the conversation, and for words that are more predictable from surrounding words (Pluymaekers *et al.* 2005a, 2005b). All of these support the idea that speakers reduce where the information will not be too important, or difficult to retrieve, for the listener. Kuperman *et al.* (2007) find an opposite effect of predictability for a specific set of data, though. Greenberg (1999) finds that low-frequency words tend to be pronounced canonically, regardless of speech rate, but that high-frequency words show more reduction with faster speech rate. However, Bard *et al.* (2001) find quite surprising results by cleverly manipulating what the listener actually knows and what the speaker knows, suggesting that speakers decide how much to reduce based more on what the speaker him/herself knows than on what the listener knows. These results together suggest that information structure certainly has an effect, but not a straightforward one, on speech reduction. It seems safe to conclude that reduction is only partially motivated by how information can be conveyed efficiently.

Furthermore, reduction itself probably conveys information. Bradlow (2002) argues that CV co-articulation conveys information to the listener and is part of the speaker's intentional strategy, rather than being simply a necessary consequence of inability to move articulators instantaneously from one place to another. Ogasawara (2007) and Tucker (2007) both find that listeners make use of reduction or speech rate in the context to help them decide on the acoustic criteria for later sounds in the speech stream. Putting these separate ideas together with the commonsense notion that speakers are more likely to reduce when talking with a close friend than with a prospective employer, for example, it seems likely that part of the motivation for reduction is to convey something about the speech style to the listener, rather than to eliminate unnecessary gestures for the speaker.

Overall, there is no definitive explanation for what drives reduction. It seems very likely that articulatory factors (e.g. task dynamic stiffness, articulator movement rate), information structure (greater reduction where information is less important), and intentional use of reduction as a feature that conveys information in itself all contribute to how much reduction a given utterance contains. The possible interactions of these factors are too complex to be disentangled in a single experiment.

5 Representational consequences of reduction

5.1 Formal phonology

As discussed in the section on historical perspectives above, formal phonology (whether OT or rule-based) has largely ignored reduction. Assuming that language is divided into competence *vs.* performance, all reduction may be a performance phenomenon, outside the grammar. However, one must know how to reduce in order to be a native speaker of a language, particularly if anything about reduction is language-specific. Barry and Andreeva (2001) show cross-linguistic similarities in reduction types, but there has been little quantitative comparison of reduction across languages, and stress *vs.* syllable *vs.* mora rhythm at least are likely to cause language-specific differences in reduction. Keune *et al.* (2005) find more reduction in Belgian Dutch (Flemish) than Netherlands Dutch. To the extent that how to reduce is language-specific, it may have to be part of the grammar, although at the phonetic level (Pierrehumbert 1994).

However, formal phonology rarely reaches a detailed enough surface level to reflect reductions, nor does it attempt to generate as many variants for each word as reduction produces. Formal phonology usually takes written broad transcriptions of single-word careful pronunciations as the data to be accounted for. If a particular token of *we were* as in Figure 79.2 sounds like *were* out of context but like *we were* in context, a formal phonological analysis is likely to lose that information before the analysis ever starts. Formal phonology examines an abstract version of how a word might be pronounced (carefully), not how it was pronounced on a particular occasion.

However, the 1990s saw a huge increase in the number of formal phonological models that integrate gradient, quantitative phenomena. Pierrehumbert (1994) lays out the reasons for addressing quantitative variability within the phonological competence. Flemming's (1995) work includes constraints that specify by how many Hertz two vowels' formants must differ. Warner (2002) addresses (but argues against) the possibility of modeling low-level phonetically variable acoustic events such as epenthetic stops (e.g. [k] in *young*[k]*ster*) within Optimality Theory. Nagy and Reynolds (1997) suggest ranking constraints variably to obtain multiple possible outcomes (see also CHAPTER 92: VARIABILITY). Boersma's version of OT (Boersma and Hayes 2001) adds noise to constraint rankings, so that how an underlying form is pronounced on a particular occasion can vary in a specific distribution. None of these works focuses on reduced speech, but the overall development of including gradience provides a mechanism for modeling the variability of reduction in formal phonology.

Using Boersma and Hayes's (2001) approach, one could easily model deletions (e.g. [wɚ] for *we were*; (Warner *et al.* 2009a) by ranking the deletion-preventing constraint MAX only slightly higher than the markedness constraints that work against realization of various segments or sequences (e.g. *LAB, *COR, *NÇ, NOCODA, etc. (Kager 1999). Markedness constraints are unviolated if the relevant segments are deleted, and MAX is unviolated if the underlying segments are maintained. Thus, if MAX were ranked just slightly above a collection of many markedness constraints, the random noise which Boersma and Hayes's system adds to rankings would sometimes place MAX lower than some of the markedness constraints, making a form with deletions optimal. Since random noise is added to the constraint rankings each time a speaker produces a form, MAX would be demoted beneath a different collection of markedness constraints on different productions, modeling the variability of which deletions occur in a given token. Furthermore, by varying how much random noise is added to constraint rankings, one could model varying amounts of deletion. Perhaps spontaneous, casual speech involves adding more noise to constraint rankings at evaluation time than careful speech does, supplying a direct way to model a speaker's choice of speech styles and degree of reduction. By ranking the IDENT constraints that prevent changes to manner of articulation and voicing appropriately relative to the markedness constraints, one could perhaps model reduction of stops to approximants and other such changes.

If reduced speech becomes part of "the business" of phonology, it could be modeled in OT in the way tentatively suggested here, or in some other way, or it could be modeled in a rule-based formalism. It would require far more detailed surface forms than are typically encoded, variable rules or variably ranked constraints (Boersma and Hayes 2001), and perhaps acoustically detailed rules or constraints (e.g. "overlap the gestures by at least *x* msecs or percent").

It is an open question in the field of psycholinguistics and spoken word recognition whether words have a single underlying representation in the lexicon or a wide variety of lexical listings (discussed below). Greenberg (1999) finds that the word *that* occurs with 117 distinct pronunciations in his spontaneous speech corpus, of which the most common is [ðæ], accounting for just 11 percent of all tokens. Is there just one abstract underlying representation stored for *that*, from which 117 different forms are derived? (See CHAPTER 1: UNDERLYING REPRESENTATIONS for more discussion.) How to write phonological rules or constraints that can produce the wide array of surface forms from a single underlying representation would be a mind-boggling problem. How can one change /ðæt/ into any of 117 forms, without allowing all words to become simply [ə]? One might need rules or constraints that effectively map any vowel onto any central vowel quality, any consonant onto an approximant, and any segment onto null. While it is not true that anything is possible in reduced speech, so many different things *are* possible that any formal phonological system might severely overgenerate, if set up to generate attested reduced forms.

5.2 Articulatory Phonology

Articulatory Phonology (Browman and Goldstein 1989 and many other publications) is better equipped to handle reduction and variability than many theories, as discussed above. Articulatory Phonology allows for reduction in the size or temporal span of gestures, and allows overlap, readily accommodating most or perhaps all reductions, as discussed above. It does require that all gestures present in the underlying representation (which consists of a gestural score), and only those gestures, be present in the surface form. This means that it is not literally possible for a gesture to be deleted. However, the theory does not prevent reducing a gesture until it is not measurably different from deletion. There is ample evidence that speakers often make articulatory gestures even if there is no audible acoustic consequence (e.g. the *seve*[m] *plus* and *perfe*[k] *memory* examples discussed above, with some tongue tip gesture maintained). This supports the idea that speakers would reduce the size of gestures in reduced speech whether listeners hear them or not. Heavily overlapping gestures might remove most sudden acoustic changes in the word, leaving one long vocalic stretch with minor acoustic variation throughout and no clear segments, as in portions of Figures 79.3a, 79.4b, and 79.7. The fact that Articulatory Phonology incorporates time as a continuous scale (gesture duration, rather than simply linear order of segments or features) is an important factor in its success with speech variability.

Articulatory Phonology may seem a perfect theory for describing reduced speech. However, this may be because it is actually more adept at describing phonetic implementations than at describing most abstract phonological alternations. With its prohibition on adding or removing gestures from the underlying representation, it is not meant to account for abstract morphophonemic alternations. (As one example, Navajo has certain suffix combinations in which all segmental material of one suffix deletes, but a high tone is added (Young and Morgan 1987). This is clearly not a matter of altering underlying gestures.) Instead, the theory shows its strength in areas that traditional formal phonology might have relegated to phonetic implementation, such as casual or fast speech reduction. It is clear, though, that inclusion of speech reduction would have no *representational* consequences for Articulatory Phonology: the use of gestural scores as representations works extremely well for reduced speech. Articulatory Phonology is simply based in the gradience and variability of real speech, which includes reduced speech, whereas other phonological theories are not.

5.3 *Abstractionist theories of spoken word recognition*

Moving beyond models of phonology, we can ask what representational consequences reduced speech has for spoken word recognition models. Greenberg (1999) finds that in data from the Switchboard corpus, the word *and* has 87 distinct pronunciations, including [æn, ɛn, ən, ænt, an, əm, i]. He also finds that the most common pronunciation of *them* is [əm]. How does a listener get from the possible surface forms back to the lexical entry *and* without creating massive confusion with *ant, on, them, hand, a, I'm,* etc.? The problem is more extreme for high-frequency function words, but is not limited to them. Models of spoken word recognition such as TRACE, SHORTLIST, etc. traditionally assume that each word has a single underlying form (e.g. Norris *et al.* 2000), although listing multiple forms is sometimes adopted (Spinelli *et al.* 2003). Listeners would recognize [ɨ] as [ænd], despite their dissimilarity, through a combination of finding the closest segmental match in the lexicon and weighting high-frequency words. For example, listeners might recognize [ɪs] as *this* despite the poor segmental match, because *this* has higher frequency than *hiss*. However, many high-frequency words reduce to similar forms: *this, just,* and *is* can probably all be realized as something like [ɨs], and realizations of *and* overlapping with *ant, on, them,* and *a* above also demonstrate this. Thus best match plus frequency may not solve the problem, and neither would multiple lexical listings.

For somewhat less ambiguous forms such as [wiçə̃] *weekend* (Figure 79.7), listeners could try to apply a rule-like conversion to get from the surface string back to the single lexical representation. However, the problem would be just as for formal phonology in reverse: rules would have to allow for insertion and alteration of almost any segments. Since reduction creates such varied forms, it might be impossible to systematically derive a single invariant underlying form by working backwards from them.

Another approach is to list every possible form of each word in the lexicon, or at least several forms. Thus, the word *that* might have at least 117 underlying representations, or at least some substantial number from which the rest of the possible surface forms can be derived. The exact number is dependent on the narrowness of the transcription system, but the effect on the lexicon is the same: the number of forms would multiply greatly. This solution is a drastic departure from the traditional view of what a lexicon contains, and assumes that speakers and listeners do very little abstraction.

Taking the logic of multiple underlying listings further, the forms listed need not be limited to those that receive differing transcriptions in a phonetics lab. Why should *that* be limited to the 117 forms Greenberg (1999) found for it? Perhaps it should have a separate listing for every acoustically distinct pattern that can be realized for the word. This leads us to whole-word-based exemplar models of spoken word recognition (Goldinger 1998). In such a theory, every incoming speech token is measured on various acoustic characteristics, and placed into a "covering map" on all the relevant acoustic dimensions (Johnson 1997), with this information saved about each token a listener hears.

Exemplar models were not developed to account for reduced speech. However, they already use an unusually detailed version of lexical representation, which can be viewed as the word category plus all the acoustic information about exemplars of that word heard in the past. An exemplar model would also save information about the speech style in which the listener heard a particular token. For example, if a listener hears a highly reduced token of *and* in fast, casual speech that was realized just as [ɨ], and successfully recognizes it, the acoustic properties of this token and the fact that it was a token of *and* will be stored, and the fact that it occurred in casual, fast speech would also be stored. This might help the listener to avoid recognizing the acoustic pattern [ɨ] as *and* in slow formal speech. It remains to be explicitly tested though whether an exemplar model of spoken word recognition would do any better than other models at identifying reduced speech tokens. Whether token-specific information is saved or not, identifying [ɨs] as the word *this vs. just*, for example, would present a challenge to any model. If past exemplars of these two words happen to fall into acoustically somewhat distinct clusters, an exemplar model might succeed (but other models would as well). However, if information about long-term speech rate across the utterance and syntactic and semantic context are more important than acoustic differences within the word, then an exemplar model might have no advantage.

To sum up the issue of whether reduced speech affects our understanding of what constitutes a lexical representation, this depends on the degree of phonetic detail included in a theory's representations. Theories that include considerable detail in lexical representations or memory (e.g. Articulatory Phonology, exemplar models) can readily accommodate reduced speech without a change to what constitutes a representation, although this does not guarantee that these models would succeed in generating or recognizing the correct forms. Theories with exclusively abstract lexical representations may require moderate or large numbers of separate underlying representations for each word in order to accommodate the variety of reduced surface forms that occur in normal speech.

6 Conclusions

At the outset, I suggested that the controversy over reduced speech is not one of how phonological theories should handle a particular phenomenon, but rather one of *whether* phonetic or phonological theories even should attempt to handle the phenomenon. It is clear that reduction exists, is extremely common, and is not peripheral to the system. But is reduced speech of any interest to phonetics or phonology? We return now to the possible answers to this question.

Within applied work, there is a reason to include reduced speech: if we would like speech technology systems (speech synthesis or automatic speech recognition (ASR)) to make use of what we know about language, we should know as much as possible about the everyday connected speech which humans use. While speakers may speak somewhat clearly to ASR systems in some applications, ASR systems have to recognize a great deal of reduction. The acoustic findings on reduction show that even relatively careful speech contains quite a number of reductions (Warner and Tucker 2007), so ASR systems are unlikely to escape the task of recognizing reduced forms. Another application for findings about reduction is language teaching. Students acquiring a second language in a classroom usually hear careful speech. Many of us have had the experience of arriving in the country of our L2, only to find that we cannot understand much at all. While that problem has many sources, pronunciation variability from casual speech reduction is likely to be one of them. (See also Shockey 2003 on the importance of reduced speech for both of these applications.)

The second reason for considering reduction to be of interest to phonetics and phonology is theoretical. While no theory needs to account for every phenomenon, it seems reasonable for theories of phonology to be able to reach detailed surface forms that actually occur in corpora, unless the theory is intended to stay entirely at a level of abstraction. If the general purpose of phonology is to relate what speakers know about words to how they pronounce them, one might want the theory to be equipped to discuss and represent attested forms; if obtained from a corpus, these forms will include reduction. Some phonological theories can represent reduced forms, at least as transcriptions and perhaps with acoustic values. The modifications to these theories that allow them to model reduction usually were not introduced for that purpose, but the necessary formal mechanisms are now in place.

For a theory of phonetics to rule out reduced speech as being outside its area of interest would be surprising. Phonetic theories clearly "are responsible for" speech as it is produced and perceived. Thus, theories of gestural coordination, of segment perception, of speaker normalization, of phoneme distinctiveness, etc. should be adaptable for reduced speech. Fortunately, many phonetic theories are less sensitive to representational issues than formal phonological theories, and Articulatory Phonology already provides a thoroughly implemented theory that can represent and model reduction, as discussed above.

Models of segment perception (whether exemplar or more abstract; cf. Smits *et al.* 2006) can also potentially accommodate the detail of reduced segments, although these models have not generally been tested on reductions. One might argue that reduced speech should be considered during the development of all phonetic theories, even if reduction is not the primary interest, to see whether the theory would adapt to the speech that speakers and listeners use daily. Even if connected speech data would be too variable to test a theory on, it should at least be able to apply to more natural speech. For example, the stimuli used by Smits *et al.* (2006) are synthesized non-speech noises, because the exact controlled distribution of synthesized acoustic characteristics allows them to contrast several models' predictions. However, one could test each model on how it distinguishes the reduced realizations of flap in Figures 79.4b and 79.6 from /ð/ or from a vowel–vowel sequence, and the mechanism behind each theory could, in principle, work for reductions.

When one spends a lot of time looking at reduced, spontaneous speech, the difference between that and careful speech can seem so pervasive that one begins to wonder why there is so much research on abstract, careful forms. The exact details of, for example, locus equations (phonetic theory, involving the relationship between formant frequencies near a consonant and at vowel mid-point) or sonority sequencing (phonological theory) might be obscured or obliterated by the deletions, overlaps, mergers, and alterations of reduction. How can one determine a reliable locus equation for a place of articulation when the vowels can shift toward schwa, be deleted, or merge into a neighboring vowel, and the consonant might be deleted, or might surface with unexpected acoustic characteristics? How can one determine the sonority sequencing requirements when the number of syllables is unstable (two on the surface, five underlyingly, as in Arai 1999, or two *vs.* four in Figure 79.1), and manner and voicing of each consonant varies (e.g. Figure 79.2, where everything has become nasals and/or approximants)? The fields of phonetics and phonology have both invested considerable effort into working out the details of theories based on careful speech forms, with an assumption that those forms are representative. Spontaneous speech might make these detailed theories of careful speech seem pointless. Instead of asking whether theories should accommodate spontaneous speech reduction, one might ask whether theories should accommodate the forms of careful speech.

However, it is clear that native speakers' judgments of syllable structure are tapping into some real property of language, and that experimental findings on locus equations are as well. The same could be said of other phenomena in phonetics and phonology: spontaneous speech obscures the phenomenon, yet the phenomenon is clearly a real part of language. Also, it is easy to focus on the extreme reductions in spontaneous speech, but even casual conversation contains clearly articulated focused words as well. Furthermore, when one does not yet know much about a topic the field is starting to explore, it is certainly advisable to work with speech that is as controlled as possible. Phonetic and phonological studies that consider exclusively careful speech patterns do tell us about a real type of human language. However, what they tell us about is probably not the most common form of human language our auditory processing systems encounter in daily life, inside or outside the classroom, the lab, or our homes. Even when hearing professional newscasting, we are likely to encounter far more reduction than most phonetics experiments or phonological data consider. There are arguments for both perspectives: that reduced speech is a specific facet of phonetic implementation irrelevant to formal phonology and to most topics of phonetics, and also that it is important to test and model reduced speech in phonological and phonetic theories. What is clear, though, is that reduced speech is a normal part of our daily experience as speakers and hearers, not a rare or marginal phenomenon.

REFERENCES

Arai, Takayuki. 1999. A case study of spontaneous speech in Japanese. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee Bailey (eds.) *Proceedings of the 14th International Congress of Phonetic Sciences*, 615–618. Berkeley: Department of Linguistics, University of California, Berkeley.

Bard, Ellen Gurman, Catherine Sotillo, M. Louise Kelly & Matthew P. Aylett. 2001. Taking the hit: Leaving some lexical competition to be resolved post-lexically. *Language and Cognitive Processes* 16. 731–737.

Barry, William & Bistra Andreeva. 2001. Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association* 31. 51–66.

Berry, Jeff. 2009. Tone space reduction in Mandarin Chinese. Unpublished ms., University of Arizona.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45–86.

Bolozky, Shmuel. 1977. Fast speech as a function of tempo in Natural Generative Phonology. *Journal of Linguistics* 13. 217–238.

Bradlow, Ann R. 2002. Confluent talker- and listener-related forces in clear speech production. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology* 7, 241–273. Berlin & New York: Mouton de Gruyter.

Bradlow, Ann R. & Tessa Bent. 2002. The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America* 112. 272–284.

Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6. 201–251.

Cheng, Chierh & Yi Xu. 2008. When and how disyllables are contracted into monosyllables in Taiwan Mandarin? *Journal of the Acoustical Society of America* 123. 3864.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Connine, Cynthia M. 2004. It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review* 11. 1084–1089.

Cutler, Anne. 1998. The recognition of spoken words with variable representation. In *Proceedings of the ESCA Workshop on Sound Patterns of Spontaneous Speech*, 83–92. Aix-en-Provence: ESCA.

Dalby, Johnathan. 1986. Phonetic structure of fast speech in American English. Ph.D. dissertation, Indiana University.

Dressler, Wolfgang U. 1975. Methodisches zu Allegroregeln. In Wolfgang U. Dressler & F. V. Mareš (eds.) *Phonologica 1972*, 219–234. Munich: Fink.

Engstrand, Olle & Diana Krull. 2001. Simplification of phonotactic structures in unscripted Swedish. *Journal of the International Phonetic Association* 31. 41–50.

Ernestus, Mirjam. 2000. Voice assimilation and segment reduction in casual Dutch: A corpus-based study of the phonology–phonetics interface. Ph.D. dissertation, University of Utrecht.

Ernestus, Mirjam, R. Harald Baayen & Rob Schreuder. 2002. The recognition of reduced word forms. *Brain and Language* 81. 162–173.

Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105. 251–279.

Greenberg, Steven. 1999. Speaking in shorthand: A syllable-centric perspective for understanding pronunciation variation. *Speech Communication* 29. 159–176.

Hale, Mark & Charles Reiss. 2000. "Substance abuse" and "dysfunctionalism": Current trends in phonology. *Linguistic Inquiry* 31. 157–169.

Hicks Kennard, Catherine. 2006. Gender and command: A sociophonetic analysis of female and male drill instructors in the United States Marine Corps. Ph.D. dissertation, University of Arizona.

Hockett, Charles F. 1955. *A manual of phonology*. Baltimore: Waverly Press.

Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson & John W. Mullennix (eds.) *Talker variability in speech processing*, 145–165. San Diego: Academic Press.

Johnson, Keith. 2004. Massive reduction in conversational American English. In Kyoko Yoneyama & Kikuo Maekawa (eds.) *Spontaneous speech: Data and analysis*. 29–54. Tokyo: National Institute for Japanese Language.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Keune, Karen, Mirjam Ernestus, Roeland van Hout & R. Harald Baayen. 2005. Social, geographical, and register variation in Dutch: From written "mogelijk" to spoken "mok." *Corpus Linguistics and Linguistic Theory* 2. 183–223.

Kohler, Klaus J. 2001. Articulatory dynamics of vowels and consonants in speech communication. *Journal of the International Phonetic Association* 31. 1–16.

Koopmans-van Beinum, Florian J. 1980. *Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions*. Amsterdam: Academische Pers.

Kryter, Karl D. 1946. The effect of plugging the ears on the intelligibility of speech in noise. *Journal of the Acoustical Society of America* 18. 249.

Kuperman, Victor, Mark Pluymaekers, Mirjam Ernestus & R. Harald Baayen. 2007. Morphological predictability and acoustic duration of interfixes in Dutch compounds. *Journal of the Acoustical Society of America* 121. 2261–2271.

Ladefoged, Peter. 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA & Oxford: Wiley-Blackwell.

Ladefoged, Peter, Iris Kameny & William Brackenridge. 1976. Acoustic effects of style of speech. *Journal of the Acoustical Society of America* 59. 228.

Lennes, Mietta, Nina Alarotu & Martti Vainio. 2001. Is the phonetic quality of unaccented words unpredictable? An example from spontaneous Finnish. *Journal of the International Phonetic Association* 31. 127–138.

Löfqvist, Anders & Vincent L. Gracco. 1999. Interarticulator programming in VCV sequences: Lip and tongue movements. *Journal of the Acoustical Society of America* 105. 1864–1876.

Maekawa, Kikuo & Hideaki Kikuchi. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report. In Jeroen van de Weijer, Kensuke Nanjo & Tetsuo Nishihara (eds.) *Japanese voicing*, 205–228. Berlin & New York: Mouton de Gruyter.

McLennan, Conor T., Paul A. Luce & Jan Charles-Luce. 2003. Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29. 539–553.

McLennan, Conor T., Paul A. Luce & Jan Charles-Luce. 2005. Representation of lexical form: Evidence from studies of sublexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance* 31. 1308–1314.

Mehta, Gita & Anne Cutler. 1988. Detection of target phonemes in spontaneous and read speech. *Language and Speech* 31. 135–156.

Mitterer, Holger & Mirjam Ernestus. 2006. Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics* 34. 73–103.

Muthusamy, Yeshwant K., Ronald A. Cole & Beatrice T. Oshika. 1992. The OGI multi-language telephone speech corpus. In *Proceedings of the 1992 International Conference on Spoken Language Processing*. 895–898.

Nagy, Naomi & Bill Reynolds. 1997. Optimality Theory and variable word-final deletion in Faetar. *Language Variation and Change* 9. 37–55.

Nakamura, Masanobu, Koji Iwano & Sadaoki Furui. 2007. The effect of spectral space reduction in spontaneous speech on recognition performances. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, vol. 4, 473-476.

Nicolaidis, Katerina. 2001. An electropalatographic study of Greek spontaneous speech. *Journal of the International Phonetic Association* 31. 67–85.

Norris, Dennis, James M. McQueen & Anne Cutler. 2000. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23. 299–370.

Ogasawara, Naomi. 2007. Processing of speech variability: Vowel reduction in Japanese. Ph.D. dissertation, University of Arizona.

Pierrehumbert, Janet B. 1994. Knowledge of variation. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 30. 232–256.

Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005a. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118. 2561–2569.

Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005b. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62. 146–159.

Pluymaekers, Mark, Mirjam Ernestus, R. Harald Baayen & Geert Booij. 2010. Morphological effects on fine phonetic detail: The case of Dutch *-igheid*. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.) *Laboratory phonology* 10, 511–532. Berlin & New York: Mouton de Gruyter.

Ranbom, Larissa J. & Cynthia M. Connine. 2007. Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language* 57. 273–298.

Raymond, William D., Robin Dautricourt & Elizabeth Hume. 2006. Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change* 18. 55–97.

Richter, Elise. 1930. Beobachtungen über Anglitt und Abglitt an Sprachkurven und umgekehrt laufenden Phonogrammplatten. In Paul Menzerath (ed.) *Berichte über die I. Tagung der Internationalen Gesellschaft für experimentelle Phonetik*, 87–90. Bonn: Scheur.

Shattuck-Hufnagel, Stephanie & Nanette M. Veilleux. 2007. Robustness of acoustic landmarks in spontaneously-spoken American English. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 925–928. Saarbrücken: Saarland University.

Shockey, Linda. 2003. *Sound patterns of spoken English*. Cambridge, MA & Oxford: Blackwell.

Smiljanić, Rajka & Ann R. Bradlow. 2005. Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America* 118. 1677–1688.

Smiljanić, Rajka & Ann R. Bradlow. 2008. Clear speech intelligibility and accentedness ratings for native and non-native talkers and listeners. *Journal of the Acoustical Society of America* 123. 3883.

Smits, Roel, Joan Sereno & Allard Jongman. 2006. Categorization of sounds. *Journal of Experimental Psychology: Human Perception and Performance* 32. 733–754.

Spinelli, Elsa, James M. McQueen & Anne Cutler. 2003. Processing resyllabified words in French. *Journal of Memory and Language* 48. 233–254.

Strik, Helmer, Micha Hulsbosch & Catia Cucchiarini. 2010. Analyzing and identifying multiword expressions in spoken language. *Language Resources and Evaluation* 44. 41–58.

Tseng, Shu-Chuan. 2005. Syllable contractions in a Mandarin conversational dialogue corpus. *International Journal of Corpus Linguistics* 10. 63–83.

Tucker, Benjamin V. 2007. Spoken word recognition of the reduced American English flap. Ph.D. dissertation, University of Arizona.

Warner, Natasha. 2002. The phonology of epenthetic stops: Implications for the phonetics–phonology interface in Optimality Theory. *Linguistics* 40. 1–27.

Warner, Natasha. Forthcoming. Methods for studying spontaneous speech. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.) *Handbook of laboratory phonology*. Oxford: Oxford University Press.

Warner, Natasha & Benjamin V. Tucker. 2007. Categorical and gradient variability in intervocalic stops. Paper presented at the 81st Annual Meeting of the Linguistic Society of America, Anaheim.

Warner, Natasha, Dan Brenner, Anna Woods, Benjamin V. Tucker & Mirjam Ernestus. 2009a. Were we or are we? Perception of reduced function words in spontaneous conversations. *Journal of the Acoustical Society of America* 125. 2655.

Warner, Natasha, Amy Fountain & Benjamin V. Tucker. 2009b. Cues to perception of reduced flaps. *Journal of the Acoustical Society of America* 125. 3317–3327.

Young, Robert & William Morgan. 1987. *The Navajo language: A grammar and dictionary.* Albuquerque: University of New Mexico Press.

Zwicky, Arnold M. 1972. On casual speech. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 8. 607–615.

80 Mergers and Neutralization

ALAN C. L. YU

1 Introduction

The notions of mergers and neutralization presuppose the concept of contrast. Two sounds are phonologically contrastive if they are in opposition with each other, i.e. if they are capable of differentiating the lexical meanings of two words in a particular language. The plosives [p] and [p^h], for example, are in opposition in Cantonese (Yue Chinese) (e.g. [paː˥] 'father' *vs.* [p^{h}aː˥] 'to lay down') but [b] and [p] are not. Contrast is not restricted to pairs of segments; classes of segments contrast as well. The aspiration opposition between [p] and [p^h] finds analogs in other pairs of segments ([t] ~ [t^h], [k] ~ [k^h], [k^w] ~ [k^{wh}]). When a phonological opposition is suspended, neutralization or merger obtains. For example, Cantonese has no aspiration opposition between plain and aspirated plosives in syllable-final position; all syllable-final plosives are voiceless and unreleased (e.g. [t^{h}aːp̚˧] 'pagoda', [paːt̚˧] 'eight', [kɔːk̚˧] 'corner, horn').

The terms *merger* and *neutralization* are often employed in complementary contexts; *merger* often characterizes a diachronic and *neutralization* a synchronic collapse of contrast. The diachronic–synchronic divide between merger and neutralization is more apparent than real, however; the two notions are the two faces of the same coin. The notion of merger is often applied in the context where a contrast reduction leaves no trace of the contrast in the synchronic system; a context-free contrast reduction is the clearest example of this. Neutralization applies to context-dependent contrast reduction; traces of a contrast remain in some contexts, but not in others. Certain varieties of English, for example, merge the voiceless labial-velar fricative /ʍ/ with its voiced counterpart /w/ (Minkova 2004). Thus the words *whine* and *wine* are homophonous; no remnant of this /ʍ/ ~ /w/ contrast is evidenced in the grammar of speakers of these dialects. In certain dialects of Cantonese (most prevalently in Guangzhou, Hong Kong, and Macao; Bauer and Benedict 1997), the distinction between plain and labial velars is not maintained before the back rounded vowel /ɔ/. The collapse of the plain *vs.* labial velars distinction is referred to as a matter of neutralization because the contrast remains before vowels that are not /ɔ/ (e.g. [kɐn˥] 'tight' *vs.* [k^{w}ɐn˥] 'boil'). These instances of contrast reduction in English and Cantonese transpire diachronically, but one results in a merger (i.e. the /ʍ/ ~ /w/ merger) and the other in neutralization (i.e. /k$^{(h)}$/ ~ /k$^{w(h)}$/ neutralization). In this chapter, I shall collectively refer to mergers and neutralization in terms of *contrast reduction*. I shall further assume that the term *neutralization* refers to contrast reduction that results in alternation, while the term *merger* will refer to any reduction of contrast, both synchronically and diachronically. Thus, in the case of the /k$^{(h)}$/ ~ /k$^{w(h)}$/ contrast in Cantonese, /k$^{(h)}$/ and /k$^{w(h)}$/ *merge* before /ɔ/ diachronically. The outcome of this *merger* is the *neutralization* of /k$^{(h)}$/ and /k$^{w(h)}$/ before /ɔ/.

This chapter begins with a review of the range of contrast reduction (§2). §3 surveys several theories that attempt to explain the sources of contrast reduction. §4 concludes with a discussion of the challenges to a purely phonological conception of contrast reduction.

2 Typology of contrast reduction

Contrast reduction manifests itself in three different ways: structure-preserving reduction, structure-building reduction, and free variation.[1] *Structure-preserving reduction* characterizes scenarios where two or more distinct sounds have, after the reduction, a form that is physically similar to that of one of the sounds appearing in the position of differentiation (e.g. /k$^{(h)}$/ ~ /k$^{w(h)}$/ neutralization; cf. Kiparsky 1985; CHAPTER 76: STRUCTURE PRESERVATION: THE RESILIENCE OF DISTINCTIVE INFORMATION). Formally, a reduction of contrast *m* is structure-preserving if and only if *m* turns two (or more) distinct sounds into only one of the two sounds, to the exclusion of the other. The merger of /ʍ/ and /w/ is structure-preserving, since the result of the merger leaves /w/ as the surviving sound. Regressive assimilation of voicing is another instance of structure-preserving contrast reduction. For example, in Dutch, the distinction between voiced and voiceless plosives is suspended preconsonantally (Ernestus and Baayen 2003). However, the result of neutralization differs depending on the nature of the following consonant. For example, before a voiced plosive, the /t/ ~ /d/ contrast in *verwijten* [vɛrʋɛitən] 'reproach-INF' and *verwijden* [vɛrʋɛidən] 'widen-INF' neutralizes toward /d/ (*verwijt bijna* [vɛrʋɛid bɛinaː] 'reproach almost' *vs. verwijd bijna* [vɛrʋɛid bɛinaː] 'widen almost'). However, before a nasal, neutralization is toward /t/ (*verwijt niet* [vɛrʋɛit nit] 'reproach not' *vs. verwijd niet* [vɛrʋɛit nit] 'widen not').

Contrast reduction is structure-building when the outcome of contrast reduction is a sound intermediate between the normal realization of the two phonemes. Final-consonant voicing neutralization in Cantonese is a case in point. Stops in syllable-final position are unreleased, and thus phonetically non-contrastive in terms of aspiration (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). Another celebrated case of structure-building reduction is flapping in English. The /t/ ~ /d/ contrast in English is suspended intervocalically where the coronal in question is immediately followed by an unstressed vowel (e.g. *heed* ['hiːd] *vs. heat* ['hiːt], but *ladder* ['læɾɚ] *vs. latter* ['læɾɚ]) (CHAPTER 113: FLAPPING IN AMERICAN ENGLISH).[2]

1 Unless noted otherwise, I shall abstract away from the issue of context-sensitivity in what follows.

2 The definition of neutralization adopted here differs from Kiparsky's (1976: 169) formulation of a neutralizing rule, which states that a rule of the form A → B / XC __ DY is neutralizing iff there are strings of the form CBD in the input to the rule. Certain structure-building neutralizing rules, such as flapping in English, are not considered neutralizing from Kiparsky's perspective, since the product of the rule is not phonemic in the language.

When contrast reduction leads to a form varying between two or more variants, this is referred to as *free variation* (see CHAPTER 92: VARIABILITY). For a large number of Cantonese speakers, syllable-initial [n] is in free variation with [l] (Bauer and Benedict 1997). Thus, words like [nej˩˧] 'you' and [naːn˨˩] 'difficult' are often pronounced with initial [l], thus merging with [lej˩˧] 'Li (surname)' and [laːn˨˩] 'orchid', respectively. The rate of [n] *vs.* [l] usage varies according to age and gender of the speaker, as well as the register of speaking (e.g. read speech *vs.* conversational speech).

2.1 *Positions of contrast reduction*

Contrast (CHAPTER 2: CONTRAST) is often restricted to certain positions within the word: the syllable peak (rather than the margin; CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE), the onset (rather than the coda; CHAPTER 55: ONSETS), the stem (rather than the affix; CHAPTER 104: ROOT–AFFIX ASYMMETRIES), the stressed syllable (CHAPTER 40: THE FOOT), or the edge of the morphological domains (CHAPTER 50: TONAL ALIGNMENT). Washo (Hokan), for example, only allows voiceless liquids and nasals in onset position (Jacobsen 1964). Isthmus Zapotec (Oto-manguean) contrasts glottalized and modal-voiced vowels, but only in stressed positions (Bueno-Holle 2009). Hausa (Chadic) has a five-vowel system (/i e a o u/), with a long–short distinction which is reliably distinguished only in final position (Steriade 1994). Ngalakan (Australian) has a five-vowel system (/i e a o u/), but mid vowels in Ngalakan are restricted to the edges of roots (Baker 1999: 72–73); if there is only one mid vowel in a root, it must appear in an edgemost syllable (i.e. initial /ceraṭa/ 'woman's ceremony' or final /curuwe-/ 'rush'). If there is more than one mid vowel, they must occur in contiguous syllables (/caworo/ 'patrilineal clan') or every vowel in the root must be a mid vowel (/koweleŋ?(-mi+)/ 'beckon to'). !Xóõ (Bushman) contrasts consonants with clicks and consonants without click accompaniment, but only in initial syllables (Traill 1985). In Etung (Bantu), falling and rising tones (HL, HꜜH, LH) are restricted to the final syllable of phonological words, but there is no restriction on the occurrence of level tones (Edmondson and Bendor-Samuel 1966). In Lushootseed (Central Salishan), glottalized consonants are only found in roots and lexical suffixes; grammatical suffixes never have glottalized consonants (Urbanczyk 1996: 46). Contrast restrictions might also differ across word types. For example, in a cross-linguistic survey of 32 languages having 26 consonants or more, Willerman (1994) found that pronouns made significantly less use of the palato-alveolar, retroflex, uvular, and pharyngeal places than other places of articulation and of fewer laterals, affricates, trills, clicks, ejectives, and aspirated segments (see also CHAPTER 102: CATEGORY-SPECIFIC EFFECTS for differences between nouns and verbs).

Loci of contrast reduction are not always characterizable in structural terms. Steriade (1994) observes that languages with a retroflexion contrast in the apicals (e.g. /t/ *vs.* /ʈ/) often neutralize the contrast in initial or postconsonantal positions, but allow the contrast in post-vocalic position (CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS). The position of retroflexion neutralization is difficult to capture in prosodic terms, since post-vocalic position can be either within or across a prosodic domain (e.g. the coda of a syllable and a syllable onset in intervocalic position). Obstruents in Lithuanian contrast in terms of voicing (Senn 1966; Steriade 1997). However, the voicing contrast is supported only before sonorants (*skobnis* 'table'; *bãdmetys* 'year of famine') and not elsewhere. Voicing is neutralized word-finally (*daũg* [dauk] 'much'; *kàd* [kat] 'that') and in pre-obstruent position (*dèg-ti* [kt] 'burn-INF', *míelas draũgas* [zd] 'dear friend').

2.2 Common triggers and targets of contrast reduction

Languages with contrast reduction often exhibit striking parallelism in the direction of merger and neutralization. Non-assimilatory neutralization of laryngeal contrasts in word-final and preconsonantal positions is often structure-preserving; the preserved segments are generally voiceless. Neutralization toward voiced or ejective is rare, if not non-existent.[3] Reduction of vocalic contrasts in unstressed positions is commonplace across the world's languages (CHAPTER 26: SCHWA). The vast majority of such reductions involve the neutralization of vowel nasalization, quantity, or height. Nasal and oral vowels, for example, are often only contrastive in stressed syllables (e.g. Copala Trique (Hollenbach 1977); Guaraní (Beckman 1998: 158)). Contrasts in vocalic quantity are frequently neutralized toward the short variant in unstressed syllables. Kolami, for example, only contrasts long and short vowels in initial syllables, which are always stressed (Emeneau 1961: 6–7). Quantity contrasts may also neutralize toward the long variant under certain circumstances. For example, a vowel following a consonant–glide sequence must be long (/ak-a/ 'ask!' *vs.* /kw-a:k-a/ 'to ask'; Myers and Hansen 2005: 318) in Rwanda (Bantu), which has a contrast in vowel length ([gusi:βa] 'to be absent' *vs.* [gusiβa] 'to erase'; Kimenyi 1979: 1). Reduction in vowel height in unstressed position often favors one of two outcomes: the unstressed vowel may become either [a] or [ə]. In Belarusian, for example, mid vowels /e o/ reduce to [a] (['noɣi] 'legs' *vs.* [na'ɣa] 'leg'; ['reki] 'rivers' *vs.* [ra'ka] 'river'; Crosswhite 2004: 192); thus the five vowels found in stressed syllables, /i e a o u/, are reduced to three, [i a u], in the unstressed syllables. The seven-vowel system in Central Eastern Catalan (/i e ɛ a ɔ o u/) is only evident in stressed syllables; in unstressed syllables, only three vowel qualities, [i ə u], are allowed; underlying /e ɛ a/ become [ə], while /u o ɔ/ become [u], as shown in (1). Vocalic contrast reductions along other featural dimensions are rare and are often secondary to height neutralization in the same system (Barnes 2002).

(1) *Central Eastern Catalan* (Barnes 2002: 37)

'riw	'river'	ri'wɛt	'river (DIM)'
'new	'snow'	nə'wɛtə	'snow (DIM)'
'mɛl	'honey'	mə'lɛtə	'honey (DIM)'
'palə	'shovel'	pə'ɛtə	'shovel (DIM)'
'rɔðə	'wheel'	ru'ðetə	'wheel (DIM)'
'mɔnə	'monkey (FEM)'	mu'nɛtə	'monkey (FEM DIM)'
'kurə	'cure'	ku'rɛtə	'cure (DIM)'

The targets of assimilatory neutralization show cross-linguistic similarities as well (Cho 1990; Ohala 1990; Jun 1995; Steriade 2001; de Lacy 2002, 2006). For

[3] Yu (2004) reports a case of final neutralization toward the voiced series in Lezgian (North Caucasian); the neutralization is restricted only to monosyllabic nouns, however. The default direction of neutralization in final position is toward the voiceless aspirated series.

example, obstruents are often voiced after nasals (Pater 1999). Nasals in turn frequently assimilate to the place of articulation of the following consonant, as illustrated by the examples from Yoruba (Niger-Congo) in (2).

(2) *Yoruba nasal assimilation* (Pulleyblank 1995: 5)

a.	bá	ḿbá	'overtake'
	fɔ́	ɱ́fɔ́	'break'
b.	tà	ńtà	'sell'
	sū	ńsū	'sleep'
c.	j●	ɲ̄j●	'dance'
	jɛ	ɲ̄jɛ	'eat'
d.	kɔ	ŋ̀kɔ	'write'
	wí	ŋ́wí	'say'
e.	gbɔ́	ŋ́mgbɔ́	'hear, understand'
	kpa	ŋ́mkpa	'kill'

Among obstruents, coronals are most susceptible to place assimilation. In Korean, for example, morpheme-final coronals assimilate to dorsals or labials (3a). Morpheme-final labials assimilate to dorsals (3b), but no assimilation is observed when the following consonant is coronal. Dorsals are inert; they assimilate neither to a following labial nor to a following coronal (3c) (CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION).[4]

(3) *Korean place assimilation* (Hume 2003: 7–8)[5]

a.	/mit+ko/	[mikk'o]		'believe and'
	/mith+pota/	[mipp'ota]		'more than the bottom'
b.	/ip+ko/	[ikk'o]		'wear and'
	/nop+ta/	[nopt'a]	*[nott'a]	'high'
c.	/nok+ta/	[nokt'a]	*[nott'a]	'melt'
	/kuk+pota/	[kukp'ota]	*[kupp'ota]	'more than soup'

3 Theories of contrast reduction

Early discussions of contrast reduction focused on how to characterize the outcome of context-specific contrast reduction. That is, how would a theory of phonemics capture the fact that the contrast between two or more sounds in some positions of a word or a syllable is not maintained in other positions (CHAPTER 11: THE PHONEME)? The main analytic puzzle neutralization presents to structuralist phonemics concerns the violation of the bi-uniqueness condition (i.e. of one-to-one mapping between allophones and phonemes). The Prague School resolves this indeterminacy by positing archiphonemes in contexts of neutralization (Trubetzkoy 1939); archiphonemes are units that represent the common features of phonemes whose contrastive property is neutralized in specific contexts. In Yoruba, for example, a preconsonantal nasal would be treated as an archiphoneme, N (e.g.

4 See Silverman (2010) for a thorough review of neutralizing processes in Korean.

5 C' indicates a tense consonant.

[ḿbá] /Ńbá/ 'overtake'). The archiphonemic treatment of neutralization anticipates the underspecification treatment of neutralized segments made possible by the reconceptualization of the phonemes as sets of distinctive features. In an underspecification model (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION), a preconsonantal nasal in Yoruba, for example, would be specified for the feature [+nasal], while the surface realization of this underspecified nasal would be specified contextually.

In addition to the issue of representation, theories of neutralization also attempt to explain the causes for neutralization. That is, why do cross-linguistic parallelisms abound in cases of contrast reduction? Two main approaches have been advanced: structure-based and cue-based. This section reviews how these two approaches conceptualize the problem of contrast reduction and what mechanisms account for the observed typological tendencies.

3.1 *Licensing and markedness*

Structure-based approaches maintain that certain prosodic or structural positions disfavor the maintenance of phonological contrasts. The phonological grammar may either prohibit a contrast in a given structural position in terms of a filter constraint (4) or impose a licensing condition which specifies how a phonological contrast must be configured in order to be realized in a given position within the word (5) (see also CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS).

(4) *Positional neutralization: Filter/negative version* (Steriade 1995: 120)

$^*\alpha F$ in x, where x is defined prosodically or morphologically.

(5) *Positional neutralization: licensing/positive version* (Steriade 1995: 121)[6]

αF must be licenced in x, where x is defined prosodically or morphologically.

Codas in Pali, for example, must be the first half of a geminate structure (6a) or nasal (6b). Coda nasals must be placeless, or homorganic with the following stop.

(6) *Pali cluster simplification* (Zec 1995: 157)

a.	sup+ta	sutta	'to sleep'
	tap+ta	tatta	'to shine'
	caj+ta	catta	'give out'
b.	dam+ta	danta	'to tame'
	vam+ta	vanta	'to investigate'

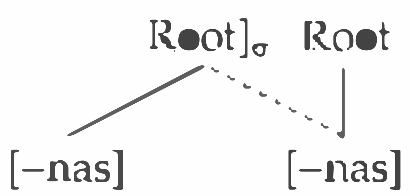
Coda constraints such as those in (7) prevent illicit codas. (7a) states that "if there is a syllable-final consonant which is singly linked, its melody cannot be [−nasal]"; (7b) states that "if there is a syllable-final consonant which is singly linked, its melody must be [+nasal]."

[6] Within Optimality Theory, two types of constraints have been posited to account for positional asymmetries in the realization of segmental features. See CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS for discussion.

(7) *Codas in Pali* (following Itô 1986)

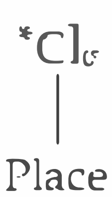
Geminates, where the melody is doubly linked both to the coda of one syllable and to the onset of the following syllable (CHAPTER 37: GEMINATES), violate neither (7a) nor (7b), because the melody is not uniquely linked to a [nasal] feature.

(8)



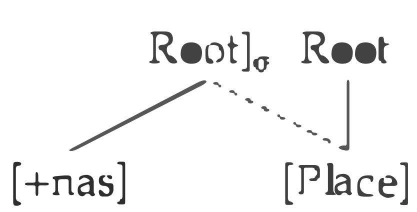
The same approach can be applied to the fact that codas in Pali are either placeless, as in the case of nasal codas, or homorganic with the following stop (9).

(9) *Coda place in Pali* (following Itô 1989: 224)



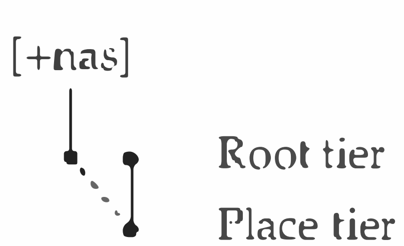
A coda consonant can be specified for place as long as the Place node is not uniquely linked to the coda consonant. If a coda nasal cannot share Place with another segment, it will remain placeless.

(10)



The restrictiveness of potential triggers and targets of neutralization have provided fruitful venues for discovering the organization of features at the phonological level. There have been many proposals for the organization of features into a hierarchical set structure within Autosegmental Phonology (see McCarthy 1988 and Clements and Hume 1995 for overviews of proposals in feature geometry; see also CHAPTER 27: THE ORGANIZATION OF FEATURES). By assuming that the different features for place of articulation are hierarchically linked to a Place node, nasal place assimilation in Yoruba can be elegantly and economically modeled in terms of the spreading of the place node (11).

(11) *Place assimilation in a feature-geometric organization* (Pulleyblank 1995: 9)



Within this type of feature-geometric framework, non-assimilatory contrast reductions are generally treated as a matter of delinking of branches of a feature free. In the Kelantan dialect of Malay (Austronesian), for example, /p t k/ neutralize to [ʔ], and /s ʃ/ become [h] (12).

(12) *Kelantan Malay place neutralization* (Teoh 1988)

/ikat/	ikaʔ	'tie'	/səsak/	səsaʔ	'crowded'
/dakap/	dakaʔ	'embrace'	/hampas/	hapah	'husk'

Debuccalization to [ʔ] and [h] can be viewed as delinking of the Place node (13). The fact that /p t k/ debuccalize to [ʔ], but /s/ to [h], can be attributed to the fact that non-place features of the underlying segment (e.g. [continuant]) are left intact.

(13) *Formalization of* /s/ → [h] *and* /p t k/ → [ʔ]

Other features [−voice, ±cont]
ǂ
Place node • Place

Adopting the framework of Optimality Theory (OT; Prince and Smolensky 1993), which determines the contrastive status of a feature F via the interaction of a constraint that requires the preservation of F and constraints on the rest of the system (Kirchner 1997), Lombardi (2001b) analyzes place neutralization such as (12) in terms of the interaction between consonantal place faithfulness and a family of universally ranked place markedness constraints ((14); cf. Prince and Smolensky 1993; Smolensky 1993; see also CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). Unlike the position-specific markedness constraints in (4) and (5), this family of markedness constraints captures the idea that pharyngeals, including /ʔ h/ (McCarthy 1994), are less marked than coronals in general, irrespective of position. The tableau in (15) illustrates a markedness-based treatment of coda place neutralization.

(14) *Place hierarchy*

*Dors/*Lab >> *Cor >> *Phar (Lombardi 2001b: 29)

(15) *Place neutralization in Kelantan Malay* (Lombardi 2001b: 31)[7]

/ikat/	Max	Dep	*Dors/*Lab	*Cor	*Phar	*Max(Place)
a. ikat			*	*!		
b. ikati		*!	*	*		
c. ika	*!		*			*
☞ d. ikaʔ			*		*	*

[7] The CodaCons constraint, which bans any Place feature in coda consonants, is omitted from this tableau, because it is not directly relevant in the present evaluation.

The position-specificity of place neutralization is captured by the universal ranking of IDENT(OnsPlace) >> MAX(Place). IDENT(OnsPlace) requires that an onset consonant have the Place of its input correspondent, while MAX(Place) requires that an underlying Place feature have an output correspondent. Assuming that ID(OnsPlace), which preserves underlying place features in the onset only, always outranks the place hierarchy in (14), a place distinction in coda position is neutralized due to the dominance of the markedness constraints in (14) over MAX(Place). Markedness violations in coda position cannot be resolved by deleting the offending coda due to the high ranking of MAX, which penalizes deletion, nor can it be resolved by the addition of a final vowel due to the high ranking of DEP, which penalizes epenthesis. Place distinctions neutralize toward [ʔ], since *PHAR, which penalizes /ʔ/, among other things, is ranked lower than the other place markedness constraints; the candidate with a /ʔ/ coda (15d) is thus preferred over the fully faithful candidate (15a), which has a coronal coda.

3.2 *Richness of cues and contrast maintenance*

As the last case study illustrates, the notion of markedness is often invoked to account for the directionality of contrast reduction (CHAPTER 4: MARKEDNESS). Laryngeal neutralization in coda position is said to favor voicelessness, because laryngeal features such as [voice] and [constricted glottis] are more marked than voicelessness (Lombardi 1991; CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). Similarly, the fact that vowels in unstressed position often neutralize to schwa is often attributed to the unmarkedness of schwa (CHAPTER 26: SCHWA). The definition of markedness is a matter of great debate, however, see CHAPTER 4: MARKEDNESS.

In an effort to provide an objective basis for markedness, some scholars have proposed to ground the notion of markedness in terms of speakers' partial understanding of the physical conditions under which speech is produced and perceived. This phonetically based notion of markedness leads to the development of a cue-based approach to contrast reduction (see Hayes *et al.* 2004 and references therein). The basic assumption of cue-based approaches to contrast reduction is that a contrast is suspended in positions where the relevant contrast-supporting cues are diminished; a contrast in such cue-impoverished environments may be maintained only at the cost of additional articulatory maneuvers. A contrast is licensed in positions that are rich in perceptual cues that maximize the contrast's perceptibility. Alveolars and retroflexes, for example, are most easily distinguished by their VC transition profiles. Positions where VC transition is impoverished or non-existent, such as word-initial and postconsonantal positions, tend to be loci where the alveolar ~ retroflex contrast is eliminated (Steriade 1994). For example, the Australian language Bunuba contrasts apical alveolar and retroflex word-medially (e.g. /biɖi/ 'thigh' *vs.* /widigi/ 'stick insect'), but only apical alveolars are found word-initially (Rumsey 2000). The only exception to this restriction is when a subsequent syllable contains [ɖ ɳ ɭ]; in such instances (e.g. /ɳaɖʌ/ 'short', /ɖuɭu/ 'heart'), long-distance retroflexion is assumed to be what licenses the presence of retroflexion word-initially (Hamann 2003). Even when VC transitions are present, however, retroflexes are often avoided in the environment of /i/. For example, retroflex fricative and affricate series in several Chinese dialects are in complementary distribution with the alveo-palatals: before a high front vowel, only alveo-palatals are found, while the retroflexes occur elsewhere (Yip 1996). Hamann (2003) explains this avoidance of retroflexes in the environment /i/ as a result of the articulatory incompatibility between the production of these segments; a flat tongue middle and retracted tongue back configuration for retroflexion cannot be combined with the high tongue middle and fronted tongue back necessary for front vowels. Languages often restrict the distribution of contour tones to phonemic long vowels (e.g. Somali and Navajo), stressed syllables (e.g. Xhosa and Jemez), and word-final positions (Zhang 2001, 2002). While it is difficult to characterize these positions in structural or prosodic terms in a unifying way, they have in common rhyme durations that are long, sonorous, and high in intensity. This fact has led some researchers to hypothesize a long sonorous rhyme duration as the unifying factor for privileged contour tone licensers (Gordon 1999, 2001; Zhang 2001). Obstruents are often voiced after a nasal, resulting in voicing neutralization (Luyia (Niger- Congo) /N + p t k ts c/ → [mb nd ŋg nz ɲj]; Herbert 1986: 236). Hayes and Stivers (1995) attribute the preference for post-nasal voicing to the effects of "velar pumping," which arises from vertical motion of a closed velum, and of "nasal leak," the leakage of air through a nearly closed velar port during the coarticulatory period between oral and nasal segments.

Structure-based accounts have difficulties accounting for languages, such as Lithuanian, which licenses laryngeal contrasts in pre-sonorant position, regardless of whether the following sonorant is tautosyllabic or heterosyllabic (see also Ancient Greek and Sanskrit; Steriade 1997). From a cue-based perspective, the reduction of laryngeal contrasts in preconsonantal and final positions follows from the fact that many of the relevant cues for the perception of voicing (closure voicing, closure duration, duration of preceding vowel, F0 and F1 values in preceding and following vowels, VOT values, burst duration, and amplitude) are endangered in those positions (see also CHAPTER 8: SONORANTS). The more impoverished the available perceptual cues are, the less sustainable the laryngeal contrast is. Thus, word-initial preconsonantal position is least hospitable to a contrast in voicing, while inter-sonorant position is most ideal for voicing realization. Formally, a cue-based account of contrast reduction may be modeled as the interaction between constraints on contrast maintenance and markedness constraints induced from phonetic knowledge (Steriade 1997; Hayes 1999). Steriade (1997), for example, models [voice] neutralization in terms of the interaction between the constraint PRESERVE[voice], which demands faithfulness to input voice values, and a fixed hierarchy of *VOICE constraints, aligned to a voice perceptibility scale (16).

(16) *Scale of obstruent voicing perceptibility according to context* (Steriade 1997: 11)[8]

V __ [+son] > V __ # > V __ [−son] > [−son] __ [−son], [−son] __ #,
__ [−son]

A language with voicing licensed only before sonorants would have the following ranking:

[8] The > symbol in (16) indicates that voicing is more perceptible in the context to its left than to the context on its right.

(17) *Voice licensed before sonorants* (Steriade 1997: 12)

*Voice / [−son] __ [−son], [−son] __ #, # __ [−son] >> *Voice / V __ [−son] >> *Voice / V __ # >> Preserve[voice] >> *Voice / V __ [+son]

Given that the ranking of constraints projected from a phonetically grounded perceptibility scale has been argued to be universal (CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY), such a model makes strong predictions about the typology of laryngeal neutralization patterns. For example, it predicts that a language with a voicing contrast in word-initial preconsonantal position must also allow a voicing contrast in word-initial, intervocalic, and word-final positions. The Mon-Khmer language Khasi, spoken in the Assam province of India, shows that such a strong prediction does not obtain. As illustrated in (18), Khasi contrasts voiced and voiceless plosives in word-initial preconsonantal position.

(18) *Voicing contrast in initial clusters in Khasi* (Henderson 1992: 62)

bti		'to lead by the hand'	*pdot*	'throat'
bthi		'sticky'	*pdeng*	'middle'
dkar		'tortoise'	*tbian*	'floor'
dkhar		'plainsman'	*tba*	'to feel'
dpei		'ashes'	*pjah*	'cold'
bshad	[bʃaːt]	'civet'	*bdi*	'twenty'

Using evidence from a Frøkjær-Jensen combined oscilloscope and mingograph, Henderson (1992) confirmed the voicing contrast in word-initial preconsonantal position and ruled out the possibility of a svarabhakti vowel between the two stops. What is of interest here is the fact that in syllable-final position there is no distinction between voiced and voiceless stops; final stops are unreleased and frequently accompanied by simultaneous glottal constriction (Henderson 1967: 567). Since a voicing contrast is allowed word-initially before another obstruent, a highly impoverished environment for the maintenance of a voicing contrast, a cue-based approach that maintains the universality of voicing perceptibility necessarily predicts that a voicing contrast should also be maintained in less impoverished environments, such as post-vocalic word-final positions. It is worth noting that counterexamples of this sort do not obviate the validity of a cue-based approach to contrast reduction per se, since the assumption of the universality of cue perceptibility is logically independent of the claim that cue maintenance is the driving force behind contrast maintenance and reduction (see Hume and Johnson 2001b for discussion on the language-specificity of speech perception).

Some cue-based theorists eschew the notion of markedness at the level of the individual segment or feature, and favor instead a contrast maximization account. Dubbed "Dispersion Theory" (Flemming 1995, 1996; Ní Chiosáin and Padgett 2001; Padgett 2003) after Lindblom's (1986, 1990) Theory of Adaptive Dispersion, such a theory of contrast maintains that the selection of a phonological contrast is subject to three functional goals (see Martinet 1952, 1955, 1964 for early formulations of these functional ideas; cf. Silverman 1996, 2004, 2006):

(19) a. Maximize the distinctiveness of contrasts.
b. Minimize articulatory effort.
c. Maximize the number of contrasts.

From this perspective, the dispreference for sound *x* is conceptualized as a dispreference for the sub-maximally distinct contrasts between *x* and other sounds in the particular sound system. As schematized in the ranking in (20), a contrast is formally neutralized in some context if it cannot be realized with a distinctiveness of *d* without violating *Effort, an effort-minimization constraint penalizing some articulation.

(20) MinDist=*d*, *Effort >> MaximizeContrasts

In Belarusian, for example, a five-vowel inventory /i e a o u/ is observed in stressed syllables. In unstressed syllables, /e a o/ reduce to [a] or [ɐ], depending on the position of the vowel relative to the stressed syllable (Barnes 2002: 65). Flemming (2004) argues that this type of vowel reduction is motivated by difficulties in producing distinct F1 contrasts in unstressed positions. Specifically, increasing difficulty in producing a low vowel as a result of vowel duration shortening in unstressed positions leads to the raising of short low vowels; the smaller range of the F1 dimensions for distinguishing F1 contrast then leads to the selection of a smaller number of contrasts. Flemming captures this intuition in terms of the ranking in (21).

(21) Unstressed vowels are short, *ShortLowV, MinDist=F1:3 >>
MaximizeContrasts >> MinDist=F1:4

The constraint Unstressed vowels are short requires unstressed vowels to be shorter than stressed ones. This constraint will be omitted in the subsequent discussion, since it is assumed to be undominated, so that no vowel systems violating this constraint will be permitted in the present context. *ShortLowV (abbreviated *a) is an effort-minimization constraint that penalizes low vowels. The MinDist=Y:X constraints are satisfied by contrasting sounds that differ by at least X distance on the Y dimension. The highest-ranking MinDist constraint that outranks the MaximizeContrasts constraint sets the threshold distance, and the optimal inventory is the one that packs the most contrasting vowels onto the relevant dimension (here F1) without any pair being closer than this threshold.

With the relative positioning of vowels on the F1 dimension stated in (22), Belarusian's three-way vowel height distinction in stressed syllables is predicted in (23). Since the present evaluation concerns only distinctions in vowel height, the back counterparts of vowels in the inventory candidate set are left out for ease of reference. The tableau in (23) shows that a four-way height distinction is suboptimal (23c), because vowels are not distinct enough according to the constraint, MinDist=F1:3. Reducing the height inventory too much (23a) results in excessive contrast reduction, thus incurring more MaximizeContrasts violations relative to the optimal inventory set (23b).

(22)	F1	7	6	5	4	3	2	1
		a	ɐ	ɛ	e	e̝	ɪ	i
				a	ə			

(23) *Belarusian: Vowels in stressed syllables*

	*a	MinDist= F1:3	Maximize Contrasts	MinDist= F1:4
a. 'i 'a			✓✓!	
☞ b. 'i 'e 'a			✓✓✓	**
c. 'i 'e 'ɛ 'a		*!	✓✓✓	****

In the unstressed syllables, the constraint *ShortLowV (*a) becomes applicable. It rules out the candidate vowel inventory [i e a], because of the presence of [a]. The three-way height distinction cannot be maintained even if the low vowel [a] is avoided, the distance between [e] and [ɐ] being insufficient, due to the high ranking MinDist=F1:3 constraint. The winning candidate has only two vowel heights, which fares worse by MaximizeContrasts, but satisfies the higher-ranked minimum distance requirements.

(24) *Belarusian: Vowels in unstressed syllables*

	*a	MinDist= F1:3	MinDist= F1:4	Maximize Contrasts
☞ a. i ɐ				✓✓
b. i e ɐ		*!	**	✓✓✓
c. i e a	*!		**	✓✓✓

Within Dispersion Theory, the objects of analysis are systems of oppositions. The notion of contrast reduction is thus given a genuine expression in such an analysis. Whereas most other approaches view mergers and neutralization as the results of the application of constraints or rules that prevent the expression of individual segments or features, Dispersion Theory holds that mergers and neutralization follow from the number of oppositions a language makes available in different contexts. It should be noted that, because of its insistence on looking at systems of contrast from the perspective of the language as a whole, Dispersion Theory raises questions regarding how phonological derivation is implemented in such a model (Boersma 1998: 361; but see Ní Chiosáin and Padgett 2001 and Padgett 2003 for a response to this problem).

This section has reviewed major theories of contrast reduction, showing that proposals range from completely structure-dependent accounts to theories that embrace the full phonetic substance of sound patterns. The debate on what a proper theory of contrast reduction is, however, might ultimately rest on resolving a more fundamental question – does synchronic contrast reduction truly exist? This is the topic of the next section.

4 Do real synchronic mergers and neutralization exist?

Until recently, most theories of phonology have assumed some form of lexical minimality (the minimization of lexically stored information; Chomsky and Halle

1968: 381; Steriade 1995: 114; see also CHAPTER 1: UNDERLYING REPRESENTATIONS) and feature economy (the minimization the ratio of features to segments in an "alphabet"; Clements 2003; see also CHAPTER 17: DISTINCTIVE FEATURES). In early generative phonology, for example, the underlying alphabet is the minimal sound set needed to express surface differences between distinct morphemes; at the level of the underlying representation, no allophonic variants are present. Theories differ in the number of levels of representation allowed (e.g. Lexical Phonology and Morphology (Kiparsky 1982, 1985; Mohanan 1982) recognizes three levels of representations: underlying, lexical, and phonetic) and the degree of minimality assumed at each level. Common to these early theories of phonology, however, is the premise that, out of the vast sea of phonetic signals, only a small subset of phonetic properties are contrastive in a given language (Sapir 1933; Trubetzkoy 1939; Jakobson *et al.* 1952; Hockett 1955; Chomsky and Halle 1968; Kiparsky 1982, 1985). Contrast is encoded in terms of a difference between $+F_i$ and $-F_i$ for some finite set of features F_i, and contrast reduction corresponds to the elimination of this difference (i.e. the outcome of such a reduction is either $+F_i$, $-F_i$, or null).

Non-distinctive phonetic properties are treated in one of two ways. To begin with, features that do not distinguish lexical items may be underspecified in the lexical entries (e.g. Archangeli 1988; Pulleyblank 1995; Steriade 1995; Clements 2003; CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION). The feature [voice] in sonorants, for example, is non-contrastive, and thus redundant, in languages such as English, which do not distinguish between voiced and voiceless sonorants (CHAPTER 8: SONORANTS; CHAPTER 13: THE STRICTURE FEATURES). Sonorants are underspecified for voicing, i.e. sonorants bear no value for the feature [voice]. Such an assumption of non-contrastive feature underspecification has important theoretical consequences for the treatment of transparency effects in the phonology of the feature [voice]. For example, as seen earlier, nasals do not induce regressive voicing assimilation in Dutch, but voiced obstruents do, suggesting that only voiced obstruents are underlyingly specified for the feature [voice]. The other treatment of non-distinctive phonetic properties is to exclude them from the feature pool altogether. For example, vowels are longer before voiced stops than before voiceless ones in American English (*bat* [bæt] *vs. bad* [bæˑd]). Peterson and Lehiste (1960) suggest that the ratio of vocoid duration before voiceless consonants to that before voiced consonants in American English is 2 : 3. Such a difference in vocoid duration which covaries with the voicing of the following consonant is generally dismissed as the effect of automatic phonetics, and thus assumed to play no role in any phonological analysis; features such as [slightly long] would not be part of the universe of phonological features.[9]

As Labov *et al.* (1991: 38) point out, the assumptions that "contrasts were discrete and binary, that there was no such thing as a small difference in sound, that production and perception were symmetrical, and that introspections were reliable" have received increased scrutiny in recent years (CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY). For example, Dispersion Theory's admission of phonological constraints that regulate features

[9] While these subfeatural cues might not be distinctive, they may nonetheless have enhancing functions (Stevens and Keyser 1989; Stevens *et al.* 1986; Keyser and Stevens 2001).

along scalar dimensions, rather than in terms of binary oppositions, already foreshadows the move away from a discrete and binary notion of contrasts (e.g. MinDist=Y:X constraints evaluate distances along some phonetic dimensions such as F1). Mounting evidence for near mergers and incomplete neutralization raises further questions about the validity of these abovementioned assumptions. This is the topic of the next section.

4.1 Near mergers and incomplete neutralization

Near merger describes the situation where speakers consistently report that two classes of sounds are the same, yet consistently differentiate them in production at a better than chance level. Labov *et al.* (1972: ch. 6), for example, reports that speakers in New York City differentiate words such as *source* and *sauce* in production, but report no distinction between them in perception. Similar near mergers have been reported in other varieties of English (e.g. *fool* and *full* in Albuquerque (Di Paolo 1988); *too vs. toe* and *beer vs. bear* in Norwich (Trudgill 1974); *line vs. loin* in Essex (Labov 1971; Nunberg 1980); *meat vs. mate* in Belfast (Milroy and Harris 1980; Harris 1985)). Near mergers are not restricted to segmental contrasts. Yu (2007b), for example, demonstrates that derived mid-rising tones in Cantonese show a small but statistically significant difference in F0 from underived mid-rising tones. Similar to near mergers, incomplete neutralization refers to reports of small but consistent phonetic differences between segments that are supposedly neutralized in certain environments. Flapping is often cited as a neutralizing phonological alternation in American English; underlying /t/ and /d/ surface as dental flaps or taps when followed by an unstressed vowel (chapter 113: flapping in american english). Word-final and preconsonantal obstruent devoicing is another classic example of a neutralizing sound pattern (chapter 69: final devoicing and final laryngeal neutralization).

Incomplete neutralization has been reported outside the domain of obstruent voicing as well. In Eastern Andalusian Spanish, for example, the combined effect of word-internal coda aspiration and the gemination of the consonant following the aspirated coda leads to potential neutralization (e.g. [kahtːa] for both /kasta/ 'caste' and /kapta/ 's/he captures'). Gerfen (2002), however, reports that aspirating an /s/ results in a longer duration of aspiration, while aspirating a /p/ or /k/ results in longer medial consonant gemination (see also Gerfen and Hall 2001). Bishop (2007) found that listeners make use of the length of the consonant following aspiration as a cue for making phonemic decisions regarding the nature of the underlying coda. In many languages, an epenthetic stop can occur within nasal–fricative or heterorganic nasal–stop clusters (e.g. English *dreamt* [dɹɛmt] ~ [dɹɛmpt]; *prince* [pɹɪns] ~ [pɹɪnts]). Several studies have found that such epenthetic stops are phonetically different from underlying stops in the same environment. Fourakis and Port (1986), for example, found that underlying /t/ in words like *prints* [pɹɪnts] are significantly longer and the neighboring nasal significantly shorter than epenthetic [t] in words like *prince*. Dinnsen (1985), citing Rudin (1980), reports that long vowels deriving from underlying /VgV/ sequences in Turkish are 13 percent longer than the underlying long vowel /Vː/. Simonet *et al.* (2008) report that the so-called /r/ ~ /l/ neutralization in post-nuclear position in Puerto Rican Spanish (e.g. /'arma/ → ['alma] 'weapon' *vs.* /'alma/ → ['alma] 'soul') is incomplete. Based on measurements of duration of the vowel

+ liquid sequences and examination of formant values and trajectories, Simonet *et al.* (2008) conclude that, while post-nuclear /r/ is similar to post-nuclear /l/, there nonetheless exist systematic durational and spectral differences, suggesting that the two liquids have not completely merged.

Since traditional theories of the phonetics–phonology interface assume that phonological representations in the lexicon are categorical, contrastive elements, and since the phonetic implementation component computes the degree and timing of articulatory gestures, which are gradient and variable, the discovery of near mergers and incomplete neutralization presents a curious conundrum. For a given underlying distinction +F and –F, how can an output –F that corresponds to an underlying +F display systematically different surface phonetic realization from an output –F that corresponds to an underlying –F, when information flow is supposed to be strictly unidirectional? In such a model, no articulatory plan can look backward to phonological encoding, nor can phonological encoding look back to the lexical level. No lexical information can influence the phonetic implementation directly either, bypassing the level of phonological encoding. On this view, the categorical form of a lexeme wholly determines the phonetic outcome. Phonetic variations on the surface are considered artifacts of the context or performance-induced anomalies.

In light of such conceptual difficulties, many have sought to explain away the observed sub-phonemic phonetic differences as a consequence of orthographic influence or as variation in speaking style. For example, it has been found that the less the experimental design emphasizes the role of orthography, the smaller the durational effects (Fourakis and Iverson 1984; Jassem and Richter 1989). Port and Crawford (1989) found that discriminant analysis to classify productions by underlying final voicing was most successful (78 percent correct) when speakers dictated the words, but least successful (55 percent correct) when target words were embedded in sentences that do not draw attention to the minimal pairs (whether read or repeated orally). But not all cases of near mergers and incomplete neutralization can be attributed to performance factors. Warner *et al.* (2004), for example, found sub-phonemic durational differences in the case of final devoicing in Dutch, even when possible orthographic influence was controlled for as a confound. Yu (2007b) found incomplete merger of underived and morphologically derived mid-rising tones in Cantonese, a language whose orthography does not indicate tone. Further support for the existence of a suspended contrast comes from the fact that speakers appear to have some access to subtle phonetic differences. As noted earlier, Bishop (2007) found that Andalusian Spanish speakers can make use of subtle closure duration differences to recover underlying coda consonants. In the case of final devoicing in Dutch, listeners not only can perceive durational differences (Warner *et al.* 2004), they even use these sub-phonemic distinctions to hypothesize which past tense allomorph nonce forms would take (Ernestus and Baayen 2003; CHAPTER 99: PHONOLOGICALLY CONDITIONED ALLOMORPH SELECTION).

4.2 *Approaches to sub-phonemic phonetic differences*

Sub-phonemic distinctions have been analyzed as the result of paradigm uniformity among morphologically related neighbors (e.g. phonetic analogy; Steriade 2000; Yu 2007a; CHAPTER 83: PARADIGMS; CHAPTER 87: NEIGHBORHOOD EFFECTS).

Steriade (2000), for example, argues that grammars prefer words within a paradigm to be uniform.[10] Steriade extends this paradigm uniformity preference to the phonetic level. French, for example, has an optional schwa deletion which creates ostensibly homophonous strings (e.g. *bas retrouvé* [baʁətʁuve] 'stocking found again' → *bas r'trouvé* [baʁtʁuve] *vs. bar trouvé* [baʁtʁuve] 'bar found'). Various studies have shown that the consonant to the left of the syllable of the deleted schwa maintains phonetic qualities that would only be expected if the schwa were still present (Rialland 1986; Fougeron and Steriade 1997). Steriade (2000) interprets such unexpected phonetic differences as the results of phonetic analogy; forms with schwa deletion are influenced phonetically by the corresponding schwa-full forms (e.g. /r/ in *bas r'trouvé* [baʁtʁuve] takes on onset-like articulation from the /r/ in the related phrase *bas retrouvé* [baʁətʁuve]).

Van Oostendorp (2008) argues that incomplete neutralization in final devoicing can be captured within a Containment model of OT in terms of a turbid representation of phonological outputs (Goldrick 2001). Output structures are characterized in terms of two types of relations: a Projection relation, which is an abstract structural relationship holding between a segment and the feature (represented by ↑ in (25)), and a Pronunciation relation, an output relationship that holds between the feature and the segment and describes the output realization of a structure (represented by ↓ in (25)). On this conception, a three-way distinction obtains between segments that are underlyingly voiceless (i.e. they lack the feature [voice]), segments that are underlyingly voiced and pronounced voiced, and segments that are underlyingly voiced, but are not realized as voiced on the surface (25).

(25) *A three-way voicing distinction using turbidity theory*

a.	taːt	b.	tad	c.	taːd̥
			↑↓		↑
			[voice]		[voice]

The selection of a representation like (25c) would be determined by the interaction between markedness constraints that disfavor coda voicing and the constraint RECIPROCITY(X,F), which holds that if a segment X entertains a projection relation with a feature F, then F must entertain a pronunciation relation with the segment X. Because of their structural differences, (25a)–(25c) will show different surface phonetic realizations.

These phonological approaches assume that cases of incomplete neutralization are in fact complete at the phonological level and that the output segment is phonologically unvoiced. The sub-phonemic differences observed would either be due to analogical influences from related forms that retain voicing or to covert structural differences among outputs. Is a complete neutralization interpretation of incomplete neutralization a necessity, or even desirable? The answer to this question hinges on the conception of the phonetics–phonology interface and, specifically, the nature of allophony. What should be considered extrinsic allophones (i.e. allophones that are phonologically governed), and what should be considered intrinsic (i.e. those introduced by phonetic variability; Wang and Fillmore 1961;

[10] A paradigm is defined here as "a set of words sharing a morpheme, e.g. {*bomb, bomb-ing, bomb-ard*, . . . }, or a set of phrases sharing a word, e.g. {*bomb, the bomb*, . . . }" (Steriade 2000).

Ladefoged 1971; Tatham 1971)? Must extrinsic allophones be governed by changes in discrete distinctive feature values, or can extrinsic allophones be gradient? The next section offers an alternative interpretation of near mergers and incomplete neutralization, which appeals to the notion of a covert contrast.

4.3 Sub-phonemic distinctions as covert contrasts

Near mergers and incomplete neutralization are problematic from the point of view of a model of the interface between phonetics and phonology sketched above, because, if the phonetic implementation component accounts only for variations due to biomechanical and aerodynamic factors, it is anomalous, to say the least, that speakers of a language with [voice] neutralization vary the realization of the neutralized sounds in accordance with the feature value of their non-neutralized counterparts. The above model of the phonetics–phonology interface is arguably simplistic, however. Kingston and Diehl (1994) articulate a model of the phonetics–phonology interface that affords the phonological component greater control over the range of variability in the phonetic implementation of contrasts. Elasto-inertial, biomechanical, aerodynamic, psychoacoustic, and perceptual constraints delimit what a speaker (or listener) *can* do, but not what they *must* do. Within this conception of the phonetics–phonology interface, a phonemic contrast is taken to be "any difference in the feature content or arrangement of an utterance's phonological representation which may convey a difference in semantic interpretation" and allophones are "any phonetic variant of a distinctive feature specification or arrangement of such specification that occurs in a particular context" (1994: 420, fn. 2). To illustrate this framework more concretely, consider Kingston and Diehl's summary of the phonetic variants of English stops contrasting for [voice] (see also Silverman 2004).

Table 80.1 illustrates the fact that the contrastive feature [+voice] in English shows great variability in its phonetic realization. In word-initial position, for example, [+voice] stops are often realized as voiceless unaspirated, even when the preceding word ends in a vowel (Caisse 1982; Docherty 1989). Kingston and Diehl (1994) interpret such data as showing that speakers choose between two active articulations in producing initial [+voice] stops in English: delay glottal closure until the stop release, or close the glottis but expand the oral cavity to overcome the difficulty of initiating voicing. Such controlled variation is made possible by the fact that there are typically multiple, auditorily independent correlates that serve as distinct bases for a minimal phonological distinction. As noted in Stevens and Blumstein (1981), [+voice] consonants are characterized by the "presence of low-frequency spectral energy or periodicity over a time interval of 20 to 30 msecs in the vicinity of the acoustic discontinuity that precedes or follows the consonantal constriction interval" (1981: 29). This low-frequency property, as Kingston and Diehl (1994) call it, has multiple supporting sub-properties such as voicing during the consonant constriction interval, a low F1 near the constriction interval, and a low F0 in the same region, as well as enhancing properties such as the duration ratio between a consonant and its preceding vowel. These properties do not all surface in all positions. Crucially, while [+voice] stops do not show prevoicing in word-initial position, the [voice] contrast is nonetheless maintained because [–voice] stops tend to have longer VOT, stronger burst energy and higher F1 and F0 following the consonant constriction interval.

Table 80.1 Summary of the phonetic variants of English stops that contrast for [voice]

	[+voice]	[−voice]
Utterance-initial or pre-tonic	short lag VOT	long lag VOT
	F1 lower	F1 higher
	F0 lower	F0 higher
	weaker burst	stronger burst
Intervocalic or post-tonic	closure voicing	no closure voicing
	short closure	longer closure
	longer preceding vowel	shorter preceding vowel
	F1 lower	F1 higher
	F0 lower	F0 higher
Utterance-final and post-vocalic	longer preceding vowel	shorter preceding vowel
	closure voicing possible	no closure voicing
	short closure	longer closure
	F1 lower	F1 higher

From this perspective of the phonetics–phonology interface, sub-phonemic differences observed in near mergers and incomplete neutralization are no more different from those observed between allophones appearing in different phonetic contexts. As noted in Steriade (1997), the percept of voicing hinges on a multitude of acoustic cues: burst amplitude, closure duration, voicing during the closure period, voice onset time, and vowel onset and offset. Phonetic cues that support a [voice] contrast in word-final positions are intrinsically impoverished relative to cues available in word-initial and word-medial positions. Nonetheless, many languages maintain the contrast in word-final positions because there remain sufficient cues that can differentiate the underlying phonological contrast. (See CHAPTER 113: FLAPPING IN AMERICAN ENGLISH for additional evidence on this interpretation.)

The interpretation of near mergers and incomplete neutralization advocated here suggests that traditional methods of introspection and field elicitation may not be adequate in detecting covert contrast (CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY). Self-introspection faces inherent problems of analyst bias and thus should not be taken as a definitive source of information. The phonetician's ears are, after all, human ears. Commutation tests are essentially armchair psycholinguistic tasks that require language consultants to perform a same–different task, with minimal control for potential confounds. Subject responses are inherently probabilistic; analysts insisting on dichotomizing a continuous function will find confident responses when the samples have a wide separation in the sample space. Samples that straddle regions of great overlap, as in the case of near mergers and incomplete neutralization, will elicit more ambiguous responses. Contrasts not detected by linguists using traditional methods of elicitation may nonetheless be detected by native speakers, as demonstrated in laboratory studies reviewed above.

4.4 *Covert contrasts as systems in transition*

The existence of covert contrast is readily understandable from the perspective of sound change and phonologization of phonetic variation (CHAPTER 93: SOUND CHANGE). In his seminal work on phonologization, Hyman (1976) conceptualizes the emergence of phonemic tonal distinctions as a three-stage process. At Stage 1, a language displays physiologically based consonantal voicing-induced pitch perturbations on the neighboring vowel. A language reaches Stage 2 when pitch perturbation becomes exaggerated to such an extent that the pitch variation cannot be attributed entirely to the physiological properties of the preceding consonant's voicing (e.g. *[pa] > [pá] and *[ba] > [bà]). The transition from Stage 1 to Stage 2 – when an intrinsic, thus unintended, variation in pitch associated with consonantal realization becomes an extrinsic feature of the vowel – is phonologization. A language reaches Stage 3 when the voicing distinction is lost completely, and the pitch distinction on vowels becomes the sole feature that signals a meaning difference between words. That is, the language has undergone the phonemicization of tone (i.e. */pa/ > [pá] and */ba/ > [pà]). From the perspective of this model of sound change, covert contrast represents a language at Stage 2 and possibly in transition to Stage 3. That is, the old contrast (e.g. obstruent voicing) has not completely disappeared (i.e. been neutralized), but the new contrast (e.g. tonal distinction) has not fully emerged either. A language in Stage 2 is in principle unstable. As Hyman points out, "accompanying every phonologization is a potential dephonologization" (Hyman 1976: 410). The emergence of a tonal distinction as a result of the phonologization of intrinsic pitch perturbation of obstruent voicing entails the eventual destruction (i.e. neutralization) of the original voicing contrast. The evolution of vowel duration and consonant voicing covariation provides an instructive example of phonologization and its connection to the emergence of covert contrast. As reviewed in Solé (2007), languages differ in the amount of control the speakers have over the maintenance of this sub-phonemic duration difference. Solé (2007) found that English speakers actively maintain durational differences before voiced and voiceless stops, regardless of speaking rates, while speakers of Catalan and Arabic do not exhibit similar control over such sub-phonemic duration differences. Her findings suggest that English has already partially phonologized the effect of consonant voicing on vowel duration, while Catalan and Arabic have not. Recall that one commonly observed feature of the incomplete neutralization of final devoicing is a vowel duration difference. Following Hyman's dictum that the phonologization of one feature carries the seeds of the destruction of another, the phonologization of a sub-phonemic vowel duration difference entails an eventual loss of the voicing contrast in the following stops. The reasons why such a correlation exists are still a matter of debate. Two factors are noteworthy in this context. First, the longer vowel before voiced stops and the shorter vowel before voiceless stops are, strictly speaking, in complementary distribution. Likewise, post-vocalic voiced and voiceless stops are also in complementary distribution, since they do not appear in the same context. This type of analytic ambiguity (i.e. between vowel duration and consonantal voicing) is typical of a language undergoing phonologization. Second, research on auditory category learning has shown that listeners are not only sensitive to the distributional information of the category cues, but also acquire unidimensional contrasts more readily than multidimensional ones (Goudbeek 2006; Clayards 2008; Clayards *et al.* 2008; Goudbeek *et al.* 2008). Such results suggest that, all else being equal, listeners will rely more heavily on a single cue for category identification even when multiple cues are available in the signal. For example, as voicing during stop closure becomes less prominent as a feature of voiced stops in final position, vowel length becomes the more reliable contrastive feature. When voicing in closure ceases to be a feature of final obstruents altogether, a contrast in vowel length is expected to emerge. Friulian, a Romance language spoken in northeastern Italy, provides an instructive example of this type of cue trade-off in phonologization. In Friulian, vowel length (CHAPTER 20: THE REPRESENTATION OF VOWEL LENGTH) is only distinctive in a stressed word-final syllable closed by a single consonant.

(26) *Vowel length distinction in Friulian* (Baroni and Vanelli 2000: 16)

a.	['la:t]	'gone (MASC)'	b.	['lat]	'milk'
	['bru:t]	'brother, mother-in-law'		['brut]	'ugly'
	[fi'ni:t]	'finished (MASC)'		['frit]	'fried (MASC)'
	['pa:s]	'peace'		['pas]	'step'
	['fu:k]	'fire'		['tɔk]	'piece'

Stressed vowels are always phonologically long before [r] (['la:rk] 'large (MASC)') and always short when they are not in the last syllable of a word ([kan'tade] 'sung (FEM)'), when they occur in the final open syllable ([ku'si] 'so'), and when they are in a final syllable closed by a consonant cluster, nasal, or affricate (['gust] 'taste', ['maŋ] 'hand', ['brat͡ʃ] 'arm'). Of particular relevance here is the fact that vowel length in word-final syllables before obstruents is predictable: the stressed vowel is long if the following consonant is realized as voiced in intervocalic position (27a); if the following consonant is voiceless intervocalically, the stressed vowel is short (27b).

(27) *Vowel length and consonant voicing* (Baroni and Vanelli 2000: 17)

a.	['la:t]	'gone (MASC)'	['lade]	'gone (FEM)'
	[fi'ni:t]	'finished (MASC)'	[fi'nide]	'finished (FEM)'
	['pe:s]	'snow'	[pe'za]	'to snow'
	['fu:k]	'fire'	[fogo'la:r]	'fireplace'
b.	['lat]	'milk'	[la'te]	'to breastfeed'
	['pas]	'pass'	[pa'sa]	'to pass'
	[pa'taf]	'slap'	[pata'fa]	'to slap'
	['tɔk]	'piece'	[tu'kut]	'little piece'

Based on acoustic evidence, Baroni and Vanelli (2000) establish that long vowels are more than twice as long as the short ones and word-final obstruents are indeed voiceless (i.e. no voicing during closure). While final [t] corresponding to medial [d] is significantly shorter than final [t] corresponding to medial [t], this difference is only observed after certain vowel qualities. Their findings suggest that, while Friulian final obstruent devoicing is incomplete (i.e. there remains some difference between underlying /d/ and underlying /t/ in final positions), this difference is mainly carried by the closure duration of the obstruent and only in very restricted contexts. On the other hand, a full-blown vowel quantity difference has emerged in its place. The salience of this vowel length contrast is exemplified by the behavior of vowel length in loanword adaptation. Friulian does not preserve the consonantal length contrast in borrowed Italian words. However, longer vowels before single consonants in Italian are treated as long in Friulian if they occur in word-final position: [impje'gaːt] 'clerk (MASC)', from Italian [impje'gaːto]. When such long vowels occur in word-internal position, they become short and the following obstruent is voiced ([impje'gade] 'clerk (FEM)'). When borrowed short vowels occur in word-internal position, the obstruents remain voiceless (e.g. [a'fiːt]/[afi'tut] 'rent/little rent'; Italian [a'fːitːo]). This loanword evidence suggests that Friulian has restructured its system to one with a limited vowel length contrast; voicing variation has become secondary to vowel length difference.

The possibility of a covert contrast in purported cases of neutralization raises questions about the existence of genuine instances of neutralization. Kim and Jongman (1996), for example, report that coda neutralization (i.e. word-final coronal obstruents (e.g. /t t^h s/) are all phonetically realized as [t]) is complete (CHAPTER 111: LARYNGEAL CONTRAST IN KOREAN). Based on both production and perceptual data, they conclude that complete neutralization is observed despite the fact that Korean orthography distinguishes between the different underlying consonants. The difference between genuine neutralization *vs.* covert contrasts might also be related to the nature of the evidence supporting the claim of neutralization. Evidence for neutralization may come from distributional information alone (e.g. laryngeal neutralization in coda position in Cantonese) or may additionally be supported by morphological means (e.g. obstruent devoicing in Dutch). All reported cases of incomplete neutralization pertain to morphologically sensitive neutralization. In sum, the question of how pervasive covert contrasts and complete neutralization are is ultimately an empirical one. More systematic phonetic and psycholinguistic investigations are needed to answer this fundamental question in contrast reduction research.

REFERENCES

Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5. 183–207.

Baker, Brett. 1999. Word structure in Ngalakgan. Ph.D. dissertation, University of Sydney.

Barnes, Jonathan. 2002. Positional neutralization: A phonologization approach to typological patterns. Ph.D. dissertation, University of California, Berkeley.

Baroni, Marco & Laura Vanelli. 2000. The relations between vowel length and consonantal voicing in Friulian. In Lori Repetti (ed.) *Phonological theory and the dialects of Italy*, 13–44. Amsterdam & Philadelphia: John Benjamins.

Bauer, Robert S. & Paul K. Benedict. 1997. *Modern Cantonese phonology*. Berlin & New York: Mouton de Gruyter.

Beckman, Jill N. 1998. Positional faithfulness. Ph.D. dissertation, University of Massachusetts, Amherst.

Bishop, Jason. 2007. Incomplete neutralization in Eastern Andalusian Spanish: Perceptual consequences of durational differences involved in s-aspiration. In Trouvain & Barry (2007), 1765–1768.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Bueno-Holle, Juan. 2009. Isthmus Zapotec tone and stress. M.A. thesis, University of Chicago.

Caisse, Michelle. 1982. Cross-linguistic differences in fundamental frequency perturbation induced by voiceless unaspirated stops. M.A. thesis, University of California, Berkeley.

Cho, Young-Mee Yu. 1990. Parameters of consonantal assimilation. Ph.D. dissertation, Stanford University.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clayards, Meghan. 2008. The ideal listener: Making optimal use of acoustic-phonetic cues for word recognition. Ph.D. dissertation, University of Rochester.

Clayards, Meghan, Michael K. Tanenhaus, Richard N. Aslin & Robert A. Jacobs. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108. 804–809.

Clements, G. N. 2003. Feature economy in sound systems. *Phonology* 20. 287–333.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

Crosswhite, Katherine. 2004. Vowel reduction. In Hayes *et al.* (2004), 191–231.

de Lacy, Paul. 2002. The formal expression of markedness. Ph.D. dissertation, University of Massachusetts, Amherst.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.

Dinnsen, Daniel A. 1985. A re-examination of phonological neutralization. *Journal of Linguistics* 21. 265–279.

Di Paolo, Marianne. 1988. Pronunciation and categorization in sound change. In Kathleen Ferrara, Becky Brown, Keith Walters & John Baugh (eds.) *Linguistic change and contact: Proceedings of the 16th Annual Conference on New Ways of Analyzing Variation in Language*, 84–92. Austin: University of Texas.

Docherty, Gerard J. 1989. An experimental phonetic study of the timing of voicing in English obstruents. Ph.D. dissertation, University of Edinburgh.

Durand, Jacques & Francis Katamba (eds.) 1995. *Frontiers of phonology: Atoms, structures, derivations*. London & New York: Longman.

Edmondson, Thomas & John T. Bendor-Samuel. 1966. Tone patterns of Etung. *Journal of African Languages* 5. 1–6.

Emeneau, M. B. 1961. *Kolami: A Dravidian language*. Annamalainagar: Annamalai University.

Ernestus, Mirjam & R. Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79. 5–38.

Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles. Published 2002, London & New York: Routledge.

Flemming, Edward. 1996. Evidence for constraints on contrast: The dispersion theory of contrast. *UCLA Working Papers in Phonology* 1. 86–106.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Hayes *et al.* (2004), 232–276.

Fougeron, Cécile & Donca Steriade. 1997. Does deletion of French schwa lead to neutralization of lexical distinctions? *Proceedings of the 5th European Conference on Speech Communication and Technology, University of Patras*, vol. 7, 943–946.

Fourakis, Marios & Gregory K. Iverson. 1984. On the "incomplete neutralization" of German final obstruents. *Phonetica* 41. 140–149.

Fourakis, Marios & Robert F. Port. 1986. Stop epenthesis in English. *Journal of Phonetics* 14. 197–221.

Gerfen, Chip. 2002. Andalusian codas. *Probus* 14. 247–277.

Gerfen, Chip & Kathleen Currie Hall. 2001. Coda aspiration and incomplete neutralisation in Eastern Andalusian Spanish. Unpublished ms., University of North Carolina, Chapel Hill.

Goldrick, Matthew. 2001. Turbid output representations and the unity of opacity. *Papers from the Annual Meeting of the North East Linguistic Society* 30(1). 231–245.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.

Gordon, Matthew. 1999. Syllable weight: Phonetics, phonology, and typology. Ph.D. dissertation, University of California, Los Angeles.

Gordon, Matthew. 2001. A typology of contour tone restrictions. *Studies in Language* 25. 405–444.

Goudbeek, Martijn. 2006. The acquisition of auditory categories. Ph.D. dissertation, Radboud University Nijmegen.

Goudbeek, Martijn, Anne Cutler & Roel Smits. 2008. Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication* 50. 109–125.

Hamann, Silke. 2003. The phonetics and phonology of retroflexes. Ph.D. dissertation, University of Utrecht.

Harris, John. 1985. *Phonological variation and change: Studies in Hiberno-English*. Cambridge: Cambridge University Press.

Hayes, Bruce. 1999. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In Michael Darnell, Edith Moravcsik, Frederick Newmeyer, Michael Noonan & Kathleen Wheatley (eds.) *Functionalism and formalism in linguistics*, vol. 1: *General papers*, 243–285. Amsterdam & Philadelphia: John Benjamins.

Hayes, Bruce & Tanya Stivers. 1995. The phonetics of post-nasal voicing. Unpublished ms., University of California, Los Angeles.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.

Henderson, Eugénie J. A. 1967. Vowel length and vowel quality in Khasi. *Bulletin of the School of Oriental and African Studies* 30. 564–588.

Henderson, Eugénie J. A. 1992. Khasi clusters and Greenberg's universals. *Mon-Khmer Studies* 18–19. 61–66.

Herbert, Robert K. 1986. *Language universals, markedness theory, and natural phonetic processes*. Berlin: Mouton de Gruyter.

Hockett, Charles F. 1955. *A manual of phonology*. Baltimore: Waverly Press.

Hollenbach, Barbara. 1977. Phonetic vs. phonemic correspondence in two Trique dialects. In W. R. Merrifield (ed.) *Studies in Otomanguean phonology*, 35–67. Dallas: Summer Institute of Linguistics.

Hume, Elizabeth. 2003. Language specific markedness: The case of place of articulation. *Studies in Phonetics, Phonology and Morphology* 9. 295–310.

Hume, Elizabeth & Keith Johnson (eds.) 2001a. *The role of speech perception in phonology*. San Diego: Academic Press.

Hume, Elizabeth & Keith Johnson. 2001b. A model of the interplay of speech perception and phonology. In Hume & Johnson (2001a), 3–26.

Hyman, Larry M. 1976. Phonologization. In Alphonse Juilland (ed.) *Linguistic studies offered to Joseph Greenberg*, vol. 2, 407–418. Saratoga, CA: Anma Libri.

Itô, Junko. 1986. Syllable theory in prosodic phonology. Ph.D. dissertation, University of Massachusetts, Amherst. Published 1988, New York: Garland.

Itô, Junko. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7. 217–259.

Jacobsen, William H., Jr. 1964. A grammar of the Washo language. Ph.D. dissertation, University of California, Berkeley.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Jassem, Wiktor & Lutoslawa Richter. 1989. Neutralization of voicing in Polish obstruents. *Journal of Phonetics* 17. 317–325.

Jun, Jongho. 1995. Perceptual and articulatory factors in place assimilation: An Optimality Theoretic approach. Ph.D. dissertation, University of California, Los Angeles.
Keyser, Samuel J. & Kenneth N. Stevens. 2001. Enhancement revisited. In Michael Kenstowicz (ed.) *Ken Hale: A life in language*, 271–291. Cambridge, MA: MIT Press.
Kim, Hyunsoon & Allard Jongman. 1996. Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics* 24. 295–312.
Kimenyi, Alexandre. 1979. *Studies in Kinyarwanda and Bantu phonology*. Carbondale: Linguistic Research Inc.
Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70. 419–454.
Kiparsky, Paul. 1976. Abstractness, opacity, and global rules. In Andreas Koutsoudas (ed.) *The application and ordering of grammatical rules*, 160–186. The Hague: Mouton.
Kiparsky, Paul. 1982. From Cyclic Phonology to Lexical Phonology. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part I, 131–175. Dordrecht: Foris.
Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.
Kirchner, Robert. 1997. Contrastiveness and faithfulness. *Phonology* 14. 83–111.
Labov, William. 1971. Methodology. In William O. Dingwall (ed.) *A survey of linguistic science*, 412–497. College Park: University of Maryland Linguistics Program.
Labov, William, Mark Karen & Corey Miller. 1991. Near-mergers and the suspension of phonemic contrast. *Language Variation and Change* 3. 33–74.
Labov, William, Malcah Yaeger & Richard Steiner. 1972. *A quantitative study of sound change in progress*. Philadelphia: US Regional Survey.
Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics*. Chicago: Chicago University Press.
Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John J. Ohala & Jeri J. Jaeger (eds.) *Experimental phonology*, 13–44. Orlando: Academic Press.
Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds.) *Speech production and speech modeling*, 403–439. Dordrecht: Kluwer.
Lombardi, Linda. 1991. Laryngeal features and laryngeal neutralization. Ph.D. dissertation, University of Massachusetts, Amherst.
Lombardi, Linda (ed.) 2001a. *Segmental phonology in Optimality Theory: Constraints and representations*. Cambridge: Cambridge University Press.
Lombardi, Linda. 2001b. Why Place and Voice are different: Constraint-specific alternations in Optimality Theory. In Lombardi (2001a), 13–45.
Martinet, André. 1952. Function, structure, and sound change. *Word* 8. 1–32.
Martinet, André. 1955. *Économie des changements phonétiques*. Berne: Francke.
Martinet, André. 1964. *Elements of general linguistics*. Chicago: University of Chicago Press.
McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.
McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.) *Phonological structure and phonetic form: Papers in laboratory phonology III*, 191–233. Cambridge: Cambridge University Press.
Milroy, James & John Harris. 1980. When is a merger not a merger? The MEAT/MATE problem in a present-day English vernacular. *English World-Wide* 1. 199–210.
Minkova, Donka. 2004. Philology, linguistics, and the history of [hw]–[w]. In Anne Curzan & Kimberly Emmons (eds.) *Studies in the history of the English language II: Unfolding conversations*, 7–46. Berlin & New York: Mouton de Gruyter.
Mohanan, K. P. 1982. Lexical Phonology. Ph.D. dissertation, MIT. Distributed by Indiana University Linguistics Club.
Myers, Scott & Benjamin B. Hansen. 2005. The origin of vowel-length neutralisation in vocoid sequences: Evidence from Finnish speakers. *Phonology* 22. 317–344.
Ní Chiosáin, Máire & Jaye Padgett. 2001. Markedness, segment realization, and locality in spreading. In Lombardi (2001a), 118–156.

Nunberg, Geoffrey. 1980. A falsely reported merger in eighteenth-century English: A study in diachronic variation. In William Labov (ed.) *Locating language in time and space*, 221–250. New York: Academic Press.

Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 258–275. Cambridge: Cambridge University Press.

Oostendorp, Marc van. 2008. Incomplete devoicing in formal phonology. *Lingua* 118. 1362–1374.

Padgett, Jaye. 2003. Contrast and post-velar fronting in Russian. *Natural Language and Linguistic Theory* 21. 39–87.

Pater, Joe. 1999. Austronesian nasal substitution and other NÇ effects. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*, 310–343. Cambridge University Press.

Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32. 693–703.

Port, Robert F. & Penny Crawford. 1989. Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17. 257–282.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Pulleyblank, Douglas. 1995. Feature geometry and underspecification. In Durand & Katamba (1995), 3–33.

Rialland, Annie. 1986. Schwa et syllabes en Français. In W. Leo Wetzels & Engin Sezer (eds.) *Studies in compensatory lengthening*, 187–226. Dordrecht: Foris.

Rudin, C. 1980. Phonetic evidence for a phonological rule: G-deletion in Turkish. *Research in Phonetics* 1. 217–232.

Rumsey, Alan. 2000. Bunuba. In R. M. W. Dixon & Barry J. Blake (eds.) *Handbook of Australian languages*, vol. 5, 35–152. Oxford: Oxford University Press.

Sapir, Edward. 1933. La réalité psychologique des phonèmes. *Journal de Psychologie Normale et Pathologique* 30. 247–265. English version published 1949 as The psychological reality of phonemes in David G. Mandelbaum (ed.) *Selected writings of Edward Sapir in language, culture, and personality*, 46–60. Berkeley: University of California Press.

Senn, Alfred. 1966. *Handbuch der litauischen Sprache*. Heidelberg: Carl Winter.

Silverman, Daniel. 1996. Phonology at the interface of phonetics and morphology: Root-final laryngeals in Chong, Korean, and Sanskrit. *Journal of East Asian Linguistics* 5. 301–322.

Silverman, Daniel. 2004. On the phonetic and cognitive nature of alveolar stop allophony in American English. *Cognitive Linguistics* 15. 69–93.

Silverman, Daniel. 2006. *A critical introduction to phonology: Of sound, mind, and body*. London & New York: Continuum.

Silverman, Daniel. 2010. Neutralization and anti-homophony in Korean. *Journal of Linguistics* 46. 453–482.

Simonet, Miquel, Marcos Rohena-Madrazo & Mercedes Paz. 2008. Preliminary evidence for incomplete neutralization of coda liquids in Puerto Rican Spanish. In Laura Colantoni & Jeffrey Steele (eds.) *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, 72–86. Somerville, MA: Cascadilla Press.

Smolensky, Paul. 1993. Harmony, markedness, and phonological activity. Paper presented at the Rutgers Optimality Workshop 1, Rutgers University (ROA-87).

Solé, Maria-Josep. 2007. Controlled and mechanical properties in speech: A review of the literature. In Maria-Josep Solé, Patrice Speeter Beddor & Manjari Ohala (eds.) *Experimental approaches to phonology*, 302–321. Oxford: Oxford University Press.

Steriade, Donca. 1994. Positional neutralization and the expression of contrast. Unpublished ms., University of California, Los Angeles.

Steriade, Donca. 1995. Underspecification and markedness. In Goldsmith (1995), 114–174.

Steriade, Donca. 1997. Phonetics in phonology: The case of laryngeal neutralization. Unpublished ms., University of California, Los Angeles.

Steriade, Donca. 2000. Paradigm uniformity and the phonetics–phonology boundary. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, 313–334. Cambridge: Cambridge University Press.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Hume & Johnson (2001a), 219–250.

Stevens, Kenneth N. & Sheila Blumstein. 1981. The search for invariant acoustic correlates of phonetic features. In Peter D. Eimas & Joanne L. Miller (eds.) *Perspectives on the study of speech*, 1–38. Mahwah, NJ: Lawrence Erlbaum.

Stevens, Kenneth N. & Samuel J. Keyser. 1989. Primary features and their enhancement in consonants. *Language* 65. 81–106.

Stevens, Kenneth N., Samuel J. Keyser & Haruko Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In Joseph S. Perkell & Dennis H. Klatt (eds.) *Invariance and variability in speech processes*, 426–449. Hillsdale, NJ: Lawrence Erlbaum.

Tatham, Mark A. A. 1971. Classifying allophones. *Language and Speech* 14. 140–145.

Teoh, Boon Seong. 1988. Aspects of Malay phonology revisited: A non-linear approach. Ph.D. dissertation, University of Illinois.

Traill, Anthony. 1985. *Phonetic and phonological studies of !Xóõ Bushman*. Hamburg: Buske.

Trouvain, Jürgen & William J. Barry (eds.) 2007. *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken: Saarland University.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.

Trudgill, Peter. 1974. *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Urbanczyk, Suzanne. 1996. Patterns of reduplication in Lushootseed. Ph.D. dissertation, University of Massachusetts, Amherst.

Wang, William S.-Y. & Charles Fillmore. 1961. Intrinsic cues and consonant perception. *Journal of Speech and Hearing Research* 4. 130–136.

Warner, Natasha, Allard Jongman, Joan A. Sereno & Rachèl Kemps. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32. 251–276.

Willerman, Raquel. 1994. The phonetics of pronouns: Articulatory bases of markedness. Ph.D. dissertation, University of Texas, Austin.

Yip, Moira. 1996. Lexical optimization in languages without alternations. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, vol. 2, 757–788. Salford: ESRI.

Yu, Alan C. L. 2004. Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80. 73–97.

Yu, Alan C. L. 2007a. Tonal phonetic analogy. In Trouvain & Barry (2007), 1749–1752.

Yu, Alan C. L. 2007b. Understanding near mergers: The case of morphological tone in Cantonese. *Phonology* 24. 187–214.

Zec, Draga. 1995. The role of moraic structure in the distribution of segments within syllables. In Durand & Katamba (1995), 149–179.

Zhang, Jie. 2001. The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis. Ph.D. dissertation, University of California, Los Angeles.

Zhang, Jie. 2002. *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis*. New York: Routledge.

81 Local Assimilation

Elizabeth C. Zsiga

1 Overview

Local assimilation is a phonological alternation in which two sounds that are adjacent become more similar. Its opposite is *dissimilation*, an alternation in which two sounds that are similar become more different (see CHAPTER 60: DISSIMILATION). Local assimilation can also be contrasted with long-distance assimilation (*harmony* – see also CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS; CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS; CHAPTER 78: NASAL HARMONY; CHAPTER 118: TURKISH VOWEL HARMONY; CHAPTER 123: HUNGARIAN VOWEL HARMONY), in which sounds that are not immediately string-adjacent influence one another, and with *coalescence* (see Casali 1996; Pater 1999), in which two adjacent sounds merge into a single segment that shares properties of both.

Local assimilation can be illustrated by different forms of the English negative prefix /ɪn-/, as in (1). Examples in (1a) and (1b) illustrate common place assimilation: the basic form of the nasal consonant is /n/ (1a), but it assimilates to the place of articulation of a following stop (1b). When the prefix precedes a labiodental fricative (1c), assimilation of /n/ is optional, as it is with other prefixes or across a word boundary (1d). In words like *illegal* and *irregular* (1e), the /n/ is not pronounced at all: in Latin, the /n/ became identical to a following /l/ or /r/. These different aspects of the /ɪn-/ alternation illustrate many of the questions and issues that arise in the cross-linguistic description of local assimilation.

(1) *An example of local assimilation in English*

a. *i*[n]*ability*
   i[nh]*ospitable*
   i[ns]*olvent*

b. *i*[mp]*ossible*
   i[mb]*alance*
   i[nt]*erminable*
   i[nd]*ecisive*
   i[ŋk]*ongruent*

c.	*i*[nf]*requent*	or	*i*[ɱf]*requent*
	i[nv]*ariant*	or	*i*[ɱv]*ariant*
d.	*u*[nb]*alanced*	or	*u*[mb]*alanced*
	i[np]*assing*	or	*i*[mp]*assing*
e.	*i*[l]*egal*	(Latin *il-legalis*)	
	i[r]*egular*	(Latin *ir-regularis*)	

If two adjacent sounds come to share just one feature, or a subset of their features, the assimilation is termed *partial*, illustrated by the word *impossible*, where the consonants in prefix and root share place of articulation, but not nasality or voicing. If two adjacent sounds become identical, as in *illegalis*, the assimilation is *total* or *complete*. It is also useful to distinguish the direction of assimilation. In a sequence of sounds AB, if A changes to become more like B, the assimilation is termed *anticipatory*: A anticipates some feature of B. If B changes to become more like A, the assimilation is *perseverative*: some feature of A continues into B. (Anticipatory assimilation may also be called *regressive*, since the assimilating feature is moving backwards, and perseverative assimilation may be termed *progressive*, since the assimilating feature is moving forward.) The assimilations in (1) are all anticipatory.

Local assimilation is the most common type of phonological alternation, and as such has played an important role in phonological theory. Phonological issues that arise with respect to local assimilation include the following:

(2) a. What features assimilate?
  b. What groups of features assimilate together?
  c. How can directional asymmetries be accounted for?
  d. What is the influence of morphological and prosodic context?
  e. What are the roles of production and perception in local assimilation?
  f. How is local assimilation different from co-articulation?
  g. How should local assimilation be formalized?

These questions concerning the nature and representation of local assimilation will be addressed in the remainder of this chapter, which is divided into two parts. §2 provides a cross-linguistic sampling of types of local assimilation, providing the data for more general theoretical discussion that follows in §3. Issues addressed in §3 are directionality and perception (§3.1), production and co-articulation (§3.2), and formalism (§3.3). In addition to these larger questions that focus on the linguistic status of assimilation per se, other more specific issues often come up in the discussion of particular datasets or types of assimilation. Various processes of local assimilation have been important in providing evidence for and against phonological issues such as underspecification, privative *vs.* binary features, feature geometry, and Lexical Phonology. Such connections will not be treated in depth in this chapter, but will be noted, along with cross-references to other chapters where the issue is addressed more fully.

2 Examples of local assimilation

Local assimilation can affect nearly every phonological feature. In fact, participating in assimilation is considered prime evidence for featural status (McCarthy

1994; Hume and Odden 1996; see also CHAPTER 17: DISTINCTIVE FEATURES). Some of the most common types of local assimilation are exemplified below.

2.1 *Voicing and other laryngeal features*

When obstruent consonants become adjacent, they often come to agree in voice, and sometimes in other laryngeal features as well. For example, in Russian, a string of obstruents always agrees in voicing with the rightmost obstruent in the sequence (Jakobson 1978; Padgett 2002).

(3) *Voicing assimilation in Russian*

a.	[ot papɨ]	'from papa'
	[od babuʃki]	'from grandma'
	[od vzbutʃki]	'from a scolding '
	[ot fspleska]	'from a splash'
b.	[ot mamɨ]	'from mama'

A similar alternation is found in Yiddish (Katz 1987, cited in Lombardi 1999), illustrated in (4).

(4) *Voicing assimilation in Yiddish*

[vog]	'weight'	[vok-ʃoi]	'scale'
[bak]	'cheek'	[bag-bejn]	'cheekbone'
but		[nud-nik]	'boring person'
		[mit-niten]	'co-respondent'

In both Russian and Yiddish, the assimilation is anticipatory: consonants anticipate the voicing of the rightmost obstruent in the cluster, whether voiced or voiceless. This is the unmarked direction for assimilation (Lombardi 1999). Perseverative voicing assimilation, in which the voicing value is determined by the leftmost consonant, may also be seen, usually in the case of assimilation of a suffix to a stem (Lombardi 1999; Borowsky 2000). Examples from English and Turkish are shown in (5).

(5) a. *Voicing assimilation in English (plural and past tense)*

[ro-z]	'rows'	[ro-d]	'rowed'
[ræg-z]	'rags'	[bɛg-d]	'begged'
[rak-s]	'rocks'	[kɪk-t]	'kicked'

b. *Voicing assimilation in Turkish* (Lewis 1967)

[git-tim]	'go-PAST.1SG'
[kɨz-dɨm]	'got mad-PAST.1SG'
[komsu-muz-dan]	'neighbor-POSS-ABL'
[raf-tan]	'shelf-ABL'

Cases of voicing assimilation have been central to the debate over whether [voice] is a privative or binary feature. Cases like that of Yiddish, where either a voiced or voiceless cluster may be formed, and where assimilation is independent of syllable-final devoicing, have been crucial. Is it possible to account for alternations such as [vog] ~ [vokʃoi] without reference to a feature [−voice]? See Cho (1999), Lombardi (1999), and Wetzels and Mascaro (2001) for further discussion.

Another important point to note about voicing assimilation in consonant clusters is that sonorant consonants are often neutral with respect to voicing assimilation (see CHAPTER 8: SONORANTS). Thus there are sequences like [ot fspleska] in Russian and [mitniten] in Yiddish, where the rightmost consonant in the cluster is a (voiced) sonorant, but the other consonants are voiceless. (Though note that the behavior of Russian /w/, which alternates with [v] and participates only partially in voicing assimilation, has been the subject of much discussion: see Padgett 2002 and references therein.)

On the other hand, the laryngeal features of obstruents and sonorants do sometimes interact. Common cases include intervocalic (or intersonorant) voicing (6) and post-nasal voicing (7).

(6) *Intersonorant voicing in Korean* (Silva 1992)

[pap]	'rice'	[i-bab-i]	'this rice-NOM'
[kuk]	'soup'	[i-gug-i]	'this soup-NOM'
[tal]	'moon'	[pan-dal]	'half moon'
[palp]	'walk'	[palb-in]	'that is walking'

/motun kilim/ → [modun gilim] 'every picture'
/kulimul pota/ → [kulimul boda] 'to look at a picture'

(7) a. *Post-nasal voicing in Yao* (Nurse and Phillipson 2003)

[ku-pélék-a]	'to send'	[kuu-m-bélek-a]	'to send me'
[ku-túm-á]	'to order'	[kuu-n-dúm-a]	'to order me'
[ku-kwéél-a]	'to climb'	[kuu-ŋ-gwéel-a]	'to climb on me'

b. *Post-nasal voicing in Puyo Pungo Quechua* (Orr 1962; Rice 1993; Pater 1999)

[sinik-pa]	'porcupine's'	[wasi-ta]	'the house'
[kam-ba]	'yours'	[wakin-da]	'the others'

It may also be the case that obstruents cause devoicing in sonorants, as in high vowel devoicing in Japanese (8a), in which /i/ and /u/ devoice when surrounded by voiceless consonants, or sonorant devoicing in English (8b), in which /l/ and /r/ devoice when preceded by a voiceless aspirated consonant.

(8) a. *High vowel devoicing in Japanese* (Tsuchida 1997)

[koku̥sai]	'international'
[ki̥tai]	'expectation'
[aki̥ko]	'woman's name'
[ɸu̥ton]	'bed'

b. *Sonorant devoicing in English*

[pl̥e]	'play'
[pr̥e]	'pray'
[tr̥u]	'true'
[kl̥e]	'clay'
[kr̥o]	'crow'

The interaction or non-interaction of sonorants and obstruents in voicing assimilation has played an important role in feature theory. On the one hand, the non-participation of sonorants has been cited as evidence that sonorants are underspecified for voice in underlying representation, with later fill-in by rule (Hayes 1984; Kiparsky 1985; Itô and Mester 1986). Alternatively, it has been argued that cases of sonorant/obstruent interactions involve features other than [voice] or [sonorant]. Rice (1993) argues that sonorants are specified with a different feature, [sonorant voice], which may spread to neighboring consonants, accounting for cases of intersonorant or post-nasal voicing. In the case of devoicing, as in (8), the assimilating feature may be aspiration: [spread glottis] rather than [−voice]. For further discussion, see CHAPTER 17: DISTINCTIVE FEATURES, CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION, and CHAPTER 8: SONORANTS.

Another approach has been to argue that cases of apparent voicing assimilations between vowels and consonants are not featural assimilation at all, but phonetic co-articulation. Jun (1995), for example, argues that Korean intersonorant voicing comes about because the glottal opening gesture for lax voiceless consonants is weak, allowing vocal fold vibration to continue throughout a short closure duration. Browman and Goldstein (1988) point out that the large and late glottal opening gesture for English initial aspirated stops is sufficient to delay voice onset in a following liquid, without further addition of a rule assimilating either [−voice] or [spread glottis]. §3 below returns to the issue of disentangling co-articulation and assimilation.

When languages have multiple laryngeal contrasts, examples of the assimilation of multiple laryngeal features have been identified. Smyth (1920) gives examples of assimilation of both aspiration and voicing in Ancient Greek.

(9) *Ancient Greek assimilation of both voicing and aspiration* (Smyth 1920)

a.	[grap^h-o]	'I write'
	[gegrap-tai]	'has been written'
	[grab-den]	'writing/scraping'
b.	[trib-o]	'I rub'
	[tetrip-tai]	'has been rubbed'
	[etrip^h-t^{h}eːn]	'it was rubbed'

Sanskrit also exhibits assimilation of multiple laryngeal features. The pattern of assimilation of voicing and aspiration in Sanskrit is complex, and its description and analysis has a long history (Whitney 1889; Wackernagel 1896). The examples in (10) represent part of this interaction, and serve to illustrate local assimilation of voicing and aspiration from the coda of the verb root to the onset of the suffix.

(10) *Assimilation of voicing and aspiration in Sanskrit* (Calabrese and Keyser 2006)

/b^{h}aud^h-ta/ →	[budd^{h}a]	'awake-PST PART'
/rud^h-ta/ →	[rudd^{h}a]	'obstruct-PST PART'
/saːd^h-ta/ →	[saːdd^{h}a]	'succeed-PST PART'

Cases of simultaneous assimilation of more than one feature, such as those in (9) and (10), have been important in providing evidence for hierarchical organization of features. See §3.3 below, and CHAPTER 27: THE ORGANIZATION OF FEATURES.

2.2 Nasality

Assimilation of nasality is very common. Vowels generally become nasalized when adjacent to a nasal consonant, as illustrated in (11). Such nasalization may be anticipatory, as in English (11a), or perseverative, as in Sundanese (11b). Sundanese nasalization can also be iterative and in some cases long-distance, applying across an intervening /h/: see Cohn (1993) and CHAPTER 78: NASAL HARMONY.

(11) a. *Anticipatory nasalization in English*

[kʰæt]	'cat'	[kʰæ̃n]	'can'
[ɹɪb]	'rib'	[ɹɪ̃n]	'rim'
[θɪk]	'thick'	[θɪ̃ŋ]	'thing'

b. *Perseverative nasalization in Sundanese* (Cohn 1993)

[ŋãtur]	'arrange'
[mãrˈɪɑs]	'examine'
[ɲĩãr]	'seek'
[mãhãl]	'expensive'

Cohn (1993) argues that, in addition to differing in direction, English and Sundanese represent two distinct types of assimilation: the one categorical and phonological (Sundanese), the other gradient and phonetic (English). §3.2 below returns to this distinction.

Assimilation of nasality may also apply between adjacent consonants, as shown in (12).

(12) a. *Nasal assimilation from onset to coda in Korean* (Kim-Renaud 1991)

[pap]	'rice'	[pam mekta]	'eat rice'
[ot]	'clothes'	[on man]	'only clothes'
[jak]	'medicine'	[jaŋ mekta]	'take medicine'

b. *Nasal assimilation from prefix to root in Twi*

[bá]	'comes'	[m-má]	'does not come'
[gu]	'pours'	[ŋ-ŋu]	'does not pour'
cf.			
[pɛ]	'likes'	[m-pɛ]	'does not like'
[tɔ]	'does'	[n-tɔ]	'does not do'

2.3 Continuant

Stops often become continuants when surrounded by, or in some cases just preceded by, continuants. The change from stop to fricative, termed *spirantization*, may be considered assimilation of the feature [continuant] (see CHAPTER 28: THE REPRESENTATION OF FRICATIVES). Examples from Spanish and Italian are shown in (13).

(13) a. *Post-continuant spirantization of voiced stops in Spanish*

/la gata/	→	[la ɣata]	'the (FEM) cat'
/la data/	→	[la ðata]	'the date'
/la bola/	→	[la βola]	'the ball'
/las gata/	→	[las ɣatas]	'the (FEM) cats'
/las bolas/	→	[las βolas]	'the balls'

b. *Intervocalic spirantization of voiceless stops in Florentine Italia* (Villafana 2006)

/la kaza/	→	[la xaza]	'the house'
/la torta/	→	[la θorta]	'the cake'
/la palːa/	→	[la ɸalːa]	'the ball'

Similarly, continuants often "harden" to stops or affricates in post-nasal position, an alternation that may be considered assimilation of [−continuant] from the preceding nasal (Padgett 1994).

(14) a. *Post-nasal hardening in Setswana* (Tlale 2006)

[supa]	'point at'	[n-tsʰupa]	'point at me'
[ʃapa]	'hit'	[ɲ-tʃʰapa]	'hit me'
[xapa]	'capture'	[ŋ-kxʰapa]	'capture me'
[rut'a]	'teach'	[n-tʰut'a]	'teach me'

b. *Post-nasal hardening in Kikuyu* (Armstrong 1967; Clements 1985)

imperative	*1sg imperfect*	
[βur-a]	[m-bur-eetɛ]	'lop off'
[reh-a]	[n-deh-eetɛ]	'pay'
[ɣor-a]	[ŋ-gor-eetɛ]	'buy'

Spirantization and hardening are not necessarily considered to be cases of assimilation, however, but cases of a separate phonological process of *lenition* or *fortition*, in which features other than [continuant] may be involved. Spanish stops may weaken to more open approximant articulations (/la bola/ → [la υola]) and intervocalic /k/ in Florentine often weakens to [h] (/la kaza/ → [la haza]). Conversely, post-nasal fortition in Setswana involves changes in laryngeal features as well as in continuant. See Kirchner (1998), Lavoie (2001), Gurevich (2003), and CHAPTER 66: LENITION for numerous further examples and discussion.

2.4 Consonantal place of articulation

2.4.1 Nasal place assimilation

Assimilation of place of articulation is probably the most ubiquitous phonological alternation. Especially common is nasal place assimilation: nasals assimilate in place of articulation to a following consonant. Examples could be found in almost any language. Nasal place assimilation in English and in the African languages Yao, Twi, Setswana, and Kikuyu was seen in examples (1), (7a), (12b), (14a), and (14b) above. Additional examples are shown in (15): Catalan (a), Zoque (b), Malayalam (c), Sri Lankan Creole (d), and Zulu (e). Zulu is included to illustrate the point that in place assimilation to complex segments such as clicks and labial-velars, assimilation to the dorsal place of articulation is most common (Maddieson and Ladefoged 1989; see also CHAPTER 18: THE REPRESENTATION OF CLICKS).

(15) a. *Nasal place assimilation in Catalan* (Mascaró 1976; Kiparsky 1985)

so[n]*amics* 'they are friends'
so[m]*pocs* 'they are few'
so[ɱ]*felicos* 'they are happy'
so[n̪ d̪]*os* 'they are two'
so[ɳ]*rics* 'they are rich'
so[ɲ lʲ]*iures* 'they are free'
so[ŋ]*grans* 'they are big'

b. *Nasal place assimilation in Zoque* (Wonderly 1946, cited in Padgett 1994)

[pama]	'clothing'	[m-bama]	'my clothing'
[tatah]	'father'	[n-datah]	'my father'
[ʧima]	'calabash'	[ɲ-ʤima]	'my calabash'
[kaju]	'horse'	[ŋ-gaju]	'my horse'
[gaju]	'rooster'	[ŋ-gaju]	'my rooster'
cf. nasal deletion preceding fricatives:			
[faha]	'belt'	[faha]	'my belt'
[ʃapun]	'soap'	[ʃapun]	'my soap'
[ranʧo]	'ranch'	[ranʧo]	'my ranch'

c. *Nasal place assimilation in Malayalam* (Mohanan 1993)

[awan]	'he'
[awam-paraɲɲu]	'he said'
[awan̪-t̪aʈiccu]	'he became fat'
[awaɲ-caːʈi]	'he jumped'
[awaŋ-karaɲɲu]	'he cried'
[kamalam]	(proper name)
[kamalam-paraɲɲu]	'Kamalam said'
[kamalan̪-t̪aʈiccu]	'Kamalam became fat'
[kamalaɲ-caːʈi]	'Kamalam jumped'
[kamalaŋ-karaɲɲu]	'Kamalam cried'

d. *Assimilation of non-coronal nasals in Sri Lankan Portuguese Creole* (Hume and Tserdanelis 2002; Hume 2003)

Nom sg	*Gen sg*	*Dat sg*	*Verbal noun*	
[maːm]	[maːn-su]	[maːm-pə]	[maːŋ-ki]	'hand'
[miːtiŋ]	[miːtin-su]	[mitim-pə]	[miːtiŋ-ki]	'meeting'
[siːn]	[siːn-su]	[siːn-pə]	[siːn-ki]	'bell'

e. *Assimilation to the dorsal place of clicks in Zulu* (Doke 1926, cited in Padgett 2002)

/iziN-/	(class 10 plural prefix)
[izim-papʰɛ]	'feathers'
[izin-ti]	'sticks'
[iziŋ-kɛzɔ]	'spoons'
[iziŋ-ǀezu]	'slices'
[iziŋ-ǂuŋǂulu]	'species of bird (PL)'
[iziŋ-ǁaŋǁa]	'green frogs'

A number of cross-linguistic differences and similarities in nasal place assimilation are worth noting.

As in the Zoque and Zulu examples, it is often the case that a nasal (or nasal-final) affix undergoes obligatory place assimilation in every lexical item in which it occurs. In such cases, it may be impossible to determine empirically the basic or underlying form, and it is often argued that such nasals are unspecified for place (e.g. Kiparsky 1985). A segment specified only as [nasal], but with no underlying place features, may be symbolized /N/. Depending on how a particular alternation is formalized, however, underspecification may or may not be assumed. (See CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION.)

Relevant to the debate on underspecification is the observation that the coronal nasal assimilates more often than either labial or velar nasals. In many languages, such as Catalan and Spanish (Navarro Tomás 1970; Honorof 1999), only the coronal nasal assimilates, although for some languages such as Malayalam, assimilation of non-coronal nasals is also attested. In Sri Lankan Creole, all nasals *except* [n] assimilate. Asymmetries in place assimilation are discussed further in §3.1 and §3.2 below.

Another point of interest in nasal place assimilation is whether or not nasals assimilate to [+continuant] segments. In Catalan and Sri Lankan, nasals assimilate to both stops and continuants, but in Malayalam and Zoque, nasals assimilate only to stops. In Malayalam, unassimilated nasal–fricative clusters are tolerated, but in Zoque the nasal deletes when a fricative follows. In other languages, other processes may apply to repair disfavored nasal–fricative clusters. In Setswana and Kikuyu ((14) above), fricatives harden to stops or affricates in post-nasal position. In English, as was noted in (1), assimilation of /n/ to /f/ is optional: the nasal–fricative cluster in the word *infrequent* may be pronounced [nf] in careful speech or [ɱf] in less careful speech, but *impolite* is invariably [mp]. The propensity for place assimilation and continuant assimilation to occur together leads Padgett (1994), among others, to posit a dependency relation between features for place and the feature [continuant], though this requires a different explanation for cases like Spanish *so*[ɱ f]*elicos* and English *i*[ɱf]*requent* (see the discussion in §3.2 below).

Data on nasal place assimilation, probably more because it is so common than because of any inherent phonological property, has often been invoked in debates on domains of application. Catalan nasal place assimilation played an important role in arguments for cyclic rule application, and the distinction between lexical and post-lexical phonology (Kiparsky 1985). The observation that place assimilation applies to English /m-/ *impossible*, /kan-/ *congruent*, and /sɪn-/ *sympathy*, but not /ʌn-/ *unprepared*, provided important data for level ordering of affixes in English. The ability of nasal place assimilation to create sounds that are not part of the underlying inventory of the language, such as [ɱ] in English and Catalan, has informed debate on Structure Preservation, and on the lexical/post-lexical distinction (Kiparsky 1985; CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME).

Finally, data on nasal place assimilation has also been crucial in the theory of feature geometry (discussed in §3.3 below), and in development of the theory of Articulatory Phonology (discussed in §3.2).

2.4.2 *Other consonantal place assimilations*

Place assimilation most often involves nasals, but other consonants undergo place assimilation as well. In Korean, for example, optional place assimilation applies to certain obstruent clusters, as illustrated in (16): final [t] may assimilate to a following labial or dorsal stop, and [p] to a following dorsal.

(16) *Place assimilation in Korean obstruents*
(Kim-Renaud 1991; Kochetov and Pouplier 2008)

/pat^{h}-pota/	→	[pat p'oda]	or	[pap p'oda]	'rather than the field'
/pat^{h}-kwa/	→	[pat k'wa]	or	[pak k'wa]	'field and'
/pap-kɨlɨs/	→	[pap k'ɨrɨt˺]	or	[pak k'ɨrɨt]	'rice bowl'
/t^{h}op-k^{h}al/	→	[t^{h}op k^{h}al]	or	[t^{h}ok k^{h}al]	'handsaw'

cf.

/pap-to/	→	[papt'o]	*[patt'o]	'rice also'
/pak-to/	→	[pakt'o]	*[patt'o]	'outside also'
/kuk-po/	→	[kukp'o]	*[kupp'o]	'national treasury'

In other cases, subsidiary place features assimilate between adjacent consonants. Most often the features [anterior] and [distributed] assimilate in sequences of coronal consonants. It was seen above that Catalan nasals (15a) assimilate to a following consonant at all places of articulation. Catalan laterals also assimilate (17), but only to a following coronal. (See CHAPTER 31: LATERAL CONSONANTS for further discussion.)

(17) *Assimilation of laterals in Catalan* (Mascaró 1976)

e[l p]*a*	'the bread'
e[l̪ d̪]*ia*	'the day'
e[ɭ r]*ic*	'the rich'
e[ʎ ʒ]*erma*	'the brother'

In English, coronal stops and nasals assimilate the [–anterior] feature of a following [ɹ], or the dental articulation of a following dental fricative.

(18) *Assimilation of retroflex and dental in English*

train	[t̠ɹeɪn]
drain	[d̠ɹeɪn]
tenth	[t^{h}ɛn̪θ]
eighth	[eɪt̪θ]
width	[wɪd̪θ]

In Sanskrit, Murinbata, and other languages of India and Australia (Steriade 2001), place assimilation among coronal clusters is often perseverative: that is, the onset assimilates to the coda, as shown in (19). This reversal in expected direction is discussed further in §3.1 below.

(19) a. *Perseverative retroflex assimilation in Sanskrit* (Whitney 1889, cited in Steriade 2001)

/iʂ-ta/	→	[iʂ-ʈa]	'sacrificed'
/ʂaɳ-nam/	→	[ʂaɳ-ɳam]	'of six'
/giɽ-su/	→	[giɽ-ʂu]	'in songs'

b. *Perseverative retroflex assimilation in Murinbata* (Street and Mollinjin 1981, cited in Steriade 2001)

/pan-ʈal/	→	[pan-tal]	'cut it-3SG'
/ŋudu-ɭɛɭ-nu/	→	[ŋudu-ɭɛɭ-ɳu]	'roll-FUT'

As was noted with respect to nasal place assimilation, evidence from consonantal place assimilation in general has been crucial in the development of phonological theory. One point of particular interest is how place assimilation, which may involve a whole set or subset of different features, may be formalized as a unitary process (see §3.3 below). Another point is the problem of directionality (§3.1): why is it that codas usually assimilate to onsets, rather than vice versa? Place assimilation has also played an important role in the debate over assimilation *vs.* co-articulation, and in the development of the theory of Articulatory Phonology (§3.2).

2.4.3 *Vowel and consonant place interactions*

While place assimilation usually applies to consonant clusters, vowels and consonants may also assimilate to each other (see CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS for more discussion). Consonants often assimilate the properties of adjacent vocalic articulations. For example, in the Wakashan language Oowekyala (Howe 2000), velar and uvular consonants contrast in rounding in initial position and following most vowels (20a). Immediately following /u/, however, all velars and uvulars assimilate to the vowel's [round] feature (20b).

(20) *Oowekyala* (Howe 2000)

a. *Contrastive rounding*

[q^{w}ut'a]	'full'
[quɬa]	'bent'

b. *Rounding assimilation*

[pusq'a-x^{ʔ}it]	'to become very hungry'
[ƛ'u'xwalasu-x^{wʔ}it]	'to become sick'
[məja-gila]	'make (draw or carve) a fiish'
[ʔamastu-g^{w}ila]	'make kindling'

In the neighboring language Nuxalk, Howe reports that rounding assimilation is anticipatory rather than perseverative: velars and uvulars become round *preceding* /u/.

Another assimilation from vowel to consonant is palatalization of a consonant adjacent to front vowels and glides. Palatalization may take the form of an alveolar or dental (or sometimes velar) becoming alveopalatal, or it may take the form of secondary articulation, adding an additional high front tongue position without changing the consonant's primary place of articulation. Languages differ in the input sequences that trigger palatalization, and in the resulting outputs. Three examples are shown in (21); see CHAPTER 71: PALATALIZATION for further examples and discussion. In English (21a), alveolars become alveopalatals before /j/. In Japanese (21b), alveolars become alveopalatals before /i/, while velar and labial consonants become secondarily palatalized. In one pattern of palatalization in Polish (21c), labials receive secondary palatalization before [i] and [e], while velars and alveolars change their primary place. (Palatalization in Slavic languages is both common and complex: see CHAPTER 121: SLAVIC PALATALIZATION and CHAPTER 122: SLAVIC YERS.)

(21) a. *English palatalization before /j/*

[d ~ ʤ]	*grade*	*gradual*
[t ~ ʧ]	*habit*	*habitual*
[s ~ ʃ]	*press*	*pressure*
[z ~ ʒ]	*use*	*usual*

b. *Palatalization in Japanese* (Vance 1987; Chen 1996)

[kas-anai]	'lend-NEG'	[kaʃ-ita]	'lend-PAST'
[kat-anai]	'win-NEG'	[kaʧ-itai]	'win-VOLITIONAL'
[wak-anai]	'boil-NEG'	[wakʲ-itai]	'boil-VOLITIONAL'
[job-anai]	'call-NEG'	[jobʲ-itai]	'call-VOLITIONAL'

c. *Palatalization in Polish* (Szpyra 1989; Chen 1996)

[łupɨ]	'booty'	[łupʲ-ic]	'to rob'
[zabava]	'game'	[zabavʲ-ic]	'to entertain'
[zloto]	'gold'	[złoc-ic]	'to gild'
[kwas]	'acid'	[kwaç-ic]	'to make sour'
[rana]	'wound'	[raɲ-ic]	'to wound'
[sok]	'juice'	[soʧ-ek]	'juice-DIM'
[mex]	'moss'	[meʃ-ek]	'moss-DIM'

Vowels may also assimilate to adjacent consonants. In Russian (22a), palatalization on consonants is contrastive, and the vowels [i] and [ɨ] are in complementary distribution: [i] is found in initial position and following palatalized consonants, [ɨ] follows non-palatalized consonants. (See CHAPTER 121: SLAVIC PALATALIZATION for additional discussion.) In the Dravidian language Tulu (22b), the accusative suffix /-ɨ/ becomes round when it follows a labial consonant (or another round vowel).

(22) a. *Vowel backing in Russian* (Halle 1959; Padgett 2002)

[ivan]	(proper name)
[k-ɨvanu]	'to Ivan'
[italʲia]	'Italy'
[v-ɨtalʲiju]	'to Italy'

b. *Rounding assimilation in Tulu* (Bright 1957, cited in Kenstowicz 1994)

[katt-ɨ]	'bond-ACC'
[kapp-u]	'blackness-ACC'
[ucc-u]	'snake-ACC'

Another type of assimilation from consonant to vowel is vowel lowering. Vowel lowering after uvular, pharyngeal, and laryngeal consonants (the class of *guttural* consonants) is found in many Semitic, Caucasian, and North American languages (Herzallah 1990; Bessell 1992, 1998; McCarthy 1994; Rose 1996). In Syrian Arabic (23a), for example, the feminine suffix is realized as [a] after laryngeals, pharyngeals, and uvulars, and as [e] after all other consonants. In Oowekyala (23b), /i/ and /u/ are lowered to [e] and [o].

(23) a. *Vowel lowering in Syrian Arabic* (Cowell 1964; Rose 1996)

[daraʒ–e]	'step'
[ʃerk-e]	'society'
[madras-e]	'school'
[waːʒh-a]	'display'
[mniːħ-a]	'good'
[daggaːʀ-a]	'tanning'

b. *Vowel lowering in Oowekyala* (Howe 2000)

/ƛiq-ila/	→	[dliqχela]	'to give a name to someone'
/qusa/	→	[qosa]	'bent, crooked'
/hula/	→	[hola]	'heap up'
/ɢwiχila/	→	[ɢwiχela]	'to bake bread'

Vowel lowering has figured prominently in the debates over what features constitute the class of guttural consonants. In various cases the feature has been argued to be [low], [−high], or [pharyngeal]. See CHAPTER 25: PHARYNGEALS for further examples and discussion.

In general, the question of when and how vowels and consonants interact with each other has been important in the area of feature theory. Consonants are often transparent to long-distance vowel-to-vowel assimilation, yet they also interact with vowels in local assimilations, as has been illustrated. Transparency to vocalic alternations suggests that vowels and consonants have different features, or that consonants bear vocalic features only as secondary articulations: e.g. front vowels and palatalized consonants are [−back], while alveolar and dental consonants are [coronal]. Such an approach accounts for cases in which processes of palatalization and rounding result in secondary articulations. But it fails to account for cases where the vowel causes a change in the consonant's primary place of articulation, or where the consonant causes a change in the backness of a vowel. Such alternations have led to proposals (e.g. Clements 1993; Hume 1994; Clements and Hume 1995) that vowels and consonants share the same features: e.g. alveolar consonants, alveopalatal consonants, and front vowels are all [coronal]. See CHAPTER 19: VOWEL PLACE, CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION, and CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS.

2.5 *Complete assimilations*

Complete assimilation occurs when two adjacent sounds become identical. Complete assimilation is particularly common in clusters involving /r/ and /l/ (see CHAPTER 30: THE REPRESENTATION OF RHOTICS and CHAPTER 31: LATERAL CONSONANTS). Complete assimilation of the Latin prefix /in-/ to both /l/ and /r/ as in *illegalis* and *irregularis* was seen in (1) above. Similar cases are found in Ponapean and Korean.

(24) a. *Assimilation of /n/ to /l/ and /r/ in Ponapean* (Rehg and Sohl 1981; Rice 1993)

/nanras/	→	[narras]	'ground level of a feasthouse'
/nanleŋ/	→	[nalleŋ]	'heaven'
/pahn liŋan/	→	[pahl liŋan]	'will be beautiful'
/pahn roŋ/	→	[pahr roŋ]	'will listen'

b. *Assimilation of /n/ to /l/ in Korean* (Davis and Shin 1999)

/non-li/	→	[nolli]	'logic'
/tan-lan/	→	[tallan]	'happiness'
/ts^{h}ən-li/	→	[ts^{h}əlli]	'natural law'

In some cases, a sonorant may assimilate completely to a following obstruent. In Arabic (25a), /l/ assimilates to a following coronal, but not to consonants at other places of articulation. In Havana Spanish (25b), the /l/ assimilates completely to most following consonants. The exception is that if the following consonant is a voiceless stop, the /l/ assimilates in all features except [voice].

(25) a. *Assimilation of /l/ in Arabic* (Kenstowicz 1994)

[ʔaʃ-ʃams]	'the sun'
[ʔad-daar]	'the house'
[ʔan-nahr]	'the river'
[ʔaz-zajt]	'the oil'
cf.	
[ʔal-qamr]	'the moon'
[ʔal-kitaab]	'the book'
[ʔal-faras]	'the mare'

b. *Assimilation of /l/ in Havana Spanish* (Harris 1985)

albañil	*a*[bb]*añil*	'mason'
tal droga	*ta*[dd]*roga*	'such a drug'
pulga	*pu*[gg]*a*	'flea'
tal mata	*ta*[mm]*ata*	'such a shrub'
el fino	*e*[ff]*ino*	'the refined one'
el pobre	*e*[bp]*obre*	'the poor man'
el tres	*e*[dt]*res*	'the three'

A similar case of near-complete assimilation occurs in Kannada (26). The final consonant of the morpheme meaning "big" copies all features from the following consonant, except that the resulting cluster must be voiced, regardless of input.

(26) *Complete assimilation with voicing in Kannada* (Roca and Johnson 1999)

[tere]	'screen'	[hed-dere]	'big screen'
[kumbaḷa]	'pumpkin'	[heg-gumbaḷa]	'big pumpkin'
[dzenu]	'bee'	[hedz-dzenu]	'big bee'
[mara]	'tree'	[hem-mara]	'big tree'

Finally, complete local assimilation of one vowel to another can also be found. Many languages will not tolerate successive non-identical vowels (CHAPTER 61: HIATUS RESOLUTION). While vowel hiatus is often repaired by deleting one vowel or the other (see Casali 1996, 1997), another strategy is assimilation, as shown in (27).

(27) *Vowel assimilation in Yoruba* (Welmers 1973)

[owo]	'money'
[owe-epo]	'oil money'
[owa-ade]	'Ade's money'

2.6 *Instances where local assimilation doesn't apply*

The preceding list of types of local assimilation has been long. Nonetheless, there are situations where local assimilation is not typically found. These include, on the one hand, environments where the trigger and target of assimilation tend not to be immediately string-adjacent, as in tone and vowel harmony. On the other hand, there are features for which languages prefer an alternating pattern, such as CVC within a syllable, or stress–unstress within a foot.

Features such as [round], [back], and [advanced tongue root] often assimilate from vowel to vowel within a word, but such assimilation is usually not local at the level of the segment, since vowels are most often separated by consonants (see CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS; CHAPTER 118: TURKISH VOWEL HARMONY; CHAPTER 123: HUNGARIAN VOWEL HARMONY). Similarly, tone assimilations are quite common, and have played an important role in the development of theories of phonological representation. Tone assimilation, however, is also generally a long-distance phenomenon, applying at least from vowel to vowel across intervening consonants, and often across stretches of multiple syllables (see CHAPTER 114: BANTU TONE; CHAPTER 45: THE REPRESENTATION OF TONE; CHAPTER 107: CHINESE TONE SANDHI). Because this chapter focuses on local processes, vowel harmony and tone assimilation are not treated further here.

It has been argued that the features [consonantal] and [sonorant] do not assimilate (McCarthy 1988). Consonants do not become vowels when adjacent to vowels, and vice versa (but see Kaisse 1992 for a possible counterexample). Although consonant clusters may come to agree in sonorancy as a result of nasal assimilation or complete assimilation, the feature [sonorant] does not assimilate independently. But see Rice (1993) for discussion of a feature [sonorant voice], which is proposed to distinguish sonorants from obstruents, and to be active in cases of sonorant/obstruent voicing interactions. See also CHAPTER 8: SONORANTS and CHAPTER 13: THE STRICTURE FEATURES.

Length does not assimilate: if anything, lengthening of one segment will induce shortening of neighboring segments, or vice versa (see CHAPTER 64: COMPENSATORY LENGTHENING). Stress does not assimilate. If two stressed syllables become adjacent, languages will often resolve the "clash" by moving or deleting a stress to restore the alternating pattern. (See CHAPTER 41: THE REPRESENTATION OF WORD STRESS.)

3 General phonological issues in local assimilation

§2 provided examples of the most common kinds of local assimilations, and pointed out theoretical issues raised by specific cases, such as privativity of the feature [voice] and the featural description of the class of guttural consonants. §3 now turns to broader questions, which are applicable to many or all kinds of local assimilation. These include directionality and perception (§3.1), the relation between assimilation and co-articulation (§3.2), and the formal treatment of local assimilation (§3.3).

3.1 *Directionality and perception*

In nearly every case discussed above, there has been a preference in the directionality of assimilation. The following principles can be deduced:

(28) a. Assimilation in consonant clusters tends to be anticipatory: the specification of the rightmost consonant dominates.
b. Codas assimilate to onsets, rather than vice versa.
c. Affixes assimilate to stems and roots, rather than vice versa.

Many phonologists have analyzed these asymmetries in structural terms. Itô (1998), for example, proposes that onsets "license" place features, and that in many languages codas can only acquire place features by sharing them with an onset consonant, thus forcing assimilation. Lombardi (1999) proposes a similar argument for laryngeal features. Beckman (1998) extends the positional analysis in proposing a theory of "positional faithfulness": certain structural positions, including onset of a syllable or word, are privileged, and changes to these privileged positions are dispreferred. Stems and roots are privileged over affixes, thus affixes tend to assimilate to stems rather than vice versa. Hyman (2008) proposes a structural account of directional asymmetries in a number of Bantu languages.

Other linguists, however, argue that asymmetries in direction of assimilation can be explained by asymmetries in perceptibility, without reference to structural positions. Steriade (2001), for example, emphasizes that consonantal place of articulation is most clearly cued by the formant transitions and burst noise that occur when a closure is released into a vowel. Steriade argues that codas most often assimilate to onsets because the phonological features of a postvocalic stop are less clearly perceived than the features of a prevocalic stop, and thus a change to the coda consonant is less obvious. In cases where a particular distinction is better cued in coda position, the direction of assimilation is reversed: there is perseverative assimilation of retroflexion in Sanskrit and Murinbata consonant clusters ((19) above), because retroflexion is best cued by formant transitions on the *preceding* vowel.

Similar arguments from perception can be applied to explain why nasals and coronals so often undergo assimilation. Nasals may be especially prone to assimilate, because nasal resonances interfere with the formant information that conveys place of articulation. Coronals may more frequently assimilate because cues to coronal place of articulation are weakest, and may be overwhelmed by the stronger cues from a following stop at a different place (Kawasaki 1982; Byrd 1992). Cho and McQueen (2008) and Sohn (2008) offer perceptual accounts of Korean place assimilation. See also Paradis and Prunet (1991) and Hume (2003) for arguments respectively for and against general coronal unmarkedness, and CHAPTER 12: CORONALS.

Integral to the discussion of perception in local assimilation is the role of *mis*-perception. A speaker may produce a word or phrase in a way that is faithful to the lexical representation, but if perceptual cues to a particular contrast in a particular position are weak or non-existent, a listener may perceive something different. That is, a speaker may say [np], but the listener may hear [mp]. If the listener assumes [mp] was the intended pronunciation of /np/, the listener may postulate a phonological alternation. For further discussion see Ohala (1981), Hume and Johnson (2001), and CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY.

3.2 Assimilation and co-articulation

Processes of local assimilation are "natural," in the sense the word is used in the theory of Natural Phonology (Donegan and Stampe 1979); that is, the phonetic motivation for such processes is clear, and the motivation works in the direction of making speaking easier. While concepts like "ease of articulation" and "articulatory effort" are difficult to quantify (see Lindblom 1983; Kirchner 1998, 2000), local assimilation has an obvious phonetic basis in co-articulation.

The term *co-articulation* describes the influence segments have on one another simply by being adjacent, apart from any featural change. Because articulators cannot change position instantly, there is necessarily either some anticipatory or perseverative effect, if not both, on neighboring segments, as articulators move from one target to the next. Two examples illustrate the point. If the velum is to be fully open by the time a consonant closure is achieved, then opening must begin during the preceding vowel, resulting in some inevitable nasal resonance during the vocalic portion. If the tongue body is to reach its target vowel position by the time the onset consonant in a CV syllable is released, articulation of vowel and consonant must begin simultaneously. Thus a [k] is made further forward in the mouth when it precedes a front vowel.

Some articulatory overlap is inevitable, but degree and direction of co-articulation will differ from language to language. Given that language-specific patterns of co-articulation must be learned as part of the grammar, some linguists have argued that there is no need to state independent phonological rules of nasalization, rounding, palatalization, or place assimilation. In particular, the theory of Articulatory Phonology (Browman and Goldstein 1992) argues that all productive phonological changes can be accounted for in terms of differences in articulatory organization, particularly gestural overlap and reduction, without invoking any phonological feature change (see CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATION). Browman and Goldstein (1990), using X-ray microbeam data, show that a coronal closing gesture is still present in English phrases which sound as though a coronal nasal had become labial: for example in the phrase *seven plus* heard as assimilated *se*[vmp]*lus*. They argue that the [n] is not deleted or changed from [coronal] to [labial], but is overlapped by the following [p], according to the general pattern of consonant coordination at word boundaries in English. The [n] and [p] articulated together sound like [m] (see also Byrd 1992). Browman and Goldstein further argue that place assimilation in *tenth* ((18) above) is also the result of overlap and blending: the tongue tip cannot be both dental and alveolar at the same time, so a compromise blended position is reached. Zsiga (1995) argues for an overlap account of palatalization at word boundaries in English. The phrase *this year* may sound like *thish year*, but data from electropalatography shows that the word-final fricative is not identical to an underlying [ʃ]. Rather, it is the acoustic result of an [s] and [j] articulated at the same time, with tongue tip and blade gestures blended together. Some proponents of Articulatory Phonology incorporate gestural dynamics into constraint-based theory (Gafos 2002; Bradley 2007).

It is not clear, however, whether all local assimilations are best described in terms of gestural overlap. One distinction that is often made is that categorical phonological alternations should be represented as the result of a change in featural specification, while partial and gradient changes are attributed to gestural overlap (see CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY). Thus Cohn (1993), for example, identifies two different kinds of nasalization in English and Sundanese. Using nasal and oral airflow data, Cohn demonstrates that nasalization of a vowel in English is partial and gradient, due to co-articulation with the opening velum, and very much dependent on timing and context. In contrast, nasalization in Sundanese is categorical: a nasalized vowel must be specified with its own featural target. In a similar vein, Zsiga (1995) argues that palatalization at word boundaries in English is the gradient result of overlap, while palatalization at morpheme boundaries ((21a) above) is the categorical result of a featural change. Ladd and Scobbie (2003: 16) provide data that vowel assimilation at word boundaries in Sardinian is categorical, and conclude

> that gestural overlap is on the whole not a suitable model of most of the assimilatory external sandhi phenomena in Sardinian, and more generally that accounts of gestural overlap in some cases of English external sandhi cannot be carried over into all aspects of post-lexical phonology.

Other researchers, however, follow Browman (1995) in arguing that apparently categorical deletions and assimilations are just the endpoints of a gradient distribution: deletion being the limiting case of reduction and categorical assimilation the limiting case of overlap. Thus Kochetov and Pouplier (2008), for example, describe the categorical change of /pk/ → [kk] in Korean ((16) above), in which they show the assimilated sequence to be identical to an underlying /kk/ cluster, as full reduction of the lip closing gesture and temporal extension of the velar closing gesture. One crucial question is whether there is a theory of gestural timing and organization that is both powerful enough to account for gradient changes, and constrained enough to account for changes that result in category neutralization (see the discussion in Zsiga 1997; Ladd and Scobbie 2003; Scobbie 2007). Another challenge lies in integrating articulatory and perceptual approaches. Further discussion of co-articulation and gestural overlap can be found in CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY.

3.3 *Formalizing local assimilation*

Local assimilation has played an important role in the development of phonological formalism. McCarthy (1988: 84) states: "The goal of phonology is the construction of a theory in which cross-linguistically common and well-established processes emerge from very simple combinations of the descriptive parameters of the model." He further argues that the ubiquitous presence of assimilation, both local and long-distance, warrants assigning it a "privileged status" in phonological formalism (1988: 86). Despite its clear phonetic bases, the process of assimilation has not necessarily been simple to capture in phonological representation.

In the formal theory of Chomsky and Halle (1968), processes of assimilation were expressed with the use of alpha notation. In this formalism, Greek letters stand for variables over "+" and "–", and every instance of the variable in a rule must be filled in with the same value. Thus, a rule of obstruent voicing agreement, as would be needed for example in Yiddish (4), would be written as in (29).

(29) *Obstruents agree in voicing: α notation*

$[-\text{sonorant}] \rightarrow [\alpha\text{voice}] \;/\; \_\_ \; [-\text{sonorant}, \alpha\text{voice}]$

While the use of a special notation does convey the privileged status of the notion of "agreement," non-occurring rules can also be easily represented, with no increase in formal complexity.

(30) *Obstruent voicing must match the value for* [±back]

[–sonorant] → [αvoice] / __ [–sonorant, αback]

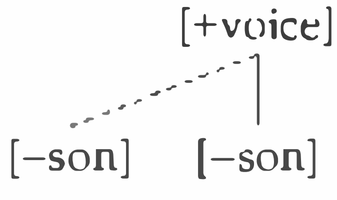
Thus, as pointed out, for example, by Bach (1968) and Anderson (1985), this rule formalism is too powerful, in that it predicts that rules (29) and (30), being equal in complexity, should be equally likely to occur. On the other hand, the common and straightforward process of nasal place assimilation (§2.4.1 above) is represented via a complicated formula (31):

(31) *Catalan nasal assimilation using alpha notation*

[+nasal] →	[αcoronal]	/ __	[αcoronal]
	[βanterior]		[βanterior]
	[γlabial]		[γlabial]
	[δback]		[δback]
	[εhigh]		[εhigh]
	[φdistributed]		[φdistributed]

It was the study of long-distance assimilation – tone and vowel harmony – that led to the introduction of autosegmental phonology (Goldsmith 1976; Clements and Sezer 1982), but this formalism was quickly adopted for local assimilations as well. In autosegmental representation, assimilation is represented by "feature spreading" through the addition of an "association line": a feature that begins as a property of one segment comes to be associated with more than one, as in the anticipatory voicing assimilation in (32):

(32) *Obstruents agree in voicing: Autosegmental notation*



Feature spreading gives assimilation a privileged status as an elementary operation, while more complicated feature switches have a correspondingly more complicated representation.

The addition of class nodes in a more elaborated *feature geometry* allows for a simple representation of rules that target a group of features. As noted by Clements (1985: 226),

> If we find that certain sets of features consistently behave as a unit with respect to certain types of rules of assimilation or resequencing, we have good reason to suppose that they constitute a unit in phonological representation.

Local place assimilation is the prime example of a set of features that behave as a unit. McCarthy (1988: 86–87) states: "The basic motivation for feature geometry [is] the naturalness of place assimilation."

Consensus has not been reached, however, on exactly which geometry is correct. The need is clear for a class node grouping consonantal place features to account for assimilations such as that in Catalan (15a), one grouping laryngeal features to account for assimilation of voicing and aspiration together as in Greek (9) and Sanskrit (10), and a ROOT node grouping all features for complete assimilation as in Ponapean (24a) or Arabic (25a). Less clear is the need for a SUPRALARYNGEAL node that groups all features except the laryngeal features. Cases like those in Havana Spanish (25b) and Kannada (26), where all features except voice assimilate, would argue for such a node (see Clements 1985); however, McCarthy (1988: 92) counters that spreading of the supralaryngeal node "is known from only one or two examples that are subject to reanalysis." Other points of contention include where to attach manner features (Padgett 1994), how to represent the class of guttural consonants (McCarthy 1994), and, probably most difficult, how to handle vowel and consonant interactions and lack of interaction. Clements and Hume (1995; see also Clements 1993; Hume 1994) suggest separate Place nodes for C-place and V-place: different patterns of interaction and transparency will depend on which nodes are targeted for assimilation. For extended further discussion, see CHAPTER 13: THE STRICTURE FEATURES; CHAPTER 14: AUTOSEGMENTS; CHAPTER 19: VOWEL PLACE; CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS; CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION; CHAPTER 25: PHARYNGEALS; CHAPTER 27: THE ORGANIZATION OF FEATURES.

Constraint-based theories (Prince and Smolensky 1993) offer a different way of formalizing assimilations. Although autosegmental representation is generally assumed, the details of feature-geometrical representations become less crucial. One way of representing local assimilation is through the mechanism of AGREE constraints: markedness constraints that state that two adjacent segments must agree with respect to the specified feature. These markedness constraints interact with constraints requiring faithfulness to underlying features, with language-specific rankings producing different patterns of assimilation (see CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). Thus Lombardi (1999) proposes the constraints in (33) to account for voicing assimilation in Yiddish ((4) above). The positional faithfulness constraint (33c) is needed to account for the fact that the coda assimilates to the onset and not vice versa.

(33) *Constraints on obstruent voicing agreement* (Lombardi 1999)

a. AGREE
   Obstruent clusters should agree in voicing.
b. IDENT(Laryngeal)
   Consonants should be faithful to underlying laryngeal specification.
c. IDENT-ONSET(Laryngeal)
   Consonants in [pre-sonorant position] should be faithful to underlying laryngeal specification.

If these are ranked such that the agreement constraint and the positional faithfulness constraint outrank general faithfulness, as in (34), the result is that the coda will assimilate in voicing to the onset.

(34) *Voicing assimilation in Yiddish* (Lombardi 1999)

/bak bejn/	AGREE	IDENT-ONSET(Lar)	IDENT(Lar)
a. bak.bejn	*!		
☞ b. bag.bejn			*
c. bak.pejn		*!	*

Steriade (2001) treats place assimilation with parallel formalism, but substitutes positional faithfulness constraints that reference differences in perceptibility rather than syllable structure (see §3.1 above).

(35) *Place assimilation with perceptibility constraints* (Steriade 2001)

/at pa/	AGREE	IDENT(Place)/C_V	IDENT(Place)/V_C
a. atpa	*!		
☞ b. appa			*
c. atta		*!	*

Place assimilation is also often handled with reference to positional markedness as well as positional faithfulness (Kager 1999). In this approach, assimilation is not driven by a constraint requiring agreement. Rather, the markedness constraint that forces the alternation is based on Itô's (1998) insight that codas may not license place features alone. Direct reference to a "coda condition" captures the insight that assimilation to the place of an adjacent onset consonant is just one way to repair the coda violation; epenthesis and deletion, which change the syllable structure rather than featural content, are others. The use of different constraints for place assimilation and voice assimilation captures the generalization that, cross-linguistically, epenthesis and deletion often occur to repair clusters that do not match in place, but they do not occur to repair clusters that do not match in voicing (see Baković 2000; Lombardi 2001). The account of nasal place assimilation in Spanish [tampoko] 'neither' in (36) and (37) is adapted from Shepherd (2003).

(36) CODA CONDITION
A coda cannot license place features.

(37) *Nasal place assimilation in Spanish*

/taN.po.ko/	CODACOND	IDENT-ONSET(Place)	IDENT(Place)
a. tan.po.ko	*!		
☞ b. tam.po.ko			*
c. tan.to.ko		*!	*

Note that in the tableaux above, there is no specific reference to feature geometry or a Place node. In keeping with a general move away from solutions based in representations and rules, the sets of features targeted for assimilation are defined within the content of the constraints, not in terms of a universal hierarchical structure that must be made to work for all cases. Padgett (1995) specifically argues against a Place node in feature geometry, proposing instead that constraints that target defined sets of features better account for partial place assimilations.

In conclusion, it may be said that questions of representation encapsulate the debates that continue over the linguistic nature of local assimilation. Phonologists are working toward finding the representation that will capture crucial cross-linguistic generalizations about assimilation in the simplest and most straightforward form, while accounting for the details of individual datasets. Debates continue over defining the features and feature classes that are active in assimilation, and whether the definition of classes should be representational or set-theoretic. It remains a question whether structural or perceptual approaches to directional asymmetries best account for the range of cross-linguistic data. Another important question is whether assimilation is featural at all: should local assimilation be defined in terms of manipulation of phonological features, in terms of articulatory organization, or in some other way? Accounting for both gradience and variability on the one hand and systematic category change on the other continues to be a challenge. Finally, theories of the phonology–morphology interface, the phonetics–phonology interface, and, most generally, theories of the overall structure and architecture of the phonological grammar continue to reference processes of local assimilation. Certainly local assimilation, the most common phonological alternation, will continue to play a central role in phonological theorizing.

ACKNOWLEDGMENT

The outline of the argument in this chapter follows that of the briefer treatment of the same subject in Zsiga (2006).

REFERENCES

Anderson, Stephen R. 1985. *Phonology in the twentieth century: Theories of rules and theories of representations*. Chicago: University of Chicago Press.

Armstrong, Lilias E. 1967. *The phonetic and tonal structure of Kikuyu*. London: Dawsons of Pall Mall for the International African Institute.

Bach, Emmon. 1968. Two proposals concerning the simplicity metric in phonology. *Glossa* 2. 128–149.

Baković, Eric. 2000. Nasal place neutralization in Spanish. In Michelle Minnick Fox, Alexander Williams & Elsi Kaiser (eds.) *Proceedings of the 24th Annual Penn Linguistics Colloquium*, 1–12. Philadelphia: University of Pennsylvania.

Beckman, Jill N. 1998. Positional faithfulness. Ph.D. dissertation, University of Massachusetts, Amherst.

Bessell, Nicola J. 1992. Towards a phonetic and phonological typology of post-velar articulations. Ph.D. dissertation, University of British Columbia.

Bessell, Nicola J. 1998. Local and non-local consonant–vowel interaction in Interior Salish. *Phonology* 15. 1–40.

Borowsky, Toni. 2000. Word faithfulness and the direction of assimilations. *The Linguistic Review* 17. 1–28.

Bradley, Travis C. 2007. Morphological derived-environment effect in gestural coordination: A case study of Norwegian clusters. *Lingua* 117. 950–985.

Bright, William. 1957. *The Karok language*. Berkeley & Los Angeles: University of California Press.

Browman, Catherine P. 1995. Assimilation as gestural overlap: Comments on Holst and Nolan. In Connell & Arvaniti (1995), 334–342.

Browman, Catherine P. & Louis Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45. 140–155.

Browman, Catherine P. & Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341–376. Cambridge: Cambridge University Press.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180.

Byrd, Dani. 1992. Perception of assimilation in consonant clusters: A gestural model. *Phonetica* 49. 1–24.

Calabrese, Andrea & Samuel J. Keyser. 2006. On the peripatetic behavior of aspiration in Sanskrit roots. In Eric Baković, Junko Ito & John J. McCarthy (eds.) *Wondering at the natural fecundity of things: Essays in honor of Alan Prince*. Linguistic Research Center, University of California, Santa Cruz. Available (June 2010) at http://escholarship.org/uc/item/96k332nm.

Casali, Roderic F. 1996. Resolving hiatus. Ph.D. dissertation, University of California, Los Angeles.

Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73. 493–533.

Chen, Su-I. 1996. A theory of palatalization and segment implementation. Ph.D. dissertation, New York State University, Stony Brook.

Cho, Taehong & James McQueen. 2008. Not all sounds in assimilation environments are perceived equally: Evidence from Korean. *Journal of Phonetics* 36. 239–249.

Cho, Young-mee Yu. 1999. *Parameters of consonantal assimilation*. Munich: Lincom Europa.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. 1993. Lieu d'articulation des consonnes de des voyelles: Une théorie unifiée. In Bernard Laks & Annie Rialland (eds.) *Architecture des représentations phonologiques*, 101–145. Paris: CNRS.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 245–306. Cambridge, MA & Oxford: Blackwell.

Clements, G. N. & Engin Sezer. 1982. Vowel and consonant disharmony in Turkish. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part II, 213–255. Dordrecht: Foris.

Cohn, Abigail C. 1993. Nasalisation in English: Phonology or phonetics. *Phonology* 10. 43–81.

Connell, Bruce & Amalia Arvaniti (eds.) 1995. *Papers in laboratory phonology IV: Phonology and phonetic evidence*. Cambridge: Cambridge University Press.

Cowell, Mark W. 1964. *A reference grammar of Syrian Arabic (based on the dialect of Damascus)*. Washington, DC: Georgetown University Press.

Davis, Stuart & Seung-Hoon Shin. 1999. The syllable contact constraint in Korean: An optimality-theoretic analysis. *Journal of East Asian Linguistics* 8. 285–312.

Doke, Clement M. 1926. *The phonetics of the Zulu language*. Johannesburg: University of the Witwatersrand Press.

Donegan, Patricia J. & David Stampe. 1979. The study of Natural Phonology. In Daniel A. Dinnsen (ed.) *Current approaches to phonological theory*, 126–173. Bloomington: Indiana University Press.

Gafos, Adamantios I. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20. 269–337.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.

Gurevich, Naomi. 2003. Functional constraints on phonetically-conditioned sound changes. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Halle, Morris. 1959. *The sound pattern of Russian: A linguistic and acoustical investigation*. The Hague: Mouton.

Harris, James W. 1985. Autosegmental phonology and liquid assimilation in Havana Spanish. In Larry D. King & Catherine A. Maley (eds.) *Selected papers from the 13th Linguistic Symposium on Romance Languages*, 127–148. Amsterdam: John Benjamins.

Hayes, Bruce. 1984. The phonetics and phonology of Russian voicing assimilation. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*, 318–328. Cambridge, MA: MIT Press.

Herzallah, Rukayyah. 1990. Aspects of Palestinian Arabic phonology: A non-linear approach. Ph.D. dissertation, Cornell University.

Honorof, Douglas. 1999. Articulatory gestures and Spanish nasal assimilation. Ph.D. dissertation, Yale University.

Howe, Darin. 2000. Oowekyala segmental phonology. Ph.D. dissertation, University of British Columbia.

Hume, Elizabeth. 1994. *Front vowels, coronal consonants and their interaction in nonlinear phonology*. New York: Garland.

Hume, Elizabeth. 2003. Language specific markedness: The case of place of articulation. *Studies in Phonetics, Phonology and Morphology* 9. 295–310.

Hume, Elizabeth & Keith Johnson (eds.) 2001. *The role of speech perception in phonology*. San Diego: Academic Press.

Hume, Elizabeth & David Odden. 1996. Reconsidering [consonantal]. *Phonology* 13. 345–376.

Hume, Elizabeth & Georgios Tserdanelis. 2002. Labial unmarkedness in Sri Lankan Portuguese Creole. *Phonology* 19. 441–458.

Hyman, Larry M. 2008. Directional asymmetries in the morphology and phonology of words, with special reference to Bantu. *Linguistics* 46. 309–350.

Itô, Junko. 1998. *Syllable theory in prosodic phonology*. New York: Garland.

Itô, Junko & Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17. 49–73.

Jakobson, Roman. 1978. Mutual assimilation of Russian voiced and voiceless consonants. *Studia Linguistica* 32. 107–110.

Jun, Sun-Ah. 1995. Asymmetrical prosodic effects on the laryngeal gesture in Korean. In Connell & Arvaniti (1995), 235–253.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kaisse, Ellen M. 1992. Can [consonantal] spread? *Language* 68. 313–332.

Katz, David. 1987. *A grammar of the Yiddish language*. London: Duckworth.

Kawasaki, Haruko. 1982. An acoustical basis for universal constraints on sound sequences. Ph.D. dissertation, University of California, Berkeley.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kim-Renaud, Young-Key. 1991. *Korean consonantal phonology*. Seoul: Hanshin.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Kirchner, Robert. 1998. An effort-based approach to consonant lenition. Ph.D. dissertation, University of California, Los Angeles.

Kirchner, Robert. 2000. Geminate inalterability and lenition. *Language* 76. 509–545.

Kochetov, Alexei & Marianne Pouplier. 2008. Phonetic variability and grammatical knowledge: An articulatory study of Korean place assimilation. *Phonology* 25. 399–431.

Ladd, D. Robert & James M. Scobbie. 2003. External sandhi as gestural overlap? Counter-evidence from Sardinian. In John Local, Richard Ogden & Rosalind Temple (eds.) *Papers in laboratory phonology* VI, 162–180. Cambridge: Cambridge University Press.

Lavoie, Lisa. 2001. Consonant strength: Phonological patterns and phonetic manifestations. Ph.D. dissertation, Cornell University.

Lewis, Geoffrey L. 1967. *Turkish grammar.* Oxford: Oxford University Press.

Lindblom, Björn. 1983. Economy of speech gestures. In Peter F. MacNeilage (ed.) *The production of speech,* 217–245. New York: Springer.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17. 267–302.

Lombardi, Linda. 2001. Why Place and Voice are different: Constraint-specific alternations in Optimality Theory. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations,* 13–45. Cambridge: Cambridge University Press.

Maddieson, Ian & Peter Ladefoged. 1989. Multiply articulated segments and the feature hierarchy. *UCLA Working Papers in Phonetics* 72. 116–138.

Mascaró, Joan. 1976. Catalan phonology and the phonological cycle. Ph.D. dissertation, MIT.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.) *Phonological structure and phonetic form: Papers in laboratory phonology III,* 191–233. Cambridge: Cambridge University Press.

Mohanan, K. P. 1993. Fields of attraction in phonology. In John A. Goldsmith (ed.) *The last phonological rule: Reflections on constraints and derivations,* 61–116. Chicago & London: University of Chicago Press.

Navarro Tomás, Tomás. 1970. *Manual de pronunciación española.* Madrid: Publicaciones de la Revista de Filología Española.

Nurse, Derek & Gérard Philippson (eds.) 2003. *The Bantu languages.* London & New York: Routledge.

Ohala, John J. 1981. The listener as a source of sound change. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 17(2), 178–203.

Orr, Carolyn. 1962. Ecuador Quichua phonology. In Benjamin F. Elson (ed.) *Studies in Ecuadorian Indian languages,* vol. 1, 60–77. Norman, OK: Summer Institute of Linguistics.

Padgett, Jaye. 1994. Stricture and nasal place assimilation. *Natural Language and Linguistic Theory* 12. 465–513.

Padgett, Jaye. 1995. Partial class behavior and nasal place assimilation. In Keiichiro Suzuki & Dirk Elzinga (eds.) *Proceedings of the 1995 Southwestern Workshop on Optimality Theory (SWOT),* 145–183. Tucson: Department of Linguistics, University of Arizona. Reprinted 2004 in John J. McCarthy (ed.) *Optimality Theory: A reader in phonology,* 379–393. Oxford: Blackwell.

Padgett, Jaye. 2002. Russian voicing assimilation, final devoicing, and the problem of [v]. Unpublished ms., University of California, Santa Cruz (ROA-528).

Paradis, Carole & Jean-François Prunet (eds.) 1991. *The special status of coronals: Internal and external evidence.* San Diego: Academic Press.

Pater, Joe. 1999. Austronesian nasal substitution and other NÇ effects. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface,* 310–343. Cambridge: Cambridge University Press.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Rehg, Kenneth L. & Damian G. Sohl. 1981. *Ponapean reference grammar.* Honolulu: University of Hawai'i Press.

Rice, Keren. 1993. A re-examination of the feature [sonorant]: The status of "sonorant obstruents." *Language* 69. 308–344.

Roca, Iggy & Wyn Johnson. 1999. *Workbook in phonology*. Malden, MA & Oxford: Wiley-Blackwell.

Rose, Sharon. 1996. Variable laryngeals and vowel lowering. *Phonology* 13. 73–117.

Scobbie, James M. 2007. Interface and overlap in phonetics and phonology. In Gillian Ramchand & Charles Reiss (eds.) *The Oxford handbook of linguistic interfaces*, 17–52. Oxford: Oxford University Press.

Shepherd, Michael. 2003. Constraint interactions in Spanish phonotactics. M.A. thesis, Cal State Northridge (ROA-639).

Silva, David J. 1992. The phonetics and phonology of stop lenition in Korean. Ph.D. dissertation, Cornell University.

Smyth, Herbert Weir. 1920. *Greek grammar*. Cambridge, MA: Harvard University Press.

Sohn, Hyang-Sook. 2008. Phonological contrast and coda saliency of sonorant assimilation in Korean. *Journal of East Asian Linguistics* 17. 33–59.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Hume & Johnson (2001), 219–250.

Street, Chester S. & Gregory P. Mollinjin. 1981. The phonology of Murinbata. In Bruce Waters (ed.) *Australian phonologies: Collected papers*, 183–244. Darwin: Summer Institute of Linguistics.

Szpyra, Jolanta. 1989. *The phonology–morphology interface: Cycles, levels and words*. London & New York: Routledge.

Tlale, One. 2006. The phonetics and phonology of Sengwato, a dialect of Setswana. Ph.D. dissertation, Georgetown University.

Tsuchida, Ayako. 1997. Phonetics and phonology of vowel devoicing. Ph.D. dissertation, Cornell University.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany: State University of New York Press.

Villafana, Christina. 2006. Consonant weakening in Florentine Italian: An acoustic study of gradient and variable sound change. Ph.D. dissertation, Georgetown University.

Wackernagel, J. 1896. *Altindische Grammatik*, vol 1. Göttingen: Vandenhoeck & Ruprecht.

Welmers, William E. 1973. *African language structures*. Berkeley: University of California Press.

Wetzels, W. Leo & Joan Mascaró. 2001. The typology of voicing and devoicing. *Language* 77. 207–244.

Whitney, William Dwight. 1889. *Sanskrit grammar, including both the classical language, and the older dialects of Veda and Brahmana*. 2nd edn, Cambridge, MA: Harvard University Press.

Wonderly, William L. 1946. Phonemic acculturation in Zoque. *International Journal of American Linguistics* 12. 92–95.

Zsiga, Elizabeth C. 1995. An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. In Connell & Arvaniti (1995), 282–302.

Zsiga, Elizabeth C. 1997. Features, gestures, and Igbo vowels: An approach to the phonology–phonetics interface. *Language* 73. 227–274.

Zsiga, Elizabeth C. 2006. Assimilation. In Keith Brown (ed.) *Encyclopedia of language and linguistics*, 2nd edn., vol. 1, 553–558. Oxford: Elsevier.

Volume I: General Issues and Segmental Phonology

Volume II: Suprasegmental and Prosodic Phonology

Volume III: Phonological Processes

Volume IV: Phonological Interfaces

Volume V: Phonology across Languages

The Blackwell Companion to Phonology is a major reference work drawing together 124 new contributions from leading scholars in the field. Led by a team of renowned international scholars, the *Companion* represents a diverse range of approaches and methodologies to key phenomena in phonological research. In contrast to other handbooks and reference works currently available for phonology, these volumes focus on phenomena and case studies that highlight past and ongoing debates in the field. Interdisciplinary connections, such as those with the social and computational sciences, are covered, as well as statistical and experimental techniques. The *Companion* will be a touchstone for future phonological theorists, giving an overview of data and insights which any theory of phonology should be able to cover.

Simultaneously offering broad coverage and a high level of detail, *The Blackwell Companion to Phonology* is a landmark work available in both print and online editions that will be indispensable to students and researchers in the field for years to come.

www.companiontophonology.com

WILEY-BLACKWELL
www.wiley.com/wiley-blackwell

VOLUME IV
Phonological Interfaces

Edited by
van Oostendorp,
Ewen, Hume,
and Rice

THE BLACKWELL COMPANION TO
Phonology

WILEY-BLACKWELL

VOLUME IV
Phonological Interfaces

Edited by
Marc van Oostendorp
Colin J. Ewen
Elizabeth Hume
Keren Rice

THE BLACKWELL COMPANION TO Phonology

WILEY-BLACKWELL

MARC VAN OOSTENDORP is Senior Researcher at the Department of Variationist Linguistics at the Meertens Institute of the Royal Netherlands Academy of Arts and Sciences, and Professor of Phonological Microvariation at the University of Leiden. He holds an M.A. in Computational Linguistics and a Ph.D. from Tilburg University.

COLIN J. EWEN is Professor and Chair of English Language and Linguistics at the University of Leiden. He holds an M.A. in English Language and Linguistics and a Ph.D. from the University of Edinburgh. He is editor (with Ellen Kaisse) of the journal *Phonology*.

ELIZABETH HUME is Professor and Chair of the Department of Linguistics at The Ohio State University. She holds a Ph.D. and M.A. in Linguistics from Cornell University, an M.A. in French and Social Psychology of Language from McMaster University, and a B.A. in French from Université Laval.

KEREN RICE is University Professor and Canada Research Chair in Linguistics and Aboriginal Studies at the University of Toronto. She holds an M.A. and Ph.D. in Linguistics from the University of Toronto and a B.A. from Cornell University.

Blackwell Companions to Linguistics series

Series Editors:
Martin Everaert and Henk van Riemsdijk

The Blackwell Companion to Syntax
Edited by Martin Everaert and Henk van Riemsdijk

The Blackwell Companion to Phonology
Edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice

The Blackwell Companion to Phonology

Edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice

Volume IV
Phonological Interfaces

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This edition first published 2011

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

The Blackwell companion to phonology / edited by Marc van Oostendorp . . . [et al.].
p. cm. — (Blackwell companions to linguistics series)
Includes bibliographical references and index.
ISBN 978-1-4051-8423-6 (hardcover : alk. paper)
1. Phonetics. 2. Grammar, Comparative and general—Phonology.
I. Oostendorp, Marc van, 1967–
P217.B53 2011
414—dc22 2010042206

A catalogue record for this book is available from the British Library.

Set in 10/12pt Palatino by Graphicraft Limited, Hong Kong

1 2011

Brief Contents

Volume I

Volume II

Volume III

Volume IV

Volume V

Full Table of Contents

Volume I

Volume II

Suprasegmental and Prosodic Phonology

Volume III

Phonological Processes

Volume IV

Phonological Interfaces

Volume V

Phonology across Languages

82 Featural Affixes

Akinbiyi Akinlabi

1 Characteristics of featural affixes

Featural affixes are phonological features that function as grammatical morphemes. The most commonly found cases are tonal (Akinlabi 1996). An example is the associative marker in Bini (Amayo 1976), exemplified in (1). (The forms before the arrow indicate the isolation forms of the nouns and the forms after the arrow are associative constructions. For clarity, the tones in the examples in (1) are indicated with both tone marks and the letters L, H for Low, High respectively. ↓ indicates a downstepped tone on the following vowel.)

(1) *Bini* (Amayo 1976)

```
òwɛ̀     ɔ̀sà        →  òwɛ́ ɔ̀sà    [òwɔ́↓sà]
| |     | |            | | | |
L L     L L            L H L L    'a chimpanzee's leg'
leg     chimpanzee

àmɛ̀     èhj̃ɛ̃́        →  àmɛ́ èhj̃ɛ̃́    [àmé↓hj̃ɛ̃́]
| |     | |            | | | |
L L     L H            L H L H    'solution of water and pepper'
water   pepper

òwɛ̀     ɔ̀nà        →  òwɛ́ ɔ̀nà    [òwɔ́↓nà]
| |     | |            | | | |
L L     L L            L H L L    'this one's leg'
leg     this one
```

However, several cases of non-tonal features functioning as grammatical morphemes have also been described in the literature. A representative list is given in (2).[1]

[1] See the references cited here for additional examples. Reviewers have pointed out a number of other examples which might have been included here. Two of them are: (a) in Coatzospan, the 2nd person familiar is marked by nasality (Gerfen 1999: 127), and (b) in Shuswap, glottalization is a floating feature (Kuipers 1974; Idsardi 1992). The list in (2) is not intended to be exhaustive.

(2) *Non-tonal examples of featural morphemes*

a. In Chaha, the 3rd masculine object is indicated by labialization. (Johnson 1975; McCarthy 1983; Hendricks 1989; Archangeli and Pulleyblank 1994; Rose 1994, 2007)

b. Nuer indicates tense/aspect distinctions with the features [continuant] and [voice]. (Crazzolara 1933; Lieber 1987; Frank 1999)

c. In Zoque, the 3rd person singular is marked by palatalization. (Wonderly 1951)

d. [nasal] is the 1st person possessive marker in Terena. (Bendor-Samuel 1960, 1966)

e. The feature of "uncontrolledness" is signaled by palatalization in Japanese. (Hamano 1986; Mester and Itô 1989; Archangeli and Pulleyblank 1994; Alderete and Kochetov 2009)

f. Noun class 5 is marked by voicing the first consonant of the root in Aka (Bantu, Zone C). (Kosseke and Sitamon 1993; Roberts 1994)

g. Noun class morphemes in Fula include the features [continuant] and [nasal]. (Arnott 1970; Lieber 1984, 1987)

h. The Athapaskan D-classifier consists solely of the feature [–continuant]. (Rice 1987)

i. In Seereer Siin, an Atlantic (Niger Congo) language, consonant mutation (involving the features [voice] and [continuant]) constitute all or part of the noun class prefix in nouns and dependent adjectives, and number in verbs. (Mc Laughlin 2000, 2005)

j. In Mafa, a central Chadic language of Cameroon, imperfectives of verbs ending in a consonant are formed with a palatal featural affix. (Ettlinger 2003, 2004)

The features in (2), like segmental morphemes, often refer to specific edges of stems, and thus are featural affixes (e.g. Chaha labialization and palatalization, Aka voicing, Zoque palatalization). While the fact that phonological features may function as grammatical morphemes is uncontroversial, the status of such features as prefixes or suffixes often remained muted in spite of traditional intuition, with some scholars contented with referring to the morphemes simply as "floating autosegments."[2] The reason why the status of featural affixes as prefixes or suffixes is often problematic is that, while segmental affixes may be phonetically realized independently, featural affixes are always phonetically realized as part of some other segment or segments of the stem. The question therefore is why featural affixes get realized as part of the stem. The answer to this is that features have to be "licensed" (i.e. their occurrences have to be sanctioned) in order to get phonetically realized, therefore featural affixes must associate with a licensor in the stem or elsewhere.

[2] Most studies on tone are exceptions to this generalization (see Clements and Goldsmith 1984; Pulleyblank 1986; Anderson 1991; van der Hulst and Snider 1993).

In this chapter I am assuming a feature geometry in which all segments have a root node, which "gathers" the features into one unit (CHAPTER 27: THE ORGANIZATION OF FEATURES). In addition, I assume that vowels (and all syllable peaks, including syllabic nasals) are dominated by a mora (CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE). Finally, I assume that class nodes, such as those for place of articulation, are monovalent. However, terminal features, such as aperture features, are bivalent. Since this chapter has a constraint-based, optimality-theoretic bias, I will not be assuming underspecification here (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION).

Universally, feature licensors can (only) be either a mora or a root node (Itô 1989; Itô and Mester 1993; etc.). Therefore, while edges in tones refer to the initial or final mora, edges in nasal harmony and the like may refer to the first or last root node; i.e. a real morphological edge, since the last licensor also coincides with the last segment of the morpheme (see Archangeli and Pulleyblank 1994).[3] But, with featural affixes, an edge does not necessarily mean a morphological edge; an edge is defined for a feature on the basis of a possible licensor in a language.

Another characteristic of featural affixes, as distinct from segmental affixes, is their domain. While most segmental affixes occur at the beginning, middle, or end of a base, featural affixes often occur throughout the base, or span it. Features that commonly have this characteristic are the "prosodic" features, in the Firthian sense of the word. As is well known, such features may include pitch, nasality, roundness, palatalization, and the like (see Firth 1948). Since these are the featural spell-out (or content) of the morphological categories in question, they are featural affixes.

In their study of alignment in (regular) segmental affixation, McCarthy and Prince (1993b: 103) observe that an alignment constraint, such as one that aligns the left edge of one morpheme with the right edge of another (as in Tagalog *-um-* prefixation) may be violated when dominated by a prosodic constraint, such as one that disallows a coda. This may force a prefix to be realized as an infix. The Tagalog affix *-um-* "falls as near as possible to the left edge of the stem, so long as it obeys the phonological requirement that its final consonant *m* not be syllabified as a coda" (McCarthy and Prince 1993b: 79). Therefore, it appears as a prefix before a vowel-initial word: /um + aral/ → [um-aral] 'teach', but as an infix when the word is consonant-initial: /um + sulat/ → [s-um-ulat] 'write', /um + gradwet/ → [gr-um-adwet] 'graduate'.

A similar characteristic is found in featural affixes. One important distinction from segmental prefixes/suffixes is that featural affixes often behave like "infixes," because they frequently do not occur at an edge of the stem. A feature may be forced away from an edge when the feature cannot co-occur with another feature(s) of the segment at the edge (see Pulleyblank 1993), leading to

[3] It should be noted that the accounts in this chapter allow for affixes which involve more than one autosegmental feature, though we do not discuss such cases here. For example, in Mokulu (Eastern Chadic, Chad Republic) the completive aspect marker consists of the features [voice] and [high] (Jungraithmayr 1990; Roberts 1994). The first consonant of the stem becomes voiced while the first vowel becomes high, even if it was a low vowel in the input. In the approach taken here, both features constitute parts of a featural prefix. However, such features may be realized on the same segment in the stem or on different segments, depending on licensing. In the case in question, licensing forces [voice] and [high] on different segments.

misalignment. A featural suffix may for example be realized elsewhere in the stem, resulting in featural infixation. However, featural affixes occur as "infixes" more often than segmental affixes.

Finally, one characteristic that has recently been observed in featural affixation is one in which a grammatical category is marked by a feature which has both segmental and featural allomorphs, as in Mafa (Ettlinger 2003, 2004).

In the following sections I illustrate each of the above characteristics of featural affixes. Each case study discussed below has been selected because it illustrates a particular characteristic or characteristics of featural affixes.

In the discussion of Chaha (§2.1), I show that a featural suffix [round] is realized as a featural infix, or even as a featural prefix, when the featural suffix is forced away from the edge. The opposite effect is illustrated with Nuer mutation (§2.3).

Tonal data from Etsako, an Edoid language, and nasalization data from Terena show situations in which featural morphemes span the entire base of affixation. In the discussions of Terena nasalization and the Etsako tone, I suggest that these are still cases of prefixation and suffixation respectively, but in conjunction with harmony. Therefore there are no special treatments of featural affixes required.

Mc Laughlin (2000, 2005) notes that, taking into consideration featural affixes, a morphological category can be expressed in one of three ways: as a segmental affix, as a featural affix, or as a combination of both segmental and featural affixes (CHAPTER 103: PHONOLOGICAL SENSITIVITY TO MORPHOLOGICAL STRUCTURE).

In summary, the primary focus in this chapter will be illustrating the characteristics of featural affixes. To do this, I will provide short descriptions of several of the featural affixes listed in (2). The characteristics include (a) marking morphological categories (like segmental affixes), (b) occurring as part of other segments rather than independently, (c) varying between prefixes and suffixes, (d) occurring elsewhere in the stem (because of feature co-occurrence constraints), (e) spanning the entire base of affixation, and (f) varying occurrence as a feature or a segment in the same language. I will argue that these characteristics of featural affixes do not require any new type of morphology, because the same machinery already developed for segmental affixes can handle them as well.

I discuss seven case studies in all, divided into four groups. The first group, Chaha and Zoque, illustrates the most basic characteristics of featural affixes mentioned above, that of directionality. Chaha illustrates suffixation and Zoque shows prefixation. The second group, Nuer and Seereer Siin, combines featural affixes with consonant mutation. Nuer is suffixal, and Seereer Siin is prefixal. The third group, Etsako and Terena, shows featural affixes that span the whole stem domain. They illustrate featural affixation combined with "harmony." Again, Etsako shows the harmony from the right (suffixal), and Terena shows it from the left (prefixal). The fourth group contains only one language, Mafa. Mafa shows a special case of affixation, in that the segment involved is at the same time a segment and a feature. I refer to this as segmental realization of a featural affix.

2 Directionality

The first case studies illustrate the need to consider featural morphemes as either prefixes or suffixes, a property that is formally accounted for by the directional component of alignment. In this light, Chaha illustrates prefixation, and Zoque illustrates suffixation.

2.1 Chaha labialization

In Chaha, a Gurage language of Ethiopia, the 3rd person masculine singular object is indicated by labialization (with the suffix /n/) (Johnson 1975; McCarthy 1983; Hendricks 1989; Archangeli and Pulleyblank 1994; Rose 1994, 2007). Labialization surfaces on the "rightmost labializable consonant" of the stem. Labializable consonants in Chaha include labial and dorsal consonants, but not coronal consonants.[4] The data in (3) (from McCarthy 1983: 179) show the surface realization of this morpheme.

(3)		without object	with 3rd masc sg object	
	a.	*Rightmost consonant of the stem is labializable*		
		dænæg	dænæg^w	'hit'
		nædæf	nædæf^w	'sting'
		nækæb	nækæb^w	'find'
	b.	*Medial consonant of the stem is labializable, final is not*		
		nækæs	næk^{w}æs	'bite'
		kæfæt	kæf^{w}æt	'open'
		bækær	bæk^{w}ær	'lack'
	c.	*Only the leftmost consonant of the stem is labializable*		
		qætær	q^{w}ætær	'kill'
		mæsær	m^{w}æsær	'seem'
		mæk^{j}ær	m^{w}æk^{j}ær	'burn'
	d.	*No labializable consonant*		
		sædæd	sædæd	'chase'

A number of observations are important here. Labialization must be realized only on the rightmost labializable consonant, and on no other. This is obvious from the third example in (3a), /nækæb/ → /nækæb^w/. Both of the last two consonants of the verb root in this example are labializable, but only the root-final consonant is labialized. The medial consonant is not labialized, because of this requirement of rightmostness. In the forms in (3b), all of the final consonants of the verb roots are coronal, e.g. /nækæs/, therefore only the root-medial consonants, which are either labial or dorsal, are rightmost; and so only these receive the labialization feature. Note further that the initial consonants in the last two examples, /kæfæt/ and /bækær/, are labializable, but again are not labialized, because of the requirement of rightmostness. In (3c) the only labializable consonants of the verb root are the leftmost consonants, /qætær/ → /q^{w}ætær/, and so by rightmostness they receive labialization. Finally, in (3d) none of the consonants is labializable and so the feature is not realized.

An explanation of the above facts is as follows. Following earlier analyses we assume that the 3rd person masculine singular object marker in Chaha is

[4] This statement is from McCarthy. Rose (2007) states the labialization rule as "labialize the rightmost velar or labial consonant, unless already palatalized." The key point in both definitions is that labialization targets dorsal and labial consonants.

the feature [round]. It must be a featural suffix, as indicated by the insistence on rightmostness. The 3rd person masculine singular object [round] aligns with (or coincides with; Zoll 1996) the right edge of the stem. In Chaha, [round] may be licensed by any consonantal root node. The position explicitly treats the morpheme as a suffix, but the segmental content is a feature [round], hence what the constraint aligns is the feature [round]. The right edge of the stem has to coincide with the feature [round], the featural content of the affix. Thus the feature [round] seeks out the rightmost consonantal root node in the verb root for licensing, given the discussion of licensing and edges above. As noted in our description of the facts, coronal consonants cannot receive the labialization feature. This means that the feature [round] cannot be articulated with a coronal consonant in Chaha. We can bar this with a feature co-occurrence constraint, which forbids [round] from linking to a root node associated to [coronal].

To conclude, there are several characteristics of featural affixes, which this affix illustrates. First, it marks a morphological category, the 3rd person masculine singular object. Second, the realization is a feature, the feature [round]. Third, it must be realized as part of another segment, a consonant, because it is not a segment. Fourth, like any affix, it has a position. However, like a featural affix it seeks the rightmost dorsal or labial consonant for licensing. Therefore it is a suffix. Fifth, like segmental affixes, it can be pushed from the suffix position. As it is a featural affix, however, co-occurring with other features is what matters. It cannot co-occur with a coronal consonant; therefore it gets pushed more and more inwards until it finds the right consonant to co-occur with. Sixth, if it does not find the right licensor, it simply does not get realized. This is comparable with the null realization of certain segmental morphemes in language, as for example where a segmental affix is not realized for some phonotactic reason. One example is Dutch, which does not have geminate consonants. Here the 3rd person singular ending [-t] is not realized on verbs which end in a coronal plosive.[5]

(4) *Dutch 3rd person suffix* [-t] *absent after verb-final* [t]

a.	*ik lees*	[ɪk les]		'I read'
	hij leest	[hɛi lest]		'he reads'
b.	*ik zie*	[ɪk zi]		'I see'
	hij ziet	[hɛi zit]		'he sees'
c.	*ik eet*	[ɪk et]		'I eat'
	hij eet	[hɛi et]	*[etː]	'he eats'

2.2 *Zoque palatalization*

In this section, I consider the process of morphological palatalization in Zoque (Zoque-Mixe of southern Mexico). Zoque palatalization contrasts with Chaha labialization (§2.1) in some crucial senses. First, while Chaha labialization illustrates a case of long-distance realization of an affix, Zoque palatalization illustrates local realization; i.e. the affix must be realized at the edge, and nowhere else (Akinlabi 1996). Second, Zoque differs from Chaha in the sense that the featural affix is a prefix as opposed to a suffix.

[5] I am grateful to Marc van Oostendorp for this example from Dutch.

Wonderly (1951: 117–118) describes a process of palatalization (CHAPTER 71: PALATALIZATION) in Zoque, which marks the 3rd person singular. He represents this morpheme as a prefix [j],[6] and treats this process of palatalization as "metathesis" of [j] and the following consonant. A rule-based treatment assuming metathesis is proposed in Dell (1980). The relevant examples are listed in (5), with the morpheme transcribed as [j], following Wonderly.[7] My interpretation here is that Wonderly's [j] is a palatal feature, which I will assume is [−back].

(5) *Zoque 3rd person singular*

a.	*With labial consonants*			
	j - pata	pʲata		'his mat'
	j - pʲesa	pʲesa		'his room'
	j - buro	bʲuro		'his burro'
	j - faha	fʲaha		'his belt'
	j - mula	mʲula		'his mule'
	j - wʲakas	wʲakas		'his cow'
b.	*With alveolar consonants*			
	j - tatah	tʲatah	[catah]	'his father'
	j - tih	nʌ tʲihu	[nʌ cihu]	'he is arriving'
	j - duratsʌhk	nʌ dʲuratsʌhku	[nʌ ʤuratsʌhku]	'it is lasting'
	j - tsʌhk	tsʲahku	[ʧahku]	'he did it'
	j - sʌk	sʲʌk	[ʃʌk]	'his beans'
	j - swerte	ʃwerte	[ʃwerte]	'his fortune'
	j - nanah	nʲanah	[ɲanah]	'his mother'
c.	*With palatal consonants (no change)*			
	j - ʧoʔngoja	ʧoʔngoja		'his rabbit'
	j - ʃapun	ʃapun		'his soap'
d.	*With velar consonants*			
	j - kama	kʲama		'his cornfield'
	j - gaju	gʲaju		'his rooster'
e.	*With laryngeal consonants*			
	j - ˀatsi	ˀʲatsi		'his older brother'
	j - hajah	hʲajah		'her husband'
	j - huj	hʲuju		'he bought it'

All words in Zoque are consonant-initial. The data in (5) show that the 3rd person singular morpheme produces secondary palatalization of the first consonant of the stem if it is labial (5a), velar (5d), or laryngeal (5e); it turns alveolars into

[6] Wonderly used the symbol /y/. I have re-transcribed Wonderly's examples to be as close as possible to the IPA.

[7] The transcription here (from Wonderly 1951) is somewhat misleading, because one can be led to believe that the morpheme here is indeed /j-/, and not a feature. However, if this were a full segment as opposed to a feature, it would be completely unnecessary for the segment to seek licensing from another segment. It would also be completely accidental that metathesis is limited to glide-consonant sequences in this language. Note that this cannot be blamed on the sonority rise in an onset (i.e. [jC] → [Cj]), because the so-called metathesis also occurs in a sequence of two glides (which in many accounts are equal in sonority): /j - wakas/ → /wʲakas/ 'his cow'.

alveopalatals in (5b), and has no phonetic effect on underlying palatals (5c). As Wonderly (1951: 118) puts it, "when y [i.e. /j/] precedes an alveopalatal consonant č, š, the y is lost." In this analysis we assume that the morpheme is not "lost," but that it has no phonetic effect if the initial consonant of the stem is palatal.

I assume that the 3rd person singular in the above data is the feature [−back] (see Sagey 1986). [−back] is licensed by any root node in Zoque. It is apparently a featural prefix, given its restriction to the first (or leftmost) consonant.

The palatalization case in Zoque is completely straightforward. All consonants participate in the palatalization, regardless of place of articulation. For example, labials are not barred from being palatalized, as coronals are barred from being labialized in Chaha.

The only set of consonants that require additional comment is the set of palatal consonants, as seen in (5c) (/[−back]-ʃapun/ → [ʃapun] 'his soap'). There are two approaches to this set of consonants. One is to assume that the [−back] 3rd singular morpheme is unparsed when the first segment is underlyingly palatal. The second approach is to assume that [−back] links vacuously to a palatal segment. I adopt the second position here, since linking [−back] to a palatal consonant will not change the consonant's realization. If palatal consonants are assumed to have underlying tokens of [−back], then linking the morpheme in this case simply implies that the [−back] specification in the surface representation corresponds to two tokens of the same feature in the input. Phonetically, it will be impossible to distinguish one or two tokens of the same feature.

In conclusion, Zoque provides evidence for featural affixes which must be realized, and which must be realized at an edge and nowhere else. In Chaha, a co-occurrence constraint forces a featural affix away from the edge. In Zoque, such co-occurrence constraints (which must be universal) have no surface effect. In Chaha, a featural affix may not be realized if none of the segments can license it. In Zoque, the affix can be licensed by all consonants, and so it is always realized.[8]

In the two case studies of featural affixes discussed above, one is a suffix (Chaha [round] or Labial), and the other is a prefix (Zoque [−back] or Coronal). Both of these involve only features. I now turn to cases in which the affix has both segmental and featural content.

3 Features plus segments: Segment mutations

Systematic alternation in homorganic segment classes that reflect morphological distinction is often called mutation (CHAPTER 65: CONSONANT MUTATION; CHAPTER 117: CELTIC MUTATIONS). The second group of case studies consists of languages which combine featural affixes with consonant mutation.

[8] There are two important issues here. First, there is a technical complication for feature geometry. If [−back] is dependent on some supralaryngeal node, and if laryngeal consonants have no supralaryngeal specification, then what does [−back] dock on? A possible explanation is that the addition of [−back] automatically generates a place node. The second issue is whether palatalized sounds occur outside of the contexts described here. If they do, it will confirm that these are not clusters, but single segments. Wonderly is silent on this question.

What is interesting about these cases is that languages with consonant mutations often combine both featural and segmental affixation. That is, the featural affix may occur by itself or with additional segments. In this section, we examine two cases, Nuer and Seereer Siin. Nuer is suffixal and Seerer Siin is prefixal.

3.1 Nuer mutation

The consonant mutation process of Nuer, a Nilo-Saharan language of Sudan, presents an interesting contrast to Chaha, in that the featural suffix must be realized at the very right edge of the verb stem rather than anywhere else in the stem. If the featural suffix cannot be realized on the last consonant of the stem due to a co-occurrence constraint, it is simply not realized at all (see Chaha palatal prosody). But our interest in Nuer mutation is that the suffixes do not just consist of features, but of segments and features.

In the Nuer verb roots, final consonant mutation is associated with various tenses and aspects in the verbal paradigms, as the following examples illustrate. The alternation is only productive in verbs and not in nouns. (All data presented here are from Crazzolara 1933: 156–160; see Frank 1999 for more details on Nuer morphology.) Rule (6) summarizes the observed consonant alternations and (7) provides examples. In the following examples each place of articulation is represented by two verb paradigms. I have converted Crazzolara's representations to IPA, following his descriptions.

(6) *Nuer final consonant alternation* (Crazzolara 1933; Lieber 1987)[9]

	labial	interdental	alveolar	palatal	velar
voiced	b	ð	d	ɟ	ɣ
voiceless continuant	f	θ	r̥	ç	h
voiceless stop	p	t̪	t	c	k

(7) *Verbal paradigms*[10]

a.	*Labial final verbs*	'to overtake a person'	'to scoop (food) hastily'
	3rd sg indic pres act	cóbɛ́jɛ̈	kɛ́bɛ́ jɛ̈
	1st pl indic pres act	cɔ̀ɔfkɔ̀ jɛ	kɛ̀afkɔ̀ jɛ
	pres pple neg	cɔ̀p	kɛp
	past pple	cof	kɛ̀f

9 Following Crazzolara's descriptions, his transcriptions have been modified as follows. [dh] and [th] (interdental) are transcribed here as [ð] and [θ] respectively. [ṛ] (a trilled alveolar continuant) is [r̥]. Finally, [y] (palatal fricative) is retranscribed as [ɟ]. Cazzolara suggests that what he writes as [b] is actually the continuant [β] in final position (Crazzolara 1933: 6). One can imagine that the same is true for what he writes as [d], since he notes that Nuer stems can have up to three forms, one ending with a voiceless stop, one with a voiceless continuant, and the third with a voiced sound which in most cases is a continuant.

10 I will not discuss the vocalic changes, since they are largely unpredictable from Crazzolara's transcriptions.

b.	*Interdental final verbs*	'to suck'	'to wade'
	3rd sg indic pres act	lóðɛ̀ jɛ̀	jǽðɛ́ jɛ̀
	1st pl indic pres act	loɔ̀θkɔ̀ jɛ	jǽθkɔ̀ jɛ
	pres pple neg	lot̪	jæt̪
	past pple	loθ	jæθ
c.	*Alveolar final verbs*	'to sharpen'	'to cut a point'
	3rd sg indic pres act	paádɛ̀ jɛ́	wɨ́lɛ́ jɛ̀
	1st pl indic pres act	páar̥kɔ́ jɛ̀	wər̥kɔ̀ jɛ
	pres pple neg	paat	wɨt
	past pple	pàar̥	wɨr̥
d.	*Palatal final verbs*	'to hit'	'to dismiss a person'
	3rd sg indic pres act	jáaʝɛ̀ jɛ	ɟjéeʝɛ̀ jɛ
	1st pl indic pres act	jáaçkɔ́ jɛ̀	ɟjáaçkɔ́ jɛ̀
	pres pple neg	jaac	ɟjèec
	past pple	jaaç	ɟjeeç
e.	*Velar final verbs*	'to throw away'	'to find'
	3rd sg indic pres act	ɟǽɣɛ́ jɛ̀	jə́ɣɛ́ jɛ̀
	1st pl indic pres act	ɟæ̀kɔ̀ jɛ	jəkɔ̀ jɛ
	pres pple neg	ɟæk	jɔk
	past pple	ɟæh	jəh

First, Crazzolara (1933: 102) notes that the verb root is monosyllabic in Nuer. Second, all verbs begin and end in consonants. I assume, following Lieber (1987), that the features implicated here are [continuant] and [voice]. I will also assume that the morphemes involved in the mutation consist of the following inputs.[11]

(8) *The Nuer suffixes*

indic pres act	=	[jɛ]
3rd sg	=	[cont] [ɛ]
1st pl	=	[cont] [kɔ]
pres pple neg	=	Ø
past pple	=	[cont]

The most important illustration of the theme of this section is the past participle morpheme, which under any analysis must include the feature [continuant], and the 1st plural morpheme, which, in addition to the feature [continuant], also includes the segment sequence [kɔ]. A comparison of all the past participle forms with the 1st plural indicative present active forms shows that the latter always include the additional [kɔ]. What is interesting is that the suffix [kɔ] also triggers spirantization of the preceding stop.[12] Therefore we must assume that this suffix has a preceding floating [continuant]. Finally, we must assume that Nuer also has intervocalic voicing, as seen in all the 3rd singular forms.

[11] But see Lieber (1987) for a different assumption on input.

[12] In the case of the forms 'throw away' and 'find', there is no spirantization. I assume that this is because the final consonant of the verb and the 1st plural suffix [kɔ] are identical. Crazzolara apparently transcribes the unspirantized sequence [hk] as a single stop [k].

It is clear from the mutation cases in Nuer that the features involved are suffixes, since in two cases the free feature [continuant] is paired with traditional segmental suffixes. In the case of the past participle morpheme, the entire content of the morpheme is the free feature [continuant]. [continuant] is licensed by a root node in Nuer. This feature links to the rightmost consonant of the verb. This association formally defines the past participle morpheme in Nuer as a suffix. This morpheme happens to have just a single featural content [continuant].

The derivation of all the forms with the past participle suffix [continuant] is the same as that of the 1st plural forms, except for the additional segments [kɔ] in the suffix. This example actually shows that if we call the segments [kɔ] a suffix, we must treat the preceding [continuant] the same way, since they, together, mark the same morpheme. And if this feature [continuant] of the 1st plural is a suffix, so is the feature [continuant] that marks the past participle alone.

Crazzolara notes that a number of segments do not undergo this mutation process in Nuer. These segments are the nasals /m n n̪ ɲ ŋ/, the liquids /l r/, and glide /w/. I will split these segments into two groups, the nasals on the one hand and the liquids and glide on the other.

I suggest that the nasals do not undergo mutation because of a co-occurrence constraint forbidding the association of [continuant] to a consonant specified for [nasal]. The examples in (9) illustrate this.

(9) *Non-alternating final consonant*

	'to see'[13]	'to hear'
3rd sg indic pres act	nɛ́ɛnɛ̈̀ jɛ	lḭ̈ŋɛ́ jɛ̂
1st pl indic pres act	nɛ́ɛ̈ankɔ́ jɛ̈̀	lie̤ŋkɔ̂ jɛ
pres pple neg	nɛ̂ɛn	liŋ
past pple	nɛ̂ɛn	liŋ

Since morphemes with final nasals never alternate, and since [continuant] does not show up anywhere else, we must assume that in these cases [continuant] must remain unrealized (i.e. unparsed). This is parallel to the case of the non-realization of [round] in Chaha.

I assume that the remaining sonorants, liquids, and glide undergo the process, though the surface forms appear invariant; i.e. [continuant] links vacuously to stems whose final consonants belong to this class, but without any apparent surface effect, since they are already continuants.[14]

In conclusion, [continuant] in Nuer provides a significant contrast to labialization in Chaha and palatalization in Zoque. In both Chaha and Nuer, the featural affix is a suffix, given the insistence on linkage to the final consonant. In both languages, the featural content of the affix cannot co-occur with a class of segments.

[13] Crazzolara (1933: 124) points out that there is a separate negative particle /ci/, which occurs before the subject clitic. Forms with nasals are the only complete paradigms that Crazzolara gives, and in these cases he provides no forms in which the first consonant is an oral stop and the second is a nasal. In all the other forms where the stem consonant does not alternate he provides the 3rd singular indicative present active and the 1st plural indicative present active for the rest of the cases.

[14] This implies that a single [continuant] specification on the final consonant on the surface corresponds to two in the input. See also the discussions of Zoque palatalization (§2.2) and Edoid tone (§4.1) for similar characteristics.

This results in the non-realization of the featural suffix on the final segment. This fact is captured by the co-occurrence constraints between the feature content of the affix and the feature content of the class of segments. Thus it is co-occurrence constraints that force featural affixes from edges. The substantive difference between the two languages is seen in Chaha's insistence on realizing the featural suffix on other segments even if it cannot be realized on the edgemost segment, while Nuer will not realize the featural suffix at all.

It is important to note that other languages with consonant mutation have been identified in the literature, e.g. Fula (Arnott 1970) and North Atlantic languages (Mc Laughlin 2000, 2005), which confirm the above analysis of Nuer mutation. These languages also differ significantly from Nuer. I will briefly discuss the case of one, Seereer Siin (Mc Laughlin 2000).

3.2 *Seereer Siin consonant mutation*

In her work on several Northern Atlantic languages of Niger Congo (Pulaar, Seereer Siin, Wolof), Mc Laughlin (2000, 2005) argues that *consonant mutation* can be viewed and accounted for as the prefixation of a floating feature to the root node of the stem-initial consonant.[15] She proposes a constraint-based account to locate the feature on the left edge of a word.

Seereer Siin consonant mutation is morphologically conditioned by *noun class* in nouns and dependent adjectives, and by number in verbs. There are two patterns of consonant mutation in Seereer: (a) voicing mutation, and (b) continuancy mutation. In each, there is a three-way homorganic range of alternations, called *grades* (Arnott 1970). I will only discuss the voicing mutation, and I will discuss only the fully mutating forms. The reader is referred to Mc Laughlin's work for the partially mutating forms, and the continuancy mutation.

In the voicing mutation, the three grades are "voiced stop," "voiceless stop," and "prenasalized voiced stop." Grade-a refers to the voiced set, grade-b to the voiceless set, and grade-c to the prenasalized set.

Seereer Siin has sixteen noun classes. Of the sixteen, classes 2, 3a, 5, 7, 8, and 10 condition the a-grade mutation, while classes 3b, 6, 12, 13, and 14 condition the c-grade mutation. The remaining classes (1, 4, 9, 11, and 15) condition the b-grade mutation.

I will now illustrate the above statements with the examples in (10), from Mc Laughlin (2000: 339–340). The numbers in parenthesis beside the forms indicate the noun classes of the forms.

(10) *Voicing mutation* (fully mutating)

voiced	*voiceless*	*nasal*[16]	
a-grade	b-grade	c-grade	
ogac (10)	akac (4)	foŋgac (13)	'stone'
ɟir (5)	acir (4)	aɲɟir (3b)	'illness'
oɓaj (10)	xaƥaj (11)	foƥaj (13)	'hand, arm'

[15] For reasons of space, only a brief summary of the facts of Seereer Siin is given here.

[16] Voiceless implosives cannot be prenasalized in Seereer Siin.

I follow Mc Laughlin in assuming that the b-grade forms constitute the "underlying" forms in the stems with voicing mutation. The stem patterns show that the features involved in the class prefixation are [+voice] and [+nasal]. [+voice] drives the voicing of underlying voiceless-initial stems, which are fully mutating. In addition, one must conclude that the class 10 prefix has both segmental and featural contents: /o [+voiced]/, as Mc Laughlin does. Finally, class 13 also has both segmental and featural contents: /fo [+nasal]/. "There is a [+voice] floating feature that drives the a-grade mutations . . . and there is a [+nasal] floating feature that drives the c-grade mutations" (Mc Laughlin 2000: 340).

Comparing the Seereer Siin forms with those from Nuer, the mutating consonants in Seereer Siin are the stem-initial consonants. The mutating features [+voiced] and [+nasal] are prefixes. They must link to the stem-initial consonant and no other. In Nuer on the other hand, the mutating feature is a suffix. The system in Seereer Siin sometimes includes featural affixes alone, and sometimes featural affixes as well as segmental affixes. As seen above, the class 10 prefix includes both segmental and featural content: /o [+voiced]/, and the class 13 prefix also has both segmental and featural content: /fo [+nasal]/. These combinations are in fact more apparent than the Nuer combinations. The segmental features causing the mutation either get associated or not, and are never pushed inwards in the stem. They only occur at the edges.

4 Harmony: Featural affixes with stem domains

The third set of case studies consists of languages that combine featural affixes with featural harmony. By "harmony," I mean featural propagation that is domain-based.

The domain of a featural affix is often the entire stem. By definition, we must take these features to be affixes, since they are the featural spell-out of some morphological category. Since the domain of the featural affix is the entire stem, I take the phenomenon to be the combination of a featural prefix or suffix, plus harmony involving the feature in question. I will illustrate with two languages. I will discuss one case involving a featural suffix (Edoid tone), and one involving a featural prefix (Terena nasalization).

4.1 Edoid associative construction

Tonal data from Edoid languages (Niger Congo, Nigeria) provide the first example of featural suffixation plus harmony. Suffixation is detectable from the fact that priority is given to right alignment, and harmony is seen in the transmission of the feature throughout the entire domain.

In several Edoid languages the "associative morpheme" is a free (floating) High tone. The list includes Etsako (Elimelech 1976), Yekhee (Elugbe 1989), Bini (Amayo 1976), Isoko (Donwa 1982), and Emai (Egbokhare 1990). In this section I will only examine Etsako (Ekpheli dialect). Several other Edoid languages have similar tonal systems to that of Etsako.

Etsako is a two-tone language, with High and Low tones (Elimelech 1976: 41). (Recall that full specification is assumed in this chapter.) In this language, the associative High tone links to the head noun, replacing all Low tones in a right-to-left manner, until it reaches a segmental High tone. The examples below consist of disyllabic nouns, but they are representative of what happens in longer forms. The forms cited here (from Elimelech 1976: 55) exhaust all possible tonal combinations of disyllabic nouns. The tones in the first row in each of (11)-(14) indicate the underlying tone pattern of the head noun in isolation, and the corresponding tones after the arrow indicate its tone pattern in an associative construction. For clarity, I have indicated the tonal pattern of the first example in each set with the tone letters H and L, in addition to the tone marks. The crucial tones to focus on are those of the first noun, since the tones of the second noun remain constant.

(11)		L			H	
	a.	àmɛ̀	èθà	→	ámɛ́èθà	[ámɛ̂θà]
		L (H)	L		H L	
		water	father			'father's water'
	b.	àmɛ̀	òké	→	ámɛ́òké	[ámɔ̂kê][17]
		water	ram			'a ram's water'
	c.	ámɛ̀	ɔ́mɔ̀	→	àmɛ́ɔ́mɔ̀	[ámɔ́mɔ̀]
		water	child			'a child's water'
	d.	àmɛ̀	ɔ́dzi	→	ámɛ́ɔ́dzi	[ámɔ́dʒi]
		water	crab			'a crab's water'

(12)		H L			H(H)	
	a.	únɔ̀	èθà	→	únɔ́èθà	[únêθà]
		H L (H)	L		H L	
		mouth	father			'father's mouth'
	b.	únɔ̀	òké	→	únɔ́òké	[únɔ́kê]
		mouth	ram			'a ram's mouth'
	c.	únɔ̀	ɔ́mɔ̀	→	únɔ́ɔ́mɔ̀	[únɔ́mɔ̀]
		mouth	child			'a child's mouth'
	d.	únɔ̀	ɔ́dzí	→	únɔ́ɔ́dzí	[únɔ́dʒî]
		mouth	crab			'a crab's mouth'

[17] At the phrasal level, a phrase-final High tone is realized as a fall, hence the final falling tones in forms with underlying final Highs such as (11b), (11d), etc.

(13)		H			H	
	a.	ódzí	èθà	→	ódzíèθà	[ódʒêθà]
		\/	\/		\/ \/	
		H (H)	L		H L	
		crab	father			'father's crab'
	b.	ódzí	òké	→	ódzíòké	[ódʒôkê]
		crab	ram			'a ram's crab'
	c.	ódzí	ɔ́mɔ́	→	ódzíɔ́mɔ̀	[ódʒɔ́mɔ̀]
		crab	child			'a child's crab'
	d.	ódzí	ódzí	→	ódzíódzí	[ódʒódʒî]
		crab	crab			'a crab's crab'

(14)		L H			L H	
	a.	ɔ̀tɛ́	èθà	→	ɔ̀tɛ́èθà	[ɔ̀têθà]
		| |	\/		| | \/	
		LH (H)	L		LHL	
		cricket	father			'father's cricket'
	b.	ɔ̀tɛ́	òké	→	ɔ̀tɛ́òké	[ɔ̀tôkê]
		cricket	ram			'a ram's cricket'
	c.	ɔ̀tɛ́	ɔ́mɔ̀	→	ɔ̀tɛ́ɔ́mɔ̀	[ɔ̀tɔ́mɔ̀]
		cricket	child			'a child's cricket'
	d.	ɔ̀tɛ́	ódzí	→	ɔ̀tɛ́ódzí	[ɔ̀tódʒî]
		cricket	crab			'a crab's cricket'

The tone changes on the head noun in associative constructions may be summarized descriptively as follows:

(15)
a. L → H (11)
b. H L → H H (12)
c. H → H (13)
d. L H → L H (14)

In (11) we assume there is a single Low tone associated with both syllables (moras) of the noun, following the Obligatory Contour Principle (Leben 1973; McCarthy 1986). The associative High tone replaces this underlying Low tone, and this Low tone itself is not realized on the surface. That the assumption made here with disyllabic forms is true of longer forms is confirmed by the trisyllabic examples in (16), where the three syllables of the head noun are now realized on a High tone in the associative constructions. Therefore all adjacent Low tone syllables become High regardless of the number of syllables.

(16)
```
L                           H
àɣòɣò      ɔ̀ké      →     áɣóɣóɔ̀ké   [áɣɔ́ɣɔ̂kê]
 \ | /      | |             \ | / | |
   L   (H)  L H               H   L H
skull      rain                        'a ram's skull'

àjɛ̀jɛ̀      èθà      →     ájɛ́jɛ́èθà    [ájɛ́jêθà]
butterfly  father                      'father's butterfly'
```

In (12) (with the HL pattern), the final Low tone of the head noun becomes High. Given the forms in (16), we assume that any number of adjacent syllables with Low tones will become High. Therefore we predict that HLL head nouns will be realized as HHH. This prediction cannot be confirmed, because our sources do not have any examples with such patterns. The forms in (13) are unremarkable, since the head noun is underlyingly High-toned. Finally, in (14), underlying LH remains the same. Our assumption here is that the associative High tone links vacuously to the final syllable of the head noun, just as [−back] links to palatal consonants in Zoque.

The above facts can be analyzed as follows. Following Elimelech I assume that "the associative marker (AM) . . . is underlyingly a High floating tone" (Elimelech 1976: 42). Tone is licensed by any mora in Etsako. Only vowels and syllabic nasals can be moraic in this language. Based on the facts in (11)–(15) above (especially (14)), as well as on facts presented in the Edoid studies cited at the beginning of this section, I suggested that the associative High tone is a featural suffix. It is suffixed to the head noun. However, a (separate) process of tonal harmony transmits the associative High tone throughout the entire head noun. Therefore the domain of the associative High tone is the entire head noun, a prosodic word (Nespor and Vogel 1986; Selkirk 1986; McCarthy and Prince 1990).

This type of phenomenon must be handled with two constraints. One is a morphological alignment constraint, the type of which we have seen so far. This alignment places the featural affix at a particular edge of the stem, characterizing it as a prefix or as a suffix (see Kirchner 1993; Pulleyblank 1993, 1996; Akinlabi 1994, 1997; Cole and Kisseberth 1994). The second is phonological feature spread: harmony. This handles feature propagation by establishing the fact that the domain of the feature is a phonological category, such as the prosodic word.

It is crucial to note that the associative High tone is different from an underlyingly linked segmental High tone of a head noun (the segmental High tone). First, while the associative High tone is a morpheme, the segmental High tone is not. And second, the segmental High tone is underlyingly linked, while the associative High tone is underlyingly free, i.e. it belongs to a morpheme with no other content. Any analysis of Etsako must recognize these differences.

4.1.1 *H-tone opacity*

In Etsako, the segmental High tone is "opaque": it blocks the propagation of the suffixal High tone. That is, the suffixal H tone cannot spread through the lexical H tone. The examples in (14) demonstrate this fact. In the LH head nouns, the output associative construction begins as LH, which does not become HH, as one would expect if the suffixal H tone were to spread through the segmental H tone. This indicates two things. First, only the suffixal H tone spreads, while the segmental H tone does not; otherwise we would once again have HH on the head noun in the output. Second, the segmental H tone is opaque to the spread of the suffixal H tone. We must assume that the constraint responsible for the association of a segmental H supersedes the one for tone spreading.

4.1.2 *German sign language*

Pfau (2000) has found a parallel of this type of affix in an unexpected place, the negative morpheme in the German Sign Language (DGS).[18] Pfau proposes an analysis of the negative headshake of DGS as an autosegment, in other words as a featural affix [headshake], which is associated with a manual form. The negative headshake, he argues, behaves in a way similar to tonal prosodies in tone languages. He proposes that this feature represents the negative morpheme, in the same way as tone functions as a grammatical morpheme associated with an entire base.

The main goal of §4.1 has been to show, first, that the domain of a featural affix may be the whole lexical category, but that it can still be identified as a prefix or suffix. Second, the featural affixation, unlike segmental affixation, may combine with harmony involving the feature itself. In §4.2, we show that non-tonal featural affixes also behave the same way, using Terena nasalization as illustration.

4.2 *Terena nasalization*

The second example of a system that combines featural affixation with harmony is Terena. In this section I argue that the feature [nasal] in Terena is a featural prefix, given the insistence on association to the initial consonant of the stem (in direct contrast to the Edoid associative High tone), and that the featural prefixation is accompanied by harmony. Terena also confirms the accounts already given in the preceding sections about both featural alignment and misalignment. In contrast to the Edoid associative marker, the lexical feature [nasal] is transparent to the propagation of the featural affix [nasal] (CHAPTER 78: NASAL HARMONY).

In Terena, an Arawakan language of Brazil (Bendor-Samuel 1960, 1966), the category of the 1st person is marked through a process of progressive nasalization. Thus the difference between the Terena examples in the first and the third columns is that the latter are marked for the 1st person.

(17) *1st person in Terena*

a.	ajo	'his brother'	àj̃õ	'my brother'
	arine	'sickness'	ãr̃ĩnẽ	'my sickness'
	unae	'boss'	ũnãè	'my boss'
	emoʔu	'his word'	ẽmõʔũ	'my word'
b.	owoku	'his house'	õw̃õᵑgu	'my house'
	iwuʔiʃo	'he rides'	ĩw̃ũʔĩⁿʒo	'I ride'
	ituke	(POSS PRON)	ĩⁿduke	(1PERS POSS PRON)
	nokone	'need'	nõᵑgone	'I need'

[18] I am deeply indebted to a reviewer for helping to make sense of this section.

c.	taki	'arm'	^{n}daki	'my arm'
	tuti	'head'	^{n}duti	'my head'
	paho	'mouth'	^{n}baho	'my mouth'
	piho	'he went'	^{m}biho	'I went'
d.	ahjaʔaʃo	'he desires'	ã^{n}ʒaʔaʃo	'I desire'
	haʔa	'father'	^{n}zaʔa	'my father'
	hjiʃoe	'dress'	^{n}ʒiʃoe	'my dress'

The descriptive generalizations from the above data are as follows. The 1st person pronoun is expressed by nasalizing the noun or verb. Nasalization affects vowels, liquids, glides, and underlying nasal consonants. Therefore, nasalization spreads through underlying nasal consonants. Laryngeal stops, but not laryngeal fricatives, are affected by nasalization. That is, nasalization may spread through a laryngeal stop, but not through a laryngeal fricative.

The examples in (17b) show that nasalization proceeds in an apparent left-to-right fashion until it reaches an obstruent. The interesting thing here is that the obstruent becomes prenasalized (and voiced), as in the first example in (17b), but nothing after it is nasalized (except of course it is an underlying nasal consonant, as in the last example in (17b)). Therefore obstruents block [nasal] spreading, but not before they become prenasalized. If a form begins with an obstruent, the effect of the 1st person progressive nasalization is to turn that obstruent into a prenasalized consonant, as in (17c), and there is no nasalization of subsequent segments. I shall not be concerned with further changes in obstruents, other than prenasalization. For example, I shall not discuss the fact that laryngeal continuants change to coronals when nasalized in (17d).

Continuing the discussion in the preceding sections, an analysis of the above Terena facts may be presented as follows. The 1st person marker is a free feature [nasal]. [nasal] can be associated with any root node in Terena, consonant or vowel. Given the insistence on associating to the first segment of the noun or verb regardless of the nature of the segment, it is a featural prefix. However, a process of harmony transmits nasality from the prefix through the stem; and thus the apparent domain of the [nasal] morpheme is the entire stem, which is a prosodic word. The surface realization of this morpheme may be accounted for the same way as tone in Etsako. An alignment constraint places [nasal] as a prefix, while a feature spread constraint accounts for spreading to the end of the word.

Just like the High tone in Etsako, [nasal] is both the featural content of a morpheme and a lexically contrastive feature in Terena. These two functions must be recognized by any analysis.

4.2.1 *Nasal transparency*

Forms like /arine/ → [ãr̃ĩnẽ] in (17a) reveal that nasal stops do not block the propagation of the [nasal] morpheme in Terena, i.e. underlying nasal stops are transparent to the morphemic [nasal] spread. Our account of this transparency is that (the constraint responsible for) the domain of association of the [nasal] morpheme takes precedence over the segmentally specified [nasal], and could therefore pass "over" the segmentally specified [nasal].

This constitutes an important difference between the underlying segmental High tone in Edoid (as exemplified by Etsako) and the segmentally specified [nasal] in Terena. While the segmental High tone in Edoid blocks the propagation of the morphemic High tone, the propagation of the [nasal] morpheme in Terena is not blocked by the segmentally specified [nasal].

4.2.2 Obstruents and co-occurrence

We now turn to account for the behavior of obstruents in Terena. As noted above, obstruents block the rightward propagation of the [nasal] morpheme, while becoming prenasalized: /owoku/ → [õw̃õ^{ŋ}gu] 'my house'. To account for this, we assume a co-occurrence constraint forbidding the co-occurrence of [−sonorant] and [+nasal] in Terena (see Pulleyblank 1989: 109). Note however that, while nasality is always barred from obstruents in general (as in Orejon; Pulleyblank 1989), Terena obstruents are partly nasalized. We can account for this by assuming that nasality is barred from the release phase of obstruents in Terena, but not from closure phase (Steriade 1993). Prenasalization in Terena can be seen as the association of the [nasal] morpheme to the closure phase of the obstruent stops, and not to the release phase.

Finally, though, the domain of the [nasal] morpheme is the entire stem (a prosodic word), like the High tone in Etsako; it is formally a featural prefix, in contrast to Etsako, where H is a featural suffix.

Gerfen (1999: 127–131) describes an interestingly similar case in Coatzospan Mixtec. In this language, the 2nd person familiar is marked by a [nasal] feature. As in Terena, the entire base is nasalized. However, unlike in Terena, the free feature [nasal] is a suffix, because the spreading is from right to left. Furthermore, if spreading is blocked, only the final vowel of the base is nasalized, indicating that the feature [nasal] links to the final vowel. Spreading is blocked when the final syllable has a voiceless obstruent onset. Finally, like in Terena, lexical nasal consonants are transparent to nasal spread.

5 Segmental realization: Mafa imperfective

In our fourth case study, the featural affix is at the same time a "feature" and a "segment." I refer to this as segmental realization of a featural affix. The case is exemplified by palatalization in Mafa. This language is interesting because of its unique morphological properties. The affix expressing the imperfective in Mafa can be characterized both as a segmental affix and as a featural affix at the same time.[19] This allomorphy gives languages like this a special place in the study of featural affixes.

Ettlinger (2003, 2004) describes the morphosyntactic process of imperfective aspect formation in Mafa, a central Chadic language of Cameroon, as follows. The imperfective is formed in one of two ways, depending on whether the final segment of the root is a vowel ([a]) or a consonant. In the case of verbs ending in [a], /j/ is suffixed to the base, as seen in (18). (All vowel-final verb stems end in an /a/ and all other suffixes are positioned after the imperfective suffix.)

[19] Another language with similar properties is Yokuts (Archangeli 1984, 1991; Archangeli and Pulleyblank 1994). In Yokuts, the glottal feature can surface as a segment or as part of another segment (or not surface at all).

(18) *Palatalization of /a/-final verbs*

gudza	'tremble'	gudzaj	'is trembling'
bəra	'insult'	bəraj	'is insulting'
^{n}da	'cut a hole'	^{n}daj	'is cutting a hole'
keða	'divide'	keðaj	'is dividing'

The imperfective of verbs ending in a consonant, however, is formed with a palatal featural suffix. Apparently, the palatal prosody targets either vowels or coronal stridents, and no more. There is one complication, regarding the vowel [u]. [u] is not palatalized (to [y]) in two contexts: (a) when it occurs after a dorsal, and (b) after a coronal strident in a disyllabic root. I will not discuss this complication here. Readers are referred to Ettlinger (2003, 2004) for an explanation.

The vowel inventory of Mafa is given in (19).

(19) *Mafa vowel inventory*

i	y	ə	u
e	œ	a	o

The surface realizations of vowels under palatalization are as follows:

(20)
/ə/ → /i/
/u/ → /y/
/o/ → /œ/
/a/ → /e/

The forms in (21a) represent monosyllabic verb roots, and those in (21b) represent disyllabic forms. The last two forms in (21) show that both vowel and coronal stridents can be palatalized, if both are present in the verb root. In the forms in (22), palatalization appears to skip some segments, while other segments are palatalized (Ettlinger 2004). This is not skipping. The skipped segments are not licensors (Akinlabi 1996) of the palatal prosody in Mafa, hence the apparent skipping.

(21) a. *Palatalization of monosyllabic consonant-final verbs*

pan-	'wash'	pen-	'is washing'
təv-	'light (VB)'	tiv-	'is lighting'
dad-	'add water to'	ded-	'is adding water to'
guts-	'squirt'	gutʃ-	'is squirting'
tsap-	'speckle'	t̪ʼep-	'is speckling with clay'
sur-	'sleep with a woman'	ʃyr-	'is sleeping with a woman'

b. *Palatalization of disyllabic consonant-final verbs*

səban-	'work'	ʃiben-	'is working'
lubat	'twist'	lybet	'is twisting'
suwdək	'miss'	ʃuwdik	'is missing'

c. *No palatalization*

gum-	'carve wood'	gum-	'is carving wood'
gud-	'search with anxiety'	gud-	'is searching with anxiety'
kurk^{w}-	'carve everywhere'	kurk^{w}-	'is searching everywhere'

Following the way featural affixes work, there is no doubt that the imperfective is a featural suffix in Mafa (Akinlabi 1996), as the vowel-final verbs show. It scans the verb root in a right-to-left manner. If the last segment of the verb root is a vowel, then the imperfective is a full segment, i.e. a suffix. If the palatal prosody finds a consonant as the final segment, then it seeks out a licensor, preferably a vowel. I assume that the coronal palatalization is just a default. This is because this is the only consonant that can be changed without actually completely changing the primary place of articulation. Finally, I suggested that the vowel [u] is blocked from change after a dorsal consonant because it shares the dorsal specification with the preceding dorsal consonant.

6 Formal insights into featural affixation

In general, there has not been much disagreement about whether features can be affixes or not. What has varied is the formal approach to featural affixes. Much of the formal work on featural affixes has been carried out within autosegmental phonology, which allows for autonomous representation of features (CHAPTER 14: AUTOSEGMENTS). The featural affix is commonly represented as a floating feature, and linked to a segment by some rule. Work done on featural affixes within this approach includes McCarthy (1983), Lieber (1984), and others.

Feature geometry (Clements 1985; Sagey 1986; Clements and Hume 1995; and others; see also CHAPTER 27: THE ORGANIZATION OF FEATURES) has also provided significant insights. For example, feature geometry provides significant insight into the grouping of features, and into why some features co-occur together and others don't. In addition, certain nodes can serve as anchors for some featural affixes while others cannot. Work like Archangeli and Pulleyblank (1994) is situated within this approach.

The formal approach to featural affixation adopted in this chapter is the constraint-based Optimality Theory (Prince and Smolensky 1993). Within this theory, grammars are composed of hierarchies of ranked and violable universal markedness and faithfulness constraints. In the theory, faithfulness constraints monitor input and output to ensure that they are the same, and markedness constraints ensure that output structures are unmarked to the highest degree possible, depending on the conflict between all markedness and faithfulness constraints.

However, there are various approaches to featural affixes within Optimality Theory itself. Variations include Zoll's (1998) subsegmental approach, which proposes that the input and output correspondence of "subsegments," including "floating features" and latent segments (undominated F-element), is monitored by MAX(subseg) (see Lombardi 1998 for similar MAX(F)), stated as in (22).

(22) MAX(subseg)

Every subsegment in the input has a correspondent in the output.

As Zoll (1998: 44) notes, featural affixes are realized as part of other segments, therefore the correspondence relation returns the output segment that hosts the feature, not the feature itself. If that is the case, Mc Laughlin (2000) argues that, since subsegments do not occur as output forms, there is no evidence for positing a DEP constraint of the sort DEP(subseg). She proposes that we employ IDENT-IO(F) to monitor subsegments in general. This may be stated as in (23):

(23) IDENT-IO(F)

Correspondent IO segments have identical values for the feature F.

Kirchner (1993), Akinlabi (1996), and Zoll (1996) suggest that features are subject to the same kind of alignment, or coincide constraints as segments. Akinlabi (1996) suggests that featural affixes are subject to the same kind of alignment constraints as non-featural morphemes. He proposes that alignments constraints account for the determination of featural affixes as prefixes or suffixes. All featural affixes, he proposes, are subject to the featural alignment in (24) (see McCarthy and Prince (1993a, 1993b). The specific morphological alignment constraint in (25) accounts for Chaha labialization (Akinlabi 1996: 246).

(24) *Featural alignment*

ALIGN(PFeat, GCat)
A prosodic feature is aligned with some grammatical category.

(25) ALIGN-3MASC-SG

ALIGN (3MASC SG, R; Stem, R)
The right edge of 3MASC SG must be aligned with the right edge of the stem.
3MASC SG is a suffix in stem.

A constraint like (25) does not say whether 3MASC SG is a segment or a feature; it simply refers to the morphological category. Therefore it should not matter whether 3MASC SG is a feature or a segment. As Akinlabi (1996: 243) points out, PFeat (in (24)) is simply the featural spell-out of the morphological category in question.

Misalignment of featural affixes is controlled by feature co-occurrence constraints (Archangeli and Pulleyblank 1994). An example of this is *NASCONT (Akinlabi 1996: 254), which forbids nasal consonants from be continuants.

(26) *NASCONT

If [nasal] then not [continuant].

The above represents the core of the grammar of featural affixes. The variations are derived from ranking the constraints. This analysis also represents the point of departure for some scholars.[20]

[20] Piggott (2000) argues against the idea that features can align to word edges, like segments. He sees featural alignment as proposed by Akinlabi (1996) as an overly powerful mechanism. He proposes instead that morphological alignment be supplemented by a provision for prosodic licensing, so that, for example, features may be incorporated into a prosodic category such as a foot or a prosodic word. See McLaughlin (2000: 341–345) and Horwood (2004) for answers to Piggott's objections. Another notable counterposition is that of Kurisu (2001), who proposes a "relational morphology theory" instead of "featural alignment." I will not discuss this here, since it is an entirely different theory.

7 Are featural affixes really featural?

I will end this discussion by examining whether "featural affixes" are really "featural." We can examine this issue from the theoretical and the empirical points of view. The traditional view of an affix is that of a "whole segment" (or segments), which marks a morphological category. The affix is dependent on, or attached to, some host, a base. The category represented could be inflectional or derivational. By segment is traditionally meant a unity of several articulatory gestures that are produced simultaneously and that paradigmatically contrast with one another. By this definition [t] and [s] are segments in *tip* and *sip*. [t] and [s] are also segmental affixes (suffixes), representing the English past tense, and 3rd person singular verbal agreement in [sækt] *sacked* and [sæks] *sacks*, respectively. In this definition of an affix, it represents a timing slot or more in the paradigmatic string.

Featural affixes on the other hand, from the cases that we have been discussing, do not always occupy a timing slot. Rather they share the same time slot with one or more of the segments in the base. For example, in Zoque palatalization (Wonderly 1951) (§2.2), palatalization simply changes an alveolar consonant to a palatal ([s] → [ʃ], in [sʌk] 'beans', [ʃʌk] 'his beans'), yet that difference signifies the distinction between 'beans' and 'his beans'. In some cases, it in fact makes no sense to talk about timing slots in the string. Such is the case in Terena nasalization (Bendor-Samuel 1960, 1966) and in Mafa palatalization (Ettlinger 2003, 2004), where the featural affix attaches to more than one segment of the base. In Mafa [lubat] 'twist', [lybet] 'is twisting', the palatal feature is attached to both vowels in the base.

But even with these facts there are problems about what a featural affix really is. The problem is those features that can be realized as full segments as well as as features. These include palatalization, labialization, nasalization, and glottalization. Note that, in Mafa and languages like it, the palatal feature can be realized as a full segment [j], when the verb is vowel-final. Does this, then, mean that this is both a segmental affix and a featural affix? Or is it a featural affix that is sometimes realized as a full segment? Mafa is intriguing because, on any account, it would satisfy the definition of a segmental affix as well as that of a featural affix. The same applies to nasalization in Seereer Siin (Mc Laughlin 2000). It is easy to assume that the nasal feature in all these cases is a full segment. However, certain features are never realized as full segments. These include voicing and continuancy. There is no other way that I know of than to analyze the feature [continuant] as a featural affix marking the past participle in Nuer (Crazzolara 1933) (§3.1).

From the theoretical point of view, this question relates to the way a "segment" is defined. If segments (or feature bundles) are the contrastive elements in a language, such that the meaning contrast between [tip] and [dip] is seen as represented by the first consonants [t] and [d] in these words, rather than by the fact that [t] and [d] differ only in that [d] is voiced and [t] is not, then there *are* featural affixes, because the elements that represent featural affixes are "less than" segments, as the empirical facts above reveal.

On the other hand, the current assumption is that the contrastive elements in language are "features," and not "feature bundles." This distinction is captured by feature theory (CHAPTER 17: DISTINCTIVE FEATURES) and the internal organization of segments (e.g. Clements 1985). On this viewpoint, the meaning contrast between [tip] and [dip] is seen as represented by voicing. If we equate minimal units with contrastive units, then there are no featural affixes. The only distinction is between affixes that are feature bundles and affixes that are single features.

Empirical data from the Mafa imperfective aspect (Ettlinger 2003, 2004) suggest that the distinction between a segmental affix and a featural affix may not be real. In this case the same feature, "palatality," sometimes behaves as a feature bundle, and sometimes as a single feature. The importance of the Mafa data is that even the distinction between "single feature" and "feature bundle" may not be real.

8 Conclusions

In summary, in this chapter I have illustrated the characteristics of featural affixes. These features include (a) marking morphological categories (like segmental affixes), (b) occurring as part of other segments rather than independently, (c) varying between prefixes and suffixes, (d) occurring inside the stem (because of feature co-occurrence constraints at edges), (e) spanning the entire base of affixation, and (f) varying occurrence as a feature or a segment in the same language. I have illustrated these with facts from Dutch, Chaha, Zoque, Nuer, Seereer Siin, Etsako, Terena, Mafa, Coatzospan Mixtec, and German Sign Language.

Comparing featural affixes with traditional regular affixes, featural affixes share four characteristics with the traditional affixes: (a) marking morphological categories, (b) varying between prefixes and suffixes, (c) (sometimes) occurring as independent segments, and (d) occurring inside the stem (because of feature co-occurrence constraints at edges). Other characteristics are unique to featural affixes alone: (a) occurring as part of other segments, (b) spanning the entire base of affixation, and (c) varying occurrence a feature or a segment in the same language.

There are a number of important lessons that the unique characteristics of "featural affixes" teach us. First, the so-called "normal affixes" always contain a timing unit, while "featural affixes" do not normally contain a timing unit. Second, they raise the question of whether segments or features are the basic elements that sound systems manipulate. Finally, they reveal that all features are not the same. Some features can be morphemic but can never be realized independently of some other segments ([continuant], [voice]), while other features that are morphemic may dock on some sound in the stem but may also become segments in their own right ([glottal], [nasal], [palatal], [labial]).

ACKNOWLEDGMENTS

I am grateful to two anonymous reviewers for insightful comments, and to the editors of the *Companion* for all their help with this chapter.

REFERENCES

Akinlabi, Akinbiyi. 1994. Alignment constraints in ATR harmony. *Studies in the Linguistic Sciences* 24. 1–18.

Akinlabi, Akinbiyi. 1996. Featural affixation. *Journal of Linguistics* 32. 239–289.

Akinlabi, Akinbiyi. 1997. Kalabari vowel harmony. *The Linguistic Review* 14. 97–138.

Alderete, John & Alexei Kochetov. 2009. Japanese mimetic palatalization revisited: Implications for conflicting directionality. *Phonology* 26. 369–388.

Amayo, Moses A. 1976. A generative phonology of Edo (Bini). Ph.D. dissertation, University of Ibadan.

Anderson, Stephen C. (ed.) 1991. *Tone in five languages of Cameroon.* Dallas: Summer Institute of Linguistics & University of Texas, Arlington.

Archangeli, Diana. 1984. Underspecification in Yawelmani phonology and morphology. Ph.D. dissertation, MIT.

Archangeli, Diana. 1991. Syllabification and prosodic templates in Yawelmani. *Natural Language and Linguistic Theory* 9. 231–283.

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology.* Cambridge, MA: MIT Press.

Arnott, David W. 1970. *The nominal and verbal systems of Fula.* Oxford: Oxford University Press.

Bendor-Samuel, John T. 1960. Some problems of segmentation in the phonological analysis of Terena. *Word* 16. 348–355.

Bendor-Samuel, John T. 1966. Some prosodic features of Terena. In C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (eds.) *In memory of J. R. Firth*, 30–39. London: Longman.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. & John A. Goldsmith (eds.) 1984. *Autosegmental studies in Bantu tone.* Dordrecht: Foris.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 245–306. Cambridge, MA & Oxford: Blackwell.

Cole, Jennifer & Charles W. Kisseberth. 1994. An optimal domains theory of vowel harmony. *Studies in the Linguistic Sciences* 24. 101–114.

Crazzolara, J. P. 1933. *Outlines of Nuer grammar.* Vienna: Verlag der Internationalen Zeitschrift "Anthropos."

Dell, François. 1980. *Generative phonology.* Cambridge: Cambridge University Press.

Donwa, Shirley O. 1982. The sound system of Isoko. Ph.D. dissertation, University of Ibadan.

Egbokhare, Francis. 1990. A phonology of Emai. Ph.D. dissertation, University of Ibadan.

Elimelech, Baruch. 1976. A tonal grammar of Etsako. *UCLA Working Papers in Phonetics* 35.

Elugbe, Benjamin O. 1989. *Comparative Edoid: Phonology and lexicon.* Port Harcourt: University of Port Harcourt Press.

Ettlinger, Marc. 2003. Aspect in Mafa: A case of featural affixation. Unpublished ms., University of California, Berkeley.

Ettlinger, Marc. 2004. Aspect in Mafa: An intriguing case of featural affixation. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 40. 73–86.

Firth, J. R. 1948. Sounds and prosodies. *Transactions of the Philological Society.* 127–152.

Frank, Wright J. 1999. Nuer morphology. M.A. thesis, State University of New York, Buffalo.

Gerfen, Chip. 1999. *Phonology and phonetics in Coatzospan Mixtec.* Dordrecht: Kluwer.

Hamano, Shoko. 1986. The sound-symbolic system of Japanese. Ph.D. dissertation, University of Florida, Gainesville.

Hendricks, Sean. 1989. Palatalization and labialization as morphemes in Chaha. Unpublished ms., Yale University.

Horwood, Graham. 2004. Relational faithfulness and position of exponence in Optimality Theory. Ph.D. dissertation, Rutgers University.

Hulst, Harry van der & Keith L. Snider (eds.) 1993. *The phonology of tone*. Berlin & New York: Mouton de Gruyter.

Idsardi, William J. 1992. The computation of prosody. Ph.D. dissertation, MIT.

Itô, Junko. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7. 217–259.

Itô, Junko & Armin Mester. 1993. Licensed segments and safe paths. *Canadian Journal of Linguistics* 38. 197–213.

Johnson, C. Douglas. 1975. Phonological channels in Chaha. *Afroasiatic Linguistics* 2. 1–13.

Jungraithmayr, Hermann. 1990. *Lexique mokilko*. Berlin: Dietrich Reimer.

Kirchner, Robert. 1993. Turkish vowel disharmony in Optimality Theory. Paper presented at the Rutgers Optimality Workshop 1, Rutgers University.

Kosseke, Dominique & Jérôme Sitamon. 1993. Aka field notes. Unpublished ms., Summer Institute of Linguistics.

Kuipers, Aert. 1974. *The Shuswap language: Grammar, texts, dictionary*. The Hague & Paris: Mouton.

Kurisu, Kazutaka. 2001. The phonology of morpheme realization. Ph.D. dissertation, University of California, Santa Cruz.

Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT.

Lieber, Rochelle. 1984. Consonant gradation in Fula: An autosegmental approach. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*, 329–346. Cambridge, MA: MIT Press.

Lieber, Rochelle. 1987. *An integrated theory of autosegmental processes*. Albany: SUNY Press.

Lombardi, Linda. 1998. Evidence for MaxFeature constraints from Japanese. *University of Maryland Working Papers in Linguistics* 7 (ROA-247).

McCarthy, John J. 1983. Consonantal morphology in the Chaha verb. *Proceedings of the West Coast Conference on Formal Linguistics* 2. 176–188.

McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17. 207–263.

McCarthy, John J. & Alan Prince. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory* 8. 209–283.

McCarthy, John J. & Alan Prince. 1993a. Prosodic morphology I: Constraint interaction and satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University.

McCarthy, John J. & Alan Prince. 1993b. Generalized alignment. *Yearbook of Morphology 1993*. 79–153.

Mc Laughlin, Fiona. 2000. Consonant mutation and reduplication in Seereer Siin. *Phonology* 17. 333–363.

Mc Laughlin, Fiona. 2005. Reduplication and consonant mutation in the Northern Atlantic languages. In Bernhard Hurch (ed.) *Studies on reduplication*, 111–133. Berlin & New York: Mouton de Gruyter.

Mester, Armin & Junko Itô. 1989. Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language* 65. 258–293.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Pfau, Ronald. 2000. The grammar of headshake: Sentential negation in German Sign Language. Unpublished ms., University of Amsterdam.

Piggott, Glyne L. 2000. Against featural alignment. *Journal of Linguistics* 36. 85–129.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology*. Dordrecht: Reidel.

Pulleyblank, Douglas. 1989. Patterns of feature co-occurrence: The case of nasality. In S. Lee Fulmer, Masahide Ishihara & Wendy Wiswall (eds.) *Proceedings of the Arizona Phonology Conference* 2, 98–115. Tucson: University of Arizona.

Pulleyblank, Douglas. 1993. Vowel harmony and Optimality Theory. *Actas do Workshop Sobre Fonologia, University of Coimbra*. 1–18.

Pulleyblank, Douglas. 1996. Neutral vowels in Optimality Theory: A comparison of Yoruba and Wolof. *Canadian Journal of Linguistics* 41. 295–347.

Rice, Keren. 1987. The function of structure preservation: Derived environments. *Papers from the Annual Meeting of the North East Linguistic Society* 17. 501–519.

Roberts, James S. 1994. Nontonal features as grammatical morphemes. *Work Papers of the Summer Institute of Linguistics, University of North Dakota* 38. 87–99.

Rose, Sharon. 1994. Palatalization, underspecification, and plane conflation in Chaha. *Proceedings of the West Coast Conference on Formal Linguistics* 12. 101–116.

Rose, Sharon. 2007. Chaha (Gurage) morphology. In A. S. Kaye (ed.) *Phonologies of Asia and Africa (including the Caucasus)*, 399–424. Winona Lake, IN: Eisenbrauns.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371–405.

Steriade, Donca. 1993. Closure, release, and nasal contours. In Marie K. Huffman & Rena A. Krakow (eds.) *Nasals, nasalization, and the velum*, 401–470. Orlando: Academic Press.

Wonderly, William L. 1951. Zoque II: Phonemes and morphophonemes. *International Journal of American Linguistics* 17. 105–123.

Zoll, Cheryl. 1996. Parsing below the segment in a constraint-based framework. Ph.D. dissertation, University of California, Berkeley.

Zoll, Cheryl. 1998. Positional asymmetries and licensing. Unpublished ms., MIT (ROA-282).

89 Gradience and Categoricality in Phonological Theory

Mirjam Ernestus

1 Introduction

Within phonological theory, important roles are assigned to the notions of "gradience" and "categoricality." The opposition qualifiies sounds and sound patterns, and is crucial both for the definition of the phonological and the phonetic components of generative grammar, and for the development of alternative types of grammatical models. This chapter discusses the assumptions generative phonology and its direct successors (including Optimality Theory) have made about the role of gradience. Moreover, it presents data supporting or contradicting these assumptions, and discusses new models accounting for the conflicting data.

The most important section of this chapter (§2) discusses the opposition between categorical sounds, which are stable and represent clear distinct phonological categories (e.g. sounds showing all characteristics of voiced segments throughout their realizations), and gradient sounds, which may change during their realizations and may simultaneously represent different phonological categories (e.g. sounds that start as voiced and end as voiceless). A shorter section (§3) discusses categorical generalizations over sounds, which are fully productive, and gradient generalizations, which are less productive. The final section (§4) provides a short conclusion.

2 Sounds

2.1 *Gradience in generative grammar*

In the early days of generative grammar, the opposition between categoricality and gradience was assumed to reflect the fundamental distinction between competence and performance. Competence described speakers' categorical knowledge about their language, abstracted away from performance factors such as vocal tract size, working memory span, articulatory effort, and so on. Performance, in contrast, described speakers' actual linguistic behavior, which could be gradient, and was not in the direct focus of linguistic research (Chomsky and Halle 1968; following Saussure 1916).

The distinction between competence and performance was reflected in the distinction between the phonological and phonetic component. The phonological component contained the speaker's competence and thus represented cognition. It was believed to be language-specific and to include the phonemes of the speaker's language and language-specific phonological processes, such as final devoicing (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION) and place assimilation (CHAPTER 81: LOCAL ASSIMILATION). This knowledge was represented in the form of categorical symbols and rules operating on these symbols. Phonetic mechanisms were responsible for the speaker's performance. These phonetic mechanisms were believed to be universal and the automatic results of speech physiology (Chomsky and Halle 1968: 293; Kenstowicz and Kisseberth 1979). They thus reflected physics and included, for instance, nasalization of vowels preceding nasal consonants, palatalization of consonants preceding high vowels, and shortening of vowels preceding voiceless obstruents. Since the phonetic component did not reflect the speaker's competence, it was considered not to be part of the grammar proper.

This view on the phonological and phonetic components changed very quickly, since various studies showed that the exact realization of an abstract symbol (e.g. a phoneme or a phonological feature) might be different in different languages. Moreover, no part of a realization appeared to be the automatic and unavoidable result of speech physiology (e.g. Keating 1985, 1990a; Kingston and Diehl 1994; see also CHAPTER 17: DISTINCTIVE FEATURES). As a consequence, the traditional definition of the phonetic component as containing only universal processes automatically resulting from speech physiology implied that this component was empty. A new distinction had to be developed which was no longer based on the notions of language-specific *vs.* language-universal and non-automatic *vs.* automatic mechanisms.

The now widely accepted definitions of the phonological and phonetic components are completely based on the opposition between categoricality and gradience (e.g. Keating 1988, 1990b; Pierrehumbert 1990; Cohn 1993; Zsiga 1997). The phonological component is assumed to deal with categorical, abstract, stable, timeless symbols, such as phonemes and phonological features. Phonological processes refer to these symbols and consequently have categorical effects: they change one symbol (e.g. [+voice]) into another one ([−voice]), or they delete or insert symbols. Phonetic processes translate the abstract symbols into articulatory and perceptual targets. This may lead to sounds with acoustic characteristics that do not perfectly represent categorical phonological symbols, but rather have intermediate values, for instance, when obstruents are partly voiced due to co-articulation. These definitions of the phonological and phonetic components have been adopted in several psycholinguistic models of speech production and comprehension (e.g. Levelt 1989; Norris 1994).

Since the distinction between gradience and categoricality is crucial in the definitions of the phonological and phonetic components, it has led to many experimental studies. The following subsections discuss their findings and their implications for phonological theory (see also CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATIONS; CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY). The first subsections discuss the domains (assimilation and segment deletion) where the evidence for gradience is most convincing but can also be relatively easily reconciled with generative grammar: the relevant processes traditionally characterized as phonological could be reclassified as phonetic. The following subsections (on incomplete neutralization and phonetic detail) discuss evidence for gradience that is less clear but has important theoretical consequences. Generative grammar cannot account for incomplete neutralization without making additional far-reaching assumptions. Further, evidence for a role for fine phonetic detail in speech processing suggests that words are not lexically represented in the form of abstract phonemes but are stored together with their detailed phonetic properties. These data have stimulated the development of accounts based on assumptions other than those of generative grammar.

2.2 *Assimilation: Data*

One of the first types of processes traditionally characterized as phonological for which researchers found evidence of gradience is formed by connected speech processes, in particular assimilation (CHAPTER 81: LOCAL ASSIMILATION). Nearly all instances of assimilation are traditionally described as the categorical spreading of a phonological feature from one segment to another segment in the phonological component. The receiving segment is assumed to be subsequently identical to segments with the same features in their underlying specifications. For instance, [m] would have exactly the same surface phonological representation and phonetic characteristics if it results from an underlying /m/ and if it results from place assimilation, as in the phrase *gree*[m b]*oat* 'green boat'.

Many articulatory studies have investigated the assumed categorical nature of place assimilation using electropalatography (EPG; Hardcastle 1972), which registers contacts between the tongue and the hard palate, or with the help of an electromagnetic midsagittal articulometer (EMMA; e.g. Perkell *et al.* 1992), which allows the tracking of individual fleshpoints by means of small transducer coils attached to various points on the speaker's vocal tract in the midsagittal plane. These studies have provided evidence for the categorical nature of some place assimilation processes. An example is regressive place assimilation in Korean, which is a characteristic of fast colloquial Korean and affects certain consonants preceding certain other consonants. For instance, the phrase /pat˺p'oda/ 'rather than the field' can be pronounced as [pap˺p'oda]. Kochetov and Pouplier (2008) showed that this assimilation results in the categorical absence of the gestures for the original articulation place of the assimilated consonant (in this example: for /t˺/) in most tokens. Another example is place assimilation of /n/ to /k/ in Italian, which categorically results in the absence of alveolar gestures (Farnetani and Busà 1994).

Other studies strongly suggest that some place assimilation processes are gradient in nature. For instance, assimilation of alveolar obstruents to the palatality of the following segments (as in American English *hi*/t j/*ou*) often does not lead to completely palatal segments ([c] in the example), but rather to segments that become more palatal during their realizations (within one and the same token) and that consequently differ in their phonetic detail from underlying palatals (e.g. Barry 1992 for Russian; Zsiga 1995 for post-lexical palatalization in American English). The same type of gradience has been reported for place assimilation of coronal obstruents in American English, as in *la*/t k/*alls* (*late calls*) produced as *la*[kː]*alls*. The assimilated obstruents often start with a coronal constriction that gradually assimilates to the articulation place of the following obstruent during their realizations (velar in the above example; Nolan 1992). Other gradient place assimilation processes include assimilation of alveolar nasals in American English (e.g. in *gree*[m b]*oat*; Ellis and Hardcastle 2002) and of /n/ to following post-alveolars in Italian (Farnetani and Busà 1994). Interestingly, some of these assimilation processes show considerable inter-speaker and intra-speaker variation. For instance, Ellis and Hardcastle (2002) found that four of their eight English speakers showed categorical place assimilation of /n/ to following velars in all tokens, two speakers showed either no or categorical assimilation, and two speakers showed gradient assimilation. Together, the data show that place assimilation processes, at least those applying across morpheme boundaries, may be gradient in nature. These processes cannot simply be accounted for by the categorical spreading of a phonological feature from one segment to another.

The evidence for gradience is clearer for place assimilation than for voice assimilation. The main reason is probably that the difference between [+voice] and [−voice] obstruents is cued by many different acoustic, and hence also articulatory, characteristics, including the duration of the preceding vowel, the duration and intensity of the obstruent, and the duration of glottal vibration during the obstruent. Voice assimilation can thus not be studied on the basis of electropalatography alone, and has been mainly investigated on the basis of the acoustic signal instead. For instance, Kuzla *et al.* (2007) studied progressive voice assimilation in German clusters consisting of a voiceless obstruent and a voiced fricative (e.g. the /tv/ cluster in *ha*/t v/*älder* 'had woods' produced as [tf]). They showed that assimilation results in shorter stretches of glottal vibration during the cluster, whereas it hardly affects the duration of the fricative, which is the most important perceptual cue to the [±voice] distinction for German fricatives. Assimilation thus does not affect all perceptual cues of the [±voice] distinction equally, and the phonetic implementation of devoiced fricatives differs from the implementation of underlyingly voiceless fricatives. This is difficult to reconcile with an abstract phonological categorical account of voice assimilation, since in such an account voice assimilation results in phonologically voiceless fricatives, which cannot be distinguished from underlyingly voiceless fricatives during phonetic implementation.

Other studies have investigated regressive voice assimilation in Dutch, that is, the voiced realizations of obstruents before voiced stops (e.g. *we*[t] 'law' is realized as *we*[d] in *wetboek* 'law book'). Ernestus *et al.* (2006) and Jansen (2007) showed that glottal vibration, which is the most important cue to the [±voice] distinction in Dutch obstruent clusters (van den Berg 1988), may be completely absent, partly present, or continuously present in clusters subject to regressive voice assimilation, suggesting that regressive voice assimilation in this language is gradient. Ernestus and colleagues (2006) also investigated the effect of a word's frequency of occurrence (i.e. the word's relative number of occurrences in speech, independent of its realization) on voice assimilation (see also CHAPTER 90: FREQUENCY EFFECTS). They found that higher frequencies correlate with shorter obstruent clusters, a perceptual cue for [+voice], but also with shorter periods of glottal vibration and longer release bursts, which are perceptual cues for [−voice]. These data also suggest that voice assimilation may result in sounds that are neither fully voiced nor fully voiceless.

In conclusion, the data on assimilation suggest that we often perceive assimilation as categorical because we are used to distinguishing between only two values of the relevant phonological feature, but that the actual results from assimilation may be gradient rather than categorical. Before discussing the theoretical implications of these data, I first discuss data showing that segment deletion may also be gradient in nature.

2.3 *Segment deletion: Data*

In addition to assimilation, many studies have investigated the nature of segment deletion (CHAPTER 68: DELETION). It is generally assumed that the absence of segments may result from three different sources. First, the lexicon may represent more than one pronunciation variant for at least some words, and segment deletion may result from speakers' selection of reduced pronunciations from their lexicons. Examples of lexicalized reduced pronunciations include English *won't* for *will not* and Dutch [tyk] for [natyrlək] 'of course' (Ernestus 2000). Second, segments may be absent due to phonological deletion processes operating on the lexically represented unreduced pronunciations. These processes result in phonological surface representations without the absent segments. Both mechanisms (i.e. selection of lexically represented pronunciation variants and phonological processes) result in pronunciation variants that do not contain any acoustic cues for the missing segments, and the absence of these segments is categorical in nature. Alternatively, segments may be absent due to gradient phonetic reduction processes, which reduce the durations and articulatory strengths of segments and make different segments overlap in time (CHAPTER 79: REDUCTION). Segments that are absent due to such reduction mechanisms typically leave some traces in the acoustic signal or in the word's articulation. In conclusion, the distinction between categoricality and gradience is also relevant for the theory of segment deletion, since it indicates which type of mechanism is responsible for a given type of deletion.

This view resulted in several studies investigating the categorical *vs.* gradient nature of segment deletion processes. Browman and Goldstein (1990) hypothesized that most highly productive casual speech reduction processes result from reduction in and overlap of articulatory gestures. They showed in an X-ray study that, for instance, the /t/ in a phrase like *perfect memory* may be acoustically absent, but still articulatorily present: speakers may close their lips for the production of the /m/ before the closure of the /t/ is released, which makes the release noise of the /t/ (its most important perceptual cue) inaudible (Browman and Goldstein 1992). Several articulatory and acoustic studies of other highly productive reduction processes support this hypothesis. Thus, Manuel (1992) and Davidson (2006) demonstrated that schwa deletion in American English is gradient (CHAPTER 26: SCHWA). They reported acoustic differences between consonant clusters resulting from schwa deletion (e.g. [sp] from schwa deletion in *support*) and underlying consonant clusters (e.g. [sp] in *sport*). For instance, clusters resulting from deletion may show aspiration, whereas underlying clusters typically do not. Similarly, Russell (2008) showed that the deletion of the first vowel of a sequence of two in Plains Cree is gradient for his two native speakers (vowels may vary in their duration on the full continuum from values typical for accented full vowels to zero, which implies that they may have clear, some, or no traces at all in the acoustic signal).

In contrast, several less productive processes appear categorical in nature. Examples are the possibly morphosyntactically governed coalescence of /a+i/ or

/aː+i/ to [eː] in Plains Cree (Russell 2008) and /e/ deletion in the highly frequent French word combination *c'était* 'it was' (Torreira and Ernestus 2009). Furthermore, segments that may also be absent in more careful speech registers are more probable to be (at least partly) categorically absent. An example is word-medial schwa in French (as in *f/ə/nêtre* 'window'; see Bürki *et al.* 2010).

2.4 Gradient assimilation and segment deletion: Theoretical implications

Together, these studies suggest that many productive connected speech processes, such as assimilation and segment deletion, are gradient in nature. If the phonological component contains only categorical processes, as is assumed in traditional versions of generative grammar, these gradient processes should be classified as phonetic, which implies a move of a substantial part of the phonological component to the phonetic component. Theoretical research is needed to understand the consequences of this move. Furthermore, the experimental data suggest that post-lexical processes, in particular, show gradience. Additional detailed articulatory and acoustic studies have to investigate whether this generalization is correct. Finally, we have to investigate why some processes are categorical and others gradient and why some processes show inter-speaker and intra-speaker variation. For instance, we have to exclude the possibility that differences result from how participants deal with the experimental situation in which they are tested, including the tools that are put in their mouths for the recording of their articulation. Some participants may show normal speech behavior, while others may adapt their speech.

The evidence for the gradient nature of many connected speech processes has stimulated the development of new theoretical accounts, which do not make a fundamental distinction between the phonological and phonetic components. One of the most influential theories is Articulatory Phonology, developed by Browman and Goldstein (1986, 1992; see also CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATIONS). This theory assumes that lexical phonological representations consist of strings of articulatory gestures (articulatory scores), which are specified for time and space, and that languages differ in how these gestures may reduce in size and overlap in time. Gradient reduction in gestural size and gradient increase in gestural overlap naturally explain the gradient natures of assimilation and segment deletion processes. For instance, nasal place assimilation in English *gree*[m b]*oat* may result from the early onset of the bilabial closure, during the realization of the preceding nasal, which makes this nasal partly bilabial. In addition, Articulatory Phonology can account for categorical connected speech processes, either by incorporating the processes in the lexical representations of the words (e.g. the French word *c'était* 'it was' may have two lexical representations: one with, and one without, the gestures for the vowel /e/), or by processes that reduce gestural sizes to zero and make gestures completely overlap in time. Note that these different types of mechanisms make Articulatory Phonology a very powerful theory, which can basically explain any reduction pattern. More research is necessary to investigate how this theory can account only for those sound patterns that are actually attested. Furthermore, detailed research is necessary to explain how listeners translate the acoustic signal into gestural scores, which are the basic units of the phonological lexical representations in Articulatory Phonology.

While Browman and Goldstein (1986, 1992) proposed Articulatory Phonology as an alternative to theories making a sharp distinction between the phonological and the phonetic components, many researchers (e.g. Byrd and Choi 2010) do not consider this theory as a competitor for these theories. Rather, they incorporate the ideas of Articulatory Phonology (especially the idea of reduction in and overlap of articulatory gestures) into the phonetic component of generative grammar. Obviously, theoretical research is necessary to investigate the consequences of this incorporation.

2.5 *Incomplete neutralization: Data*

Final devoicing is another phonological process whose possible gradient nature has received a great deal of attention in the literature (see also CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). It has usually been assumed (e.g. Booij 1995) to imply a categorical change of voiced obstruents into voiceless ones, and thus a complete neutralization of the distinction between underlyingly voiced and voiceless obstruents in their phonological surface representations and articulatory and acoustic characteristics. Within traditional generative phonology, the output of final devoicing (i.e. final voiceless obstruents) forms the input to other categorical phonological processes (see below). Hence, if final devoicing turns out to lead to incomplete neutralization (i.e. to slightly voiced obstruents) and thus, according to the definitions of generative grammar, to be phonetic in nature, this has consequences for the theoretical accounts of these other phonological processes as well. That is, a gradient nature of final devoicing would have more important theoretical consequences than the gradient nature of the connected speech processes discussed above. Consequently, the possibility of incomplete neutralization has attracted attention from many researchers.

Most experimental studies have investigated the nature of final devoicing by comparing the acoustic characteristics of words differing only in the underlying voice specifications of their final obstruents. The acoustic characteristics that are typically investigated are known to correlate with perceived voicing. They include the duration of the vowel preceding the final obstruent, the duration of the final stop's closure, the duration of this stop's burst, the complete duration of the final fricative, and the duration of glottal vibration during the final obstruent. For instance, Port and O'Dell (1985) investigated ten minimal word pairs in German (e.g. *Rat* 'counsel' *vs.* *Rad* 'wheel'), read aloud by ten speakers, and showed that all acoustic measures mentioned above provided cues to the underlying voice specification of the final obstruent. In line with this, cluster analysis could correctly classify the underlying voice specifications of the obstruents on the basis of these acoustic measurements for 63 percent of the tokens. Similar studies have provided evidence for incomplete neutralization in Polish (e.g. Slowiaczek and Dinnsen 1985) and Dutch (e.g. Warner *et al.* 2004). They report acoustic differences between underlyingly voiced and voiceless obstruents in word-final position, but also that that these differences may be very small (e.g. Warner and colleagues observed a difference in vowel duration of only 2.5 msecs).

Other studies have cast doubt on these findings. For instance, Port and Crawford (1989) recorded five native speakers of German reading three minimal word pairs in four different contexts. The underlyingly voiced final obstruents differed in their realization slightly from the underlying voiceless final obstruents in all four contexts, which is in line with the incomplete neutralization hypothesis. However, speakers differed in which acoustic cues were relevant for the distinction, and, more importantly, in whether acoustic characteristics typically cueing voiced obstruents (e.g. longer preceding vowels) were combined with underlyingly voiced or voiceless obstruents. One possible explanation for (part of) these mixed results may be the nature of one of the minimal pairs (*seid*, a form of 'to be', *vs.* *seit* 'since'), since the final obstruent of the member *seid* never occurs in onset position in Modern German, and there is consequently no synchronic evidence that this obstruent is underlyingly voiced. Another study showing mixed results was conducted by Charles-Luce (1985), who investigated eight German minimal word pairs. Each of the words appeared in four different sentences, in which it was in sentence-final or medial position. Vowel duration appeared to be the only reliable cue to underlying voicing, distinguishing /t/ and /d/ in both sentence positions, but /s/ and /z/ only in sentence-final position.

Several studies have raised the question of whether the reported evidence for incomplete neutralization may result from the experimental tasks speakers had to perform. Participants typically read sentences aloud, and their pronunciation may therefore show spelling effects. Fourakis and Iverson (1984) investigated this possibility by asking their German participants to conjugate strong verbs after having heard the infinitives (e.g. they heard *reiten* and had to form *ritt* and *geritten*). In this task, participants' attention was not drawn to the spelling of the words to be pronounced. Only 10 percent of the statistical analyses showed a significant difference between the words ending in underlyingly voiced and underlyingly voiceless obstruents. Importantly, the differences were much smaller than those obtained for the same words in a word-reading task performed by the same speakers. Dinnsen and Charles-Luce (1984) addressed the role of spelling by studying five Catalan minimal word pairs whose members differed from each other in the underlying voice specification of the final obstruent, but not in spelling (e.g. /fat/ *fat* 'fate' *vs.* /fad/ *fat* 'silly'). The words were embedded in carrier sentences, and five speakers read the sentences five times. Two speakers showed incomplete neutralization, one in the expected direction (vowels were 10 percent longer before underlyingly voiced obstruents in one context condition), and one in the unexpected direction (15 percent longer closures for underlyingly voiced obstruents). Finally, Warner *et al.* (2006) addressed the role of spelling by comparing two types of Dutch word pairs consisting of morphologically related homophones that differed underlyingly only in the presence of the singleton /t/ *vs.* the geminate /tt/. Importantly, only one of these two types of word pairs reflects the underlying difference in spelling. For instance, /het+ən/ [hetən] *heten* 'are called' *vs.* /het+tən/ [hetən] *heetten* 'were called' reflects the underlying difference, whereas /het/ [het] *heet* 'am called' *vs.* /het+t/ [het] *heet* 'is called' does not. The results suggest that only those underlying differences that are reflected in orthography lead to pronunciation differences, and that these pronunciation differences are comparable in size to the pronunciation differences induced by incomplete neutralization resulting from final devoicing. Together, these results suggest that incomplete neutralization may be completely driven by orthography.

The nature of final devoicing has also been investigated in several perception studies, addressing the question of whether listeners are sensitive to the minimal acoustic differences assumed to be present between underlyingly voiced and voiceless obstruents. If they are, this supports the hypothesis of incomplete neutralization. Participants typically listened to words in isolation and indicated which word they heard by selecting the corresponding orthographic representation (e.g. German listeners heard [rat] and indicated whether they had heard *Rat* 'counsel' or *Rad* 'wheel'). All studies showed that participants tend to choose the intended orthographic representation at just above chance level (e.g. 59 percent in Port and O'Dell 1985; 62 percent in Warner *et al.* 2006). In another type of study (Ernestus and Baayen 2007), Dutch participants rated rhymes (i.e. monosyllabic words without their onsets) as 0.7 more voiced on a scale of one to five if the final obstruent was underlyingly voiced compared to voiceless. These studies thus suggest that listeners are sensitive to the minimal cues of incomplete neutralization.

It is legitimate to wonder to what extent the results from the perception experiments are simple task effects, reflecting unnatural linguistic behavior. All studies reported above asked participants to choose between orthographic forms, and hence drew participants' attention to spelling. Moreover, participants could not perform their tasks without taking the acoustic cues to incomplete neutralization into account. Ernestus and Baayen (2006) circumvented this problem by presenting Dutch participants auditorily with non-existing verb stems and asking them to produce the corresponding past tense forms. According to Dutch regular morphology, the appropriate past tense allomorph is *-te* if the final obstruent of the verbal stem is underlyingly voiceless; otherwise it is *-de*. Earlier research had shown that participants interpret the final obstruents of nonce words on the basis of the phonologically similar existing words (Ernestus and Baayen 2003). Ernestus and Baayen (2006) showed that, if the final obstruents differ slightly in their voicing, participants interpret these acoustic differences as resulting from incomplete neutralization and use these differences as a cue for their interpretations of the final obstruents as well. They do so even if their interpretations have no consequences for the spelling of these final obstruents. These findings suggest that listeners are sensitive to incomplete neutralization also if this is not necessary for the experimental task and has no consequences for spelling.

In conclusion, several experimental studies have shown that final devoicing may be incomplete, and that listeners are sensitive to the resulting minimal acoustic differences between underlyingly voiced and voiceless obstruents. Other studies, however, have cast doubt on these findings. Further research into this issue is necessary.

2.6 Incomplete neutralization: Theoretical implications

The possibility that final devoicing may be gradient is unexpected within generative grammar, since it has always been classified as a phonological process. If final devoicing is phonetic in nature (see e.g. Port and O'Dell 1985, who suggested that final devoicing and incomplete neutralization together form one phonetic implementation process), its output cannot form the input of purely phonological processes. This complicates the theoretical account of several other processes.

One example is the devoicing of voiced fricatives following syllable-final obstruents in Dutch (e.g. while *maa*/n+v/*is* 'angelfish' is pronounced as *maa*[nv]*is*, *gou*/d+v/*is* 'goldfish' is pronounced as *gou*[tf]*is* (see e.g. Booij 1995)). In the traditional generative account, this fricative devoicing results from phonological progressive voice assimilation, which is fed by phonological final devoicing (i.e. in the example *gou*[tf]*is*, final devoicing turns word-final /d/ into [t], which triggers devoicing of the following /v/). If final devoicing is phonetic, we have to assume that the devoicing of fricatives results from a phonological process that precedes and is independent of final devoicing. Another possibility is that progressive voice assimilation is phonetic as well, an assumption for which we do not have any acoustic or articulatory support.

A second phonological process that appears to follow final devoicing is resyllabification. In Dutch, word-final obstruents form syllables with following vowel-initial clitics (CHAPTER 84: CLITICS), and the word-final obstruents then occupy onset positions (e.g. *weet ie* 'knows he' is pronounced as /ue-ti/). Importantly, these word-final obstruents are typically voiceless, independently of their underlying voice specification. If final devoicing precedes resyllabification, this is as expected. Hence, if final devoicing is part of the phonetic component, we have to assume that resyllabification is phonetic as well, or we have to assume a phonological process, independent of phonetic final devoicing, which devoices resyllabified obstruents. In summary, if final devoicing is phonetic in nature, we have to assume that other phonological processes are also phonetic, or that there are several phonological processes doing partly the same work as final devoicing.

Since both options appear unattractive, Dinnsen and Charles-Luce (1984), as well as Slowiaczek and Dinnsen (1985), suggest that phonetic implementation rules (including final devoicing) may apply before phonological rules. Note that this solution implies that phonetic processes may be of very different types. Traditional phonetic implementation processes translate segments or phonological features into phonetic scores (for articulation) that correspond well with these symbols. Final devoicing, in contrast, would change [+voice] into (almost completely) [−voice].

Given the problems facing a phonetic account of final devoicing, some researchers have proposed that the process is phonological in nature, and that incomplete neutralization results from phonetic implementation processes. These accounts have to solve the question of how the phonetic component can distinguish between obstruents that should be realized as completely voiceless and those that should be slightly voiced. Van Oostendorp (2008) proposes that obstruents may be phonologically specified as voiced ([voice]), as voiceless (no specification for voice), or as devoiced (the feature [voice] is not in a pronunciation relation), and argues that this possibility directly results from assumptions about the phonological component that are necessary for the explanation of unrelated phenomena.

A completely different account of incomplete neutralization is proposed by Ernestus and Baayen (2007). Their account is based on the assumption that the mental lexicon contains representations for all words of the language, including morphologically complex words. Thus, the Dutch lexicon contains both the singular *man*[t] 'basket' and the plural *man*[d]*en* 'baskets'. This assumption is supported by the finding that all words of high frequencies of occurrence, including morphologically inflected and derived words, are recognized and produced more quickly and with fewer errors than words of low frequencies (e.g. Baayen *et al.* 1997; Alegre and Gordon 1999; CHAPTER 90: FREQUENCY EFFECTS). If the lexicon contains all words of a language, all word-final obstruents can be lexically represented as voiceless. The information that obstruents are voiced in morphologically related words is present in the lexical representations of these related words themselves. Thus, the Dutch word for 'basket' can be lexically represented as *man*/t/, since the plural *man*/d/*en* is stored as well. In this account, incomplete neutralization may be explained in two ways. First, lexical representations may be gradient and contain detailed information about the exact pronunciations of the segments (see also §2.7). Word-final obstruents may thus be represented as slightly voiced. Second, the realization of a word may be affected by the pronunciations of phonologically and morphologically related words. If a stem-final obstruent is voiced in most words, these voiced specifications may affect the pronunciation of the stem-final obstruent in word-final position, which is consequently produced as slightly voiced. This type of lexical analogy would also explain why, in the absence of an abstract mechanism of final devoicing, the final obstruents of new words are always produced as voiceless: this results from the influence of all final voiceless obstruents in the lexicon.

In conclusion, incomplete neutralization has attracted much attention in the theoretical literature, framed both within and outside generative grammar. This may be surprising since we saw above that the phenomenon is not yet well established. Note, however, that if future research will show that incomplete neutralization is just an artifact of our experimental paradigms, we still need to explain how these experimental effects can arise in speech production and comprehension. Incomplete neutralization will therefore remain an important theoretical topic

2.7 *Fine phonetic detail in speech processing*

Within generative grammar, lexical representations are categorical in nature, as they consist of strings of phonemes or phonological features, abstracting away from phonetic detail which is not necessary to distinguish between these units. In contrast, several researchers now consider the hypotheses that lexical representations are gradient in nature and reflect fine phonetic detail (see, for example, the account of incomplete neutralization of Ernestus and Baayen 2007, mentioned above) and that one and the same word may have many lexical representations reflecting slightly different pronunciations. These hypotheses are based on the findings that phonetic detail may play an important role in speech comprehension (CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY).

First, experimental data show that listeners are sensitive to phonetic detail providing information about upcoming segments. For instance, in several languages, the relative duration of a vowel is a cue to the presence of additional syllables within the same word, as vowels are typically shorter if they are followed by more syllables. Listeners use these durational cues and predict that syllables like *ham* and *dive*, produced with relatively short vowels, are part of longer words (i.e. *hamster* and *diver*; e.g. Davis *et al.* 2002; Kemps *et al.* 2005). Similarly, listeners use fine phonetic cues in syllable onsets to predict the presence of /r/ or /s/ in syllable codas (e.g. Heinrich and Hawkins 2009).

Second, several experiments have shown that listeners remember voice characteristics and that these memory traces may affect speech processing. For instance, participants are faster in determining whether two words in a sequence are identical if these two words are presented in the same voice than if they are presented in different voices (Cole *et al.* 1974). Participants tend to complete morphological stems with those suffixes that result in words they have just heard before, especially if these complex words were produced by the same voice as the stems (Schacter and Church 1992). Furthermore, participants tend to mimic previously heard pronunciations in their phonetic detail (Goldinger 1998).

These phonetic detail effects can be accounted for within generative grammar by means of the phonetic component and performance factors. The phonetic component may translate long stretches of phonological segments, rather than single segments, into acoustic signals. Likewise, listeners may analyze acoustic signals to extract not only their segments but also information on following segments (see e.g. Norris and McQueen 2008). This would explain the existence and perceptual relevance of acoustic cues distributed over longer stretches of speech. The effects of voice characteristics may result from the storage of acoustic signals in short-term memory.

In addition, these data may be accounted for by assuming that the detailed phonetic properties of a word are stored in the mental lexicon together with all other information about that word. Thus, the lexical representation *diver* may contain the information that the first vowel is relatively short. Episodic models (e.g. Goldinger 1998) assume that the mental lexicon contains such detailed representations for all tokens of all words that a speaker has ever encountered (such representations are called exemplars). These models can easily explain the processing effects of voice characteristics: if a lexicon contains a word token with the characteristics of a given speaker, the mapping of a new token of that word produced by that same speaker with the exemplars in the mental lexicon is easier than if the mental lexicon does not already contain a token by that speaker.

Episodic models are especially popular in psycholinguistics. So far, two purely episodic models have been developed and computationally implemented for speech processing: Johnson's (1997) XMOD and Goldinger's (1998) MINERVA. The XMOD model is based on the Lexical Access from Spectra (LAFS) model developed by Klatt (1979), and assumes that the incoming speech signal is transformed into a sequence of spectra. MINERVA was originally developed by Hintzman (1986) and applied to speech by Goldinger. Both XMOD and MINERVA assume that during the recognition process, exemplars respond to an acoustic input in proportion to their similarities to this input, and that their activations spread to the abstract word nodes (XMOD) or to the working memory (MINERVA), which enables recognition.

In addition to these purely episodic models, several hybrid models have been formulated, which assume both abstract lexical representations (strings of phonemes or features) and exemplars. These models can account for all experimental evidence supporting abstract lexical representations (including categorical perception, e.g. Liberman *et al.* 1957) and for the role of fine phonetic detail in speech processing. In addition, they can account for the recent finding that speaker characteristics affect speech processing only if for some reason processing is slow. McLennan and Luce (2005) as well as Mattys and Liss (2008) showed that tokens produced by the same voice are recognized more quickly than tokens produced by different voices only if the experimental task produces delayed responses (e.g. a shadowing task with a long set response time, or a lexical decision experiment that is difficult because of the many word-like pseudowords).

An important hybrid model for speech production is proposed by Pierrehumbert (2002). She assumes that speech production involves the activation of abstract representations, the application of abstract phonological rules (e.g. Prosodic Final

Lengthening), and the activation of exemplar clouds of phonological units (e.g. phonemes and phoneme sequences). Two hybrid models for word recognition are Goldinger's (2007) Complementary Learning System and the model that McLennan *et al.* (2003) developed on the basis of the Adaptive Resonance Theory (Grossberg and Stone 1986). Both models assume that the incoming signal is first analyzed into abstract phonological units, which are matched with the abstract representations in the lexicon, and only then is the signal matched with the stored exemplars. Another hybrid model for word recognition is PolySP (Polysystemic Speech Perception), developed by Hawkins and Smith (Hawkins and Smith 2001; Hawkins 2003). This model assumes that a memory trace does not only contain acoustic information, but also multi-medial context, for instance, visual information about the speaker's gestures. In addition, the model assumes that the analysis of an acoustic input into its linguistic units (phonemes, etc.) may precede (and contribute to) or coincide with or follow word recognition or not take place at all, depending on the circumstances.

In conclusion, experimental evidence suggests that gradient acoustic characteristics play a role in speech processing. More research is necessary showing which types of acoustic characteristics are relevant, how this gradient information is accessed under which conditions, and how the role of this type of information should be accounted for in speech production and comprehension models.

2.8 *Conclusion*

Gradience appears to be a much more important characteristic of speech sounds than is traditionally assumed. Place and voice assimilation, segment deletion, and final devoicing often result in sounds showing incomplete neutralization, i.e. they result in sounds that contain characteristics of more than one phoneme or that are only partly absent. Since generative grammar assumes that gradience is a characteristic of the phonetic component, these data suggest that within this theory many processes that have always been classified as phonological actually belong to the phonetic component. Particularly in the case of final devoicing, this reclassification has consequences for the classification of other speech processes as well. Alternative theories have been developed, which assume that the phonological primitives are articulatory gestures or that lexical representations reflect the gradient nature of speech sounds. These theories are supported by data showing that fine phonetic detail affects speech processing.

3 Productive sound patterns

3.1 *Introduction*

Gradience does not only play a role in the discussion of the phonological and phonetic components and of the nature of lexical representations, but also in the theoretical discussion of the nature of productive (morpho)phonological processes. Within traditional generative phonology, a productive process applies always and to all inputs that satisfy its structural description. Productive processes are thus categorical in nature. Recent research suggests, however, that some productive processes show gradience. The following two subsections discuss evidence for gradient phonological processes, their implications for generative phonology, and alternative theories that account for the gradient data.

3.2 Phonotactic constraints

The first type of phonological processes whose categorical nature has been seriously questioned is phonotactic generalizations. Within traditional generative phonology, all illegal sequences are considered equally illegal, all legal sequences as equally legal, and there is no gradience in legality.

If this assumption is correct, differences in the frequencies of occurrence of phonemes and phoneme sequences are based on coincidence. Pierrehumbert (1994) studied the frequencies of consonants and consonant clusters at the beginning of words, at the end of words (excluding the phonological appendix), and in syllable onset and coda positions within morpheme-internal consonant clusters (e.g. the frequency of [n] in words like *vanquish*, where it is in syllable coda position within a consonant cluster, and the frequency of [st] in words like *lobster*, where it is in syllable onset position in a consonant cluster) in an American English dictionary. If the consonants are randomly distributed over the positions, the frequencies of a given consonant (cluster) in the different positions should be unrelated. This appeared not to be the case. The frequency of a morpheme-internal cluster appears highly correlated to the frequency of its first part (i.e. the consonants in coda position) in word-final position and to the frequency of its second part (i.e. the consonants in onset position) in word-initial position. Phonemes and phoneme sequences structurally differ in their frequencies in a language.

Crucially, language users reflect these frequencies in their well-formedness judgments of nonce words and parts of words (CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS). Speakers typically judge high-frequency rhymes as "phonologically" better than low-frequency rhymes (Treiman *et al.* 2000), phonotactically legal nonce words as better if they contain phoneme sequences of high frequency (e.g. Vitevitch *et al.* 1997; Frisch *et al.* 2000), and nasal-obstruent clusters as better if these clusters are more frequent (Hay *et al.* 2004). Thus, *blick* is rated as a good English word, *bnick* as an impossible word and *bwick* is rated in between. Importantly, these gradient well-formedness judgments are obtained both if participants are allowed to provide gradient responses and if they have to provide categorical judgments, with the judgments being averaged over participants (Frisch *et al.* 2000). This strongly suggests that phonotactic constraints are gradient rather than categorical.

Language users' judgments of a nonce word are also affected by the phonological distance of this word from existing words (CHAPTER 87: NEIGHBORHOOD EFFECTS). Thus, participants rate a nonce word as more well-formed if it differs in fewer phonemes from an existing word (Greenberg and Jenkins 1964; Ohala and Ohala 1986). In addition, their well-formedness judgments are related to the size of a word's phonological neighborhood (Bailey and Hahn 2001; Hammond 2004), which is typically defined as the number of existing words that can be changed into that word by the substitution, addition, or deletion of a single phoneme. Importantly, the effect of the word's phonological neighborhood is independent of the effects of the frequencies of the word's constituents (i.e. the effect is also present if words with small and larger neighborhoods are matched in the frequencies of their constituents). This shows again that well-formedness judgments are not categorical (i.e. it is not the case that a word is either completely well formed or completely ill formed). Rather, these judgments are gradient between completely well formed and completely ill formed.

Importantly, the measures affecting well-formedness judgments also play a role in other (psycho-)linguistic tasks. The frequencies of the phonemes and phoneme sequences in a word have been shown to affect speech production, recognition, and learning. For instance, participants are better at repeating nonce words made up of high-frequency rather than low-frequency phoneme sequences (Vitevitch *et al.* 1997) and at transcribing such words orthographically (Hay *et al.* 2004). Participants tend to interpret ambiguous fricatives as the most probable ones given the preceding and following segments (Pitt and McQueen 1998). Nine-month-old infants prefer to listen to words consisting of high-frequency rather than low-frequency phoneme sequences (Jusczyk *et al.* 1994). Furthermore, when both eight-month-old infants and adults are presented with continuous speech froma non-existing (artificial) language, they extract the words of this language on the assumption that frequent phoneme sequences form (parts of) words, while the less frequent ones span word boundaries (Saffran *et al.* 1996a; Saffran *et al.* 1996b). Similarly, speech production and comprehension are affected by a word's phonological neighborhood. Thus, participants recognize words with large neighborhoods more slowly in auditory lexical decision (e.g. Luce and Pisoni 1998) and produce them with more expanded vowel spaces (Munson and Solomon 2004), while pre-school-aged children produce such words more quickly and with fewer errors in picture-naming tasks (Arnold *et al.* 2005).

Several generative linguists have assumed that the gradience of well-formedness judgments may be merely a task effect, resulting from performance factors (for a discussion, see Schütze 2005). This account is in line with the finding that the variables affecting well-formedness ratings also play roles in speech production, perception, and learning, which are certainly modulated by performance factors.

In addition, there is a continuum of accounts which differ in their assumptions about the contributions of the phonological component and the mental lexicon. The models at one end of the continuum assume that the gradience of well-formedness judgments results from the gradient nature of the phonological component itself. This component would be gradient due to the probabilistic nature of its constraints or rules. For instance, Hammond (2004) frames his account of gradient well-formedness judgments within Probabilistic Optimality Theory, which is based on Stochastic Optimality Theory, developed by Boersma (1998). The idea is that the ranking of constraints is variable, and that a given (markedness or faithfulness) constraint outranks some other constraint with a certain probability. If this probability is smaller than 1, the phonological component shows variation, sometimes favoring one form and sometimes another, which results in gradient well-formedness rankings. The probability of a given ranking (and consequently the judgment of a given form) may be co-determined by the frequencies of phoneme sequences and by the exact contents of the mental lexicon. Models at the other end of the continuum assume that well-formedness judgments for a given word result only from the comparison of that word with all words in the mental lexicon and their constituents. The visual or auditory presentation of a word leads to the activation of all (phonologically) similar words in the lexicon and their constituents, and a higher total lexical activation leads to a higher well-formedness rating. In these analogical models, there is thus no role for an abstract phonological component with hardwired phonological constraints or rules (e.g. Bailey and Hahn 2001). Models positioned between the two ends typically assume that the effects of constituent frequencies result from phonotactic knowledge, and the effects of phonological neighborhood from lexical knowledge. Phonotactic knowledge is permanently stored in the phonological component, while lexical knowledge is deduced from the mental lexicon if necessary (e.g. Bailey and Hahn 2001; Albright 2009).

In summary, the evidence for gradience in well-formedness judgments is undisputed. Detailed research is necessary in different domains of phonology to establish the best theoretical account.

3.3 *Allomorphy*

A second type of productive phonological process that appears gradient is those involved in morphological processing. These morphophonological processes select affixes on the basis of the phonological properties of the words' stems (CHAPTER 99: PHONOLOGICALLY CONDITIONED ALLOMORPH SELECTION). For instance, Dutch regular past tense forms consist of a verbal stem and the suffix *-te* or *-de*. According to the traditional literature (which follows Dutch orthography), the correct allomorph is *-te* if the verbal stem ends in an underlyingly voiceless obstruent (e.g. *sta*/p+t/*e* 'stepped'), otherwise it is *-de* (e.g. *kra*/b+d/*e* 'scratched'). It has been shown recently that at least some of these apparently perfectly categorical generalizations do not do justice to the full data.

One example is the above-mentioned regular past tense formation in Dutch. Ernestus and Baayen (2004) show that the description of the selection of the past tense allomorph given in the literature is too simplistic. Speakers tend to choose the non-standard allomorph for verbal stems that are special, in that the underlying voice specification of their final obstruent is unexpected given the other stems ending in the same type of final rhyme in the lexicon. For instance, speakers often choose the non-standard allomorph for *kra*/b/ (creating *kra*/b+t/*e*), which is one of the few Dutch verb stems ending in a short vowel and a voiced (instead of voiceless) bilabial stop. The pattern "short vowel–underlyingly voiceless bilabial stop" is much more common (e.g. *sto*/p/ 'stop', *kla*/p/ 'bang', *me*/p/ 'slap', *ni*/p/ 'sip') than the pattern "short vowel–underlyingly voiced bilabial stop," and speakers tend to add the allomorph that is correct for the majority of verbs ending in a short vowel and a bilabial stop to the minority of verbs for which it is incorrect (i.e. verbs ending in a short vowel and an underlyingly voiced bilabial stop). These findings can easily be incorporated in all types of theoretical accounts, since the only adaptation necessary is that the broad generalizations are replaced or supplemented by generalizations that are more specific for the precise phonological properties of the words. Apparently, Dutch requires a generalization stating that stems ending in short vowels and bilabial stops tend to select *-te*.

Importantly, however, the facts are more complex. First, Ernestus and Baayen (2004) observe that, if participants select the standard allomorph, they do so more quickly for verbs following the majority patterns than for exceptional verbs (i.e. they produce forms of the type *stapte* more quickly than forms of the type *krabde*). Second, Ernestus and Baayen (2003, 2004; see also Ernestus 2006) find that speakers show stochastic behavior; they often do not agree with each other, and the same speaker may choose *-te* for some verbs and *-de* for other verbs of the same type. Similar results have been found, among others, for past tense formation in English (Albright and Hayes 2003), the choice of the English indefinite article (*a vs. an*; Skousen 1989), and vowel harmony in Hungarian (Hayes and Londe 2006). Apparently, the morphophonological processes that have to replace or supplement the traditional broad generalizations are not simple categorical rules that apply whenever their structural description is met. The processes are gradient in nature.

Speakers' probabilistic behavior has been accounted for in the two types of approaches (forming a continuum) that also explain the gradience of well-formedness ratings (see above). The first approach holds that constraints or rules are probabilistic in nature. Thus, in Stochastic Optimality Theory (Boersma 1998), constraint rankings are stochastic, and in the rule-based account proposed by Albright and Hayes (2003) rules differ in their confidence intervals. Both accounts assume that the probability of a constraint ranking or rule (and thus of a given form) is determined by the exact contents of the mental lexicon. While this approach can account well for the observed probabilistic effects, additional assumptions are necessary to explain why speakers are slower in selecting the standard allomorph if it receives less lexical support than the other allomorph (for a discussion, see Ernestus 2006). The second approach to speakers' stochastic behavior assumes that, when speakers select an allomorph for a word, they check all words in their lexicons online. The probability that they select a given allomorph is proportional to its support from the words in the lexicon, with words that are more similar to the target word being more influential. If the target word itself is in the lexicon as well and supports a different allomorph from the one receiving the greatest lexical support from the other words, this may result in severe competition between the two allomorphs, which may lead to the selection of the non-standard allomorph and longer response latencies (Ernestus and Baayen 2004).

In conclusion, phonologically driven allomorphy also strongly suggests that gradience is an important characteristic of phonology. The generalizations formulated in the generative literature appear too coarse-grained, given that speakers show probabilistic behavior. Several models can account for the obtained observations so far. More data are necessary to tease the different accounts apart.

4 Conclusion

In the early days of generative grammar, phonology was assumed to be completely categorical in nature. The present chapter has provided a summary of different types of corpus-based and experimental studies which strongly suggests that many processes traditionally classified as phonological are in fact gradient in nature. Sounds may contain characteristics of different categories, and speakers may show probabilistic behavior. These data have given rise to modifications of traditional generative phonology and to the development of new theories, including theories assuming different types of phonological primitives and phonological representations, and theories challenging the role of abstract generalizations. Further research is necessary to obtain a more detailed view of the role of gradience in phonology and to tease different theoretical accounts apart. Until then, we have to conclude that gradience is an important challenge for phonology.

REFERENCES

Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26. 9–41.

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119–161.

Allegre, Maria & Peter Gordon. 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40. 41–61.

Arnold, Hayley S., Edward G. Conture & Ralph N. Ohde. 2005. Phonological neighborhood density in the picture naming of young children who stutter: Preliminary study. *Journal of Fluency Disorders* 30. 125–148.

Baayen, R. Harald, Ton Dijkstra & Robert Schreuder. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language* 36. 94–117.

Bailey, Todd M. & Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44. 568–591.

Barry, Martin C. 1992. Palatalisation, assimilation and gestural weakening in connected speech. *Speech Communication* 11. 393–400.

Berg, Rob J. H. van den. 1988. *The perception of voicing in Dutch two-obstruent sequences*. Enschede: Sneldruk Enschede.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Booij, Geert. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Browman, Catherine P. & Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In Kingston & Beckman (1990), 341–376.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180.

Bürki, Audrey, Mirjam Ernestus & Ulrich Frauenfelder. 2010. Is there only one fenêtre in the production lexicon? On-line evidence about the nature of phonological representations. *Journal of Memory and Language* 62. 421–437.

Byrd, Dani & Susie Choi. 2010. At the juncture of prosody, phonology, and phonetics: The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.) *Laboratory phonology* 10, 31–59. Berlin & New York: Mouton de Gruyter.

Charles-Luce, Jan. 1985. Word-final devoicing in German: Effects of phonetic and sentential contexts. *Journal of Phonetics* 13. 309–324.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Cohn, Abigail C. 1993. Nasalisation in English: Phonology or phonetics. *Phonology* 10. 43–81.

Cole, Ronald A., Max Coltheart & Fran Allard. 1974. Memory of a speaker's voice: Reaction time to same- or different-voiced letters. *Quarterly Journal of Experimental Psychology* 26. 1–7.

Davidson, Lisa. 2006. Schwa elision in fast speech: Segmental deletion or gestural overlap. *Phonetica* 63. 79–112.

Davis, Matthew H., William D. Marslen-Wilson & M. Gareth Gaskell. 2002. Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28. 218–244.

Dinnsen, Daniel A. & Jan Charles-Luce. 1984. Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12. 49–60.

Ellis, Lucy & W. J. Hardcastle. 2002. Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics* 30. 373–396.

Ernestus, Mirjam. 2000. *Voice assimilation and segment reduction in casual Dutch: A corpus-based study of the phonology–phonetics interface*. Utrecht: LOT.

Ernestus, Mirjam. 2006. Statistically gradient generalizations in phonology. *The Linguistic Review* 23. 217–234.

Ernestus, Mirjam & R. Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79. 5–38.

Ernestus, Mirjam & R. Harald Baayen. 2004. Analogical effects in regular past tense production in Dutch. *Linguistics* 42. 873–903.

Ernestus, Mirjam & R. Harald Baayen. 2006. The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In Louis M. Goldstein, Douglas Whalen & Catherine T. Best (eds.) *Laboratory phonology* 8, 27-49. Berlin & New York: Mouton de Gruyter.

Ernestus, Mirjam & R. Harald Baayen. 2007. Intraparadigmatic effects on the perception of voice. In van de Weijer & van der Torre (2007), 153–172.

Ernestus, Mirjam, Mybeth Lahey, Femke Verhees & R. Harald Baayen. 2006. Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America* 120. 1040–1051.

Farnetani, Edda & M. Grazia Busà. 1994. Italian clusters in continuous speech. *Proceedings of the 3rd International Conference on Spoken Language Processing*, 359–361. Yokohama: Acoustical Society of Japan.

Fourakis, Marios & Gregory K. Iverson. 1984. On the "incomplete neutralization" of German final obstruents. *Phonetica* 41. 140–149.

Frisch, Stefan A., Nathan R. Large & David B. Pisoni. 2000. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42. 481–496.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105. 251–279.

Goldinger, Stephen D. 2007. A complementary-systems approach to abstract and episodic speech perception. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 49–54. Saarbrücken: Saarland University.

Greenberg, Joseph H. & James J. Jenkins. 1964. Studies in the psychological correlates of the sound system of American English. *Word* 20. 157–177.

Grossberg, Stephen & Gregory Stone. 1986. Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review* 93. 46–74.

Hammond, Michael. 2004. Gradience, phonotactics, and the lexicon in English phonology. *International Journal of English Studies* 4. 1–24.

Hardcastle, W. J. 1972. The use of electropalatography in phonetic research. *Phonetica* 25. 197–215.

Hawkins, Sarah. 2003. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31. 373–405.

Hawkins, Sarah & Rachel Smith. 2001. Polysp: A polysystemic, phonetically rich approach to speech understanding. *Rivista di Linguistica* 13. 99–188.

Hay, Jennifer, Janet B. Pierrehumbert & Mary E. Beckman. 2004. Speech perception, well-formedness and the statistics of the lexicon. In John Local, Richard Ogden & Rosalind Temple (eds.) *Phonetic interpretation: Papers in laboratory phonology VI*. Cambridge: Cambridge University Press. 58–74.

Hayes, Bruce & Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104.

Heinrich, Antje & Sarah Hawkins. 2009. The effect of r-resonance information on intelligibility. In Uther *et al.* (2009), 804–807.
Hintzman, Douglas L. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93. 411–428.
Jansen, Wouter. 2007. Dutch regressive voicing assimilation as a "low level phonetic process": Acoustic evidence. In van de Weijer & van der Torre (2007), 125–151.
Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson & John W. Mullennix (eds.) *Talker variability in speech processing*, 145–165. San Diego: Academic Press.
Jusczyk, Peter W., Paul A. Luce & Jan Charles-Luce. 1994. Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33. 630–645.
Keating, Patricia. 1985. Universal phonetics and the organization of grammars. In Victoria A. Fromkin (ed.) *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 115–132. Orlando: Academic Press.
Keating, Patricia. 1988. Underspecification in phonetics. *Phonology* 5. 275–292.
Keating, Patricia. 1990a. Phonetic representations in a generative grammar. *Journal of Phonetics* 18. 321–334.
Keating, Patricia. 1990b. The window model of coarticulation: Articulatory evidence. In Kingston & Beckman (1990), 451–470.
Kemps, Rachèl, Lee H. Wurm, Mirjam Ernestus, Robert Schreuder & R. Harald Baayen. 2005. Prosodic cues for morphological complexity: Comparatives and agent nouns in Dutch and English. *Language and Cognitive Processes* 20. 43–73.
Kenstowicz, Michael & Charles W. Kisseberth. 1979. *Generative phonology: Description and theory*. New York: Academic Press.
Kingston, John & Mary E. Beckman (eds.) 1990. *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press.
Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70. 419–454.
Klatt, Dennis H. 1979. Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics* 7. 279–312.
Kochetov, Alexei & Marianne Pouplier. 2008. Phonetic variability and grammatical knowledge: An articulatory study of Korean place assimilation. *Phonology* 25. 399–431.
Kuzla, Claudia, Taehong Cho & Mirjam Ernestus. 2007. Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics* 35. 301–320.
Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Liberman, Alvin, Katherine S. Harris, Howard S. Hoffman & Belver C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54. 358–368.
Luce, Paul A. & David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19. 1–36.
Manuel, Sharon Y. 1992. Recovery of "deleted" schwa. In Olle Engstrand & Catharina Kylander (eds.) *Perilus XIV: Papers from the Symposium on Current Phonetic Research Paradigms for Speech Motor Control*, 115–118. Stockholm: University of Stockholm.
Mattys, Sven L. & Julie M. Liss. 2008. On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality. *Perception and Psychophysics* 70. 1235–1242.
McLennan, Conor T. & Paul A. Luce. 2005. Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition* 31. 306–321.
McLennan, Conor T., Paul A. Luce & Jan Charles-Luce. 2003. Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory and Cognition* 29. 539–553.
Munson, Benjamin & Nancy P. Solomon. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47. 1048–1058.

Nolan, Francis. 1992. The descriptive role of segments: Evidence from assimilation. In Gerard J. Docherty & D. Robert Ladd (eds.) *Papers in laboratory phonology II: Gesture, segment, prosody*, 261–280. Cambridge: Cambridge University Press.

Norris, Dennis. 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52. 189–234.

Norris, Dennis & James M. McQueen. 2008. Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review* 115. 357–395.

Ohala, John J. & Manjari Ohala. 1986. Testing hypotheses regarding the psychological manifestation of morpheme structure constraints. In John J. Ohala & Jeri Jaeger (eds.) *Experimental phonology*, 239–252. Orlando: Academic Press.

Oostendorp, Marc van. 2008. Incomplete devoicing in formal phonology. *Lingua* 118. 1362–1374.

Perkell, Joseph S., Marc H. Cohen, Mario A. Svirsky, Melanie L. Matthies, Iñaki Garabieta & Michael T. T. Jackson. 1992. Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America* 92. 3078–3096.

Pierrehumbert, Janet B. 1990. Phonological and phonetic representation. *Journal of Phonetics* 18. 375–394.

Pierrehumbert, Janet B. 1994. Syllable structure and word structure: A study of triconsonantal clusters in English. In Patricia Keating (ed.) *Phonological structure and phonetic form: Papers in laboratory phonology III*, 168–188. Cambridge: Cambridge University Press.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology 7*, 101–139. Berlin & New York: Mouton de Gruyter.

Pitt, Mark A. & James M. McQueen. 1998. Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language* 39. 347–370.

Port, Robert F. & Penny Crawford. 1989. Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17. 257–282.

Port, Robert F. & Michael O'Dell. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13. 455–471.

Russell, Kevin. 2008. Sandhi in Plains Cree. *Journal of Phonetics* 36. 450–464.

Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996a. Statistical learning by 8-month-old infants. *Science* 274. 1926–1928.

Saffran, Jenny R., Elissa L. Newport & Richard N. Aslin. 1996b. Word segmentation: The role of distributional cues. *Journal of Memory and Language* 35. 606–621.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Lausanne & Paris: Payot.

Schacter, Daniel L. & Barbara A. Church. 1992. Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18. 915–930.

Schütze, Carson T. 2005. Thinking about what we are asking speakers to do. In Stephan Kepser & Marga Reis (eds.) *Linguistic evidence: Empirical, theoretical, and computational perspectives*, 457–484. Berlin & New York: Mouton de Gruyter.

Skousen, Royal. 1989. *Analogical modeling of language*. Dordrecht: Kluwer.

Slowiaczek, Louisa M. & Daniel A. Dinnsen. 1985. On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13. 325–341.

Torreira, Francisco & Mirjam Ernestus. 2009. Vowel elision in connected French: The case of vowel /e/ in the word *c'était*. In Uther *et al.* (2009), 448–451.

Treiman, Rebecca, Brett Kessler, Stephanie Knewasser, Ruth Tincoff & Margo Bowman. 2000. English speakers' sensitivity to phonotactic patterns. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, 269–282. Cambridge: Cambridge University Press.

Uther, Maria, Roger Moore & Stephen Cox (eds.) 2009. *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*. Brighton: Causal Productions.

Vitevitch, Michael S., Paul A. Luce, Jan Charles-Luce & David Kemmerer. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40. 47–62.
Warner, Natasha, Allard Jongman, Joan A. Sereno & Rachèl Kemps. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32. 251–276.
Warner, Natasha, Erin Good, Allard Jongman & Joan A. Sereno. 2006. Orthographic *vs.* morphological incomplete neutralization effects. *Journal of Phonetics* 34. 285–293.
Weijer, Jeroen van de & Eric Jan van der Torre (eds.) 2007. *Voicing in Dutch: (De)voicing – Phonology, phonetics, and psycholinguistics.* Amsterdam & Philadelphia: John Benjamins.
Zsiga, Elizabeth C. 1995. An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In Bruce Connell & Amalia Arvanti (eds.) *Papers in laboratory phonology IV: Phonology and phonetic evidence*, 282–302. Cambridge: Cambridge University Press.
Zsiga, Elizabeth C. 1997. Features, gestures, and Igbo vowels: An approach to the phonology–phonetics interface. *Language* 73. 227–274.

83 Paradigms

Adam Albright

1 Introduction

Morphological paradigms are a mainstay of traditional descriptions of inflectional systems and of diachronic change. Only in recent years, however, have paradigms played a formal role in the grammatical analysis of phonological systems, in the form of correspondence constraints and contrast constraints on paradigmatically related forms. In this chapter I review some of the evidence that has been taken to indicate that paradigm structure plays an active role in synchronic phonology, and discuss some of the grammatical mechanisms that have been proposed to capture such effects, focusing especially on work within Optimality Theory (OT: Prince and Smolensky 2004).

It must be acknowledged at the outset that one cannot meaningfully discuss phonological paradigm effects without a precise definition of "paradigm." I begin by adopting a very general and widely assumed definition: a paradigm is the exhaustive set of inflected forms that share a single root or stem – e.g. inflected case and number forms of a noun, or person, number or tense/aspect/mood forms of a verb. In some cases, phonology treats all inflected forms of a root alike, and this broad definition suffices. In many cases, however, it is necessary to restrict the domain of discussion to a specific subset of inflected forms that constitute a local subparadigm – e.g. the set of verb forms that share present tense subjunctive inflection (the "present subjunctive paradigm"), or the set of noun forms that share dual inflection (the "dual paradigm"). For the purpose of illustrating paradigm effects and their analysis, I will simply stipulate the scope of the effect as necessary; in §4, I return to the issue of the formal definition of paradigms.

An important related issue is whether phonological paradigm effects are conditioned by the same type of paradigm structure that is posited by paradigm-based theories of morphology, such as Word-and-Paradigm Morphology (Hockett 1954; Matthews 1965; Zwicky 1985; Anderson 1986) or Paradigm Function Morphology (Stump 2001). On the face of it, paradigm-based theories of morphology seem especially well suited for capturing phonological paradigm effects, since they provide a representational unit (the "paradigm") that can be used to condition morphological and phonological grammar distributions. The use of phonological paradigm constraints does not presuppose the existence of paradigms as morphological representations, however. In fact, the phonological constraints described in §3 make very limited use of morphological structure: if two forms share a root and all derivational affixes, then they are treated by the phonology as members of the same inflectional paradigm. Although arguments against morphological and phonological uses of paradigms are often presented side by side (e.g. Bobaljik 2008), it appears that the question of whether morphology is paradigm-based may be orthogonal to the question of whether phonology imposes constraints on relations between inflectionally related forms. The current chapter focuses on the role that paradigmatic relations play in conditioning phonological distributions, and leaves aside arguments for and against paradigms as morphological representations. A long-term goal, beyond the scope of this chapter, would be to establish whether morphological and phonological distributions rely on a common set of representational units, or whether phonological constraints make only indirect reference to morphological representations.

In this chapter, I consider a variety of effects that demonstrate a role for paradigms in phonological grammar. Before turning to the analysis of synchronic paradigm effects, however, it is useful to review briefly the type of data that have traditionally been taken as evidence that phonological processes are sensitive to paradigm structure, drawn from the domain of language change.

1.1 *Paradigms as a factor in conditioning diachronic change*

It is often convenient to present complex inflectional systems in tabular form, with each cell in the paradigm representing a particular combination of morphosyntactic features. In some theories of morphology, paradigms are taken to be not merely a matter of descriptive convenience, but also a grammatically relevant representation of the distribution of inflectional markers (Hockett 1954; Matthews 1965; Anderson 1986; Wurzel 1989; Stump 2001; Blevins 2003; Ackerman *et al.* 2009). According to proponents of paradigm-based models, representations in terms of paradigms permit concise or insightful statements about morphological distribution that would be difficult to capture if each morpheme were represented individually. For instance, patterns of syncretism and alternation are often shared across multiple inflection classes, suggesting that a paradigmatic template is in force (Williams 1994; Baerman *et al.* 2005; Maiden 2005; though see Bobaljik 2002 and Harley 2008 for alternative approaches). An example is seen in Modern German, which has stem vowel alternations in the 2nd and 3rd singular present tense forms of many verbs. These alternations can be traced back to historically unrelated raising processes, but the conditioning context (originally, high vowels in the suffix) is no longer present in the modern language, nor can the various changes found in the 2nd and 3rd singular be straightforwardly unified as the same featural change. Thus it appears that the broadest generalization one can make is that the 2nd and 3rd singular differ from the rest of the paradigm in having a raised and/or fronted vowel. This relation cannot be captured by a single (morpho)phonological rule, but can be characterized by a template in which the 2nd and 3rd singular forms differ from the remaining forms.

(1) *Vowel alternations in German present tense indicative verb forms*

Verb	'travel'	'run'	'give'	'see'
INF	faːʁən	laʊfən	geːbən	zeːən
1SG	faːʁ(ə)	laʊf(ə)	geːb(ə)	zeː(ə)
2SG	f<u>ɛː</u>ʁst	l<u>ɔɪ</u>fst	g<u>ɪ</u>b̥st	z<u>iː</u>st
3SG	f<u>ɛː</u>ʁt	l<u>ɔɪ</u>ft	g<u>ɪ</u>b̥t	z<u>iː</u>t
1PL	faːʁən	laʊfən	geːbən	zeːən
2PL	faːʁt	laʊft	geːb̥t	zeːt
3PL	faːʁən	laʊfən	geːbən	zeːən

A compelling source of evidence for a paradigmatic effect comes from language change: in fact, verbs like [geːbən] 'give' and [zeːən] 'see' originally showed a different distribution, in which the 1st singular had the same vowel as the 2nd and 3rd singular: [gibə], [gib̥st], [gib̥t]. The change of 1st singular [gibə]→ [geːbə] to match the vowel of the plural and infinitive appears to have been motivated by the influence of verbs like [faːʁən] 'travel', in which raising has always been limited to the 2nd and 3rd singular (Paul *et al.* 1989: 243).[1] Put differently, a prevalent pattern of alternation within the paradigm (2nd and 3rd singular *vs.* others) was generalized to verbs with a similar but less robustly attested pattern (singular *vs.* plural). Such changes are often taken as evidence that speakers evaluate the relations between forms within paradigms, and that language change may enforce or regularize such relations.

1.2 *Paradigm uniformity in language change*

One common way in which paradigmatic relations are strengthened is by loss (or "leveling") of alternations among inflectionally related forms. A widely discussed example comes from immediately pre-classical Latin, in which [s] ~ [r] alternations created by rhotacism of intervocalic stridents were leveled to invariant [r] (Hock 1991: 179–190; Barr 1994; Kenstowicz 1996; Hale *et al.* 1997; Kiparsky 1998; Baldi 1999: 323; Albright 2005).

(2) *Leveling of rhotacism alternations in Latin* honor

	Stage 1	*Stage 2*
NOM SG	honoː<u>s</u>	hono<u>r</u>
GEN SG	honoːris	honoːris
DAT SG	honoːriː	honoːriː
ACC SG	honoːrem	honoːrem
ABL SG	honoːre	honoːre

[1] An alternative view, put forward by Joesten (1931) and discussed by Dammers *et al.* (1988: 449) and by Hartweg and Wegera (1989: 129), holds that the distribution in (1) is etymologically expected for all of these verbs, including *geben*, and that the use of 1st singular *gibe* in literary Middle High German represents a partial leveling of the vowel from the 2nd and 3rd singular to the 1st singular in certain dialects. Either way, the difference between dialects with 1st singular *gibe* and those with *gebe* requires some form of analogical change within the paradigm to match a pattern found in other verbs.

Crucially, the change of [s] to [r] in [hono:s] → [honor] took place only within the inflectional paradigm, while derivationally related forms such as [hones-tus] 'honorable', [hones-te:] 'honorably', and [hones-ta:s] 'honorableness' remained unchanged. Furthermore, [s]-final nouns that had no inflected forms with [r], such as the indeclinable noun [nefa:s] 'sacrilege' (related to the derived adjective [nefa:r-ius] 'wicked'), also remained unchanged, as did [s]-final words in other parts of speech, such as the adverb [nimis] 'excessively'. These facts show that the change of [s] to [r] was not part of a broader sound change extending rhotacism to the word-final position. Furthermore, if we assume that forms like [hones-tus] remained synchronically linked to the related noun [honor], the change cannot be viewed as a re-analysis to /honor/ by learners missing the presence of [s] ~ [r] alternations (Vincent 1974). Such examples are often taken as evidence that the domain of analogical leveling is the inflectional paradigm.

Another common feature of paradigm leveling is that it may affect only a subset of the inflectionally related forms. For example, several classes of Middle High German (MHG) verbs showed vowel alternations between the singular and plural, with the infinitive always matching the vowel of the plural, and the past participle frequently showing yet a different vowel. In Yiddish, these alternations have been eliminated completely in some verbs (3a), while they have been leveled only among present tense forms for others (3b). We may plausibly suppose that the infinitive and finite present tense forms of /visən/ 'know' remain synchronically related, but the identity of the 1st/3rd plural and infinitive has been abandoned in favor of an invariant verb stem within the subparadigm of present tense forms (Albright 2010).

(3) *Full and partial leveling in Yiddish*

a.	'need'	MHG	Yiddish	b.	'know'	MHG	Yiddish
	INF	*dürf-en*	d<u>a</u>rf-n̩		INF	*wiȝȝ-en*	vis-n̩
	1SG	*darf*	darf		1SG	*weiȝ*	veɪs
	2SG	*darf-t*	darf-st		2SG	*weis-t*	veɪs-t
	3SG	*darf*	darf-(t)		3SG	*weiȝ*	veɪs
	1PL	*dürf-en*	d<u>a</u>rf-n̩		1PL	*wiȝȝ-en*	v<u>eɪ</u>s-n̩
	2PL	*dürf-t*	d<u>a</u>rf-t		2PL	*wis-t*	v<u>eɪ</u>s-t
	3PL	*dürf-en*	d<u>a</u>rf-n̩		3PL	*wiȝȝ-en*	v<u>eɪ</u>s-n̩
	PAST PRT	*ge-dorft*	ge-d<u>a</u>rft		PAST PRT	*ge-wus-t*	ge-vus-t

Language change provides numerous examples of paradigmatic conditioning. However, we frequently cannot be certain whether the changes reflect a synchronic preference for non-alternating paradigms, or whether they represent lexical re-analyses on the part of learners. Such changes are difficult to interpret, because we cannot be sure that Latin speakers synchronically derived forms like [hono:s] and [hones-tus] (without rhotacism) from a single stem /hono:s/. If [hono:s] and [honestus] were derived from different stems, we could interpret the change as a re-analysis of the nominal stem from /hono:s/ to /honor/ (perhaps motivated by the preponderance of rhotacized [hono:r-] forms), while the adjectival stem /hones/ remained intact. Similarly, although there is no reason to think that MHG speakers treated the infinitive and plural stems *wiȝȝ-* and *wiȝȝ-* as distinct, if for some reason they did encode them as separate stems, then we could simply say that Yiddish lost the plural stem while retaining the infinitive stem. In order to show that a phonological process misapplies within inflectional paradigms, we must demonstrate that the forms in question are synchronically derived from the same stem, and that the process in question continues to apply except when trumped by identity among paradigmatically related forms. In the following section, we review several cases with exactly this flavor: a phonological process continues to apply straightforwardly and productively in derived forms, but is overridden just in case greater identity between inflected forms would result.

2 Synchronic paradigm effects

The literature on synchronic paradigm effects has identified two opposing ways in which phonological relations among inflected words may be regulated: on the one hand, there is a tendency to demand identity among inflected forms, so that elements of shared meaning (the stem, shared inflectional markers) have a consistent phonological form throughout the paradigm (*uniformity*). At the same time, there is a tendency to avoid total identity between forms that are morphologically distinct (*anti-homophony*) (see also CHAPTER 103: PHONOLOGICAL SENSITIVITY TO MORPHOLOGICAL STRUCTURE). In both cases, it is claimed that the requirement of identity or distinctness may be enforced within the domain of the paradigm, while violations of identity between derivationally related forms, or violations of anti-homophony among unrelated forms, are not enforced. I consider these two tendencies in turn. Paradigmatic conditions in phonology have been explored extensively in the past decade, and it would not be possible in a chapter of this length to do justice to the range and intricacy of empirical cases that have been brought to bear on the issue. For additional examples, the reader is referred to the collections of papers in Hermans and van Oostendorp (1999), Downing *et al.* (2005b), and Bachrach and Nevins (2008).

Frequently, affixation creates or destroys the context for a phonological process to apply, leading to the possibility of alternations. This section discusses cases in which regularly expected alternations are avoided, and paradigmatic identity is seen instead. As is standard in the literature on identity effects in other domains, such as reduplication (McCarthy and Prince 1995), we may distinguish between cases where identity is achieved by applying a process outside its regular context (*overapplication*), by failing to apply a process within its regular context (*underapplication*), or by applying a process differently from how it would normally apply to a given phonological string (*misapplication*). It is also important to bear in mind that these phenomena are not substantively different from those that arise in cases of cyclicity in derived forms; therefore, much of the discussion of parallel facts in CHAPTER 85: CYCLICITY is relevant here as well.

2.1 *Overapplication*

Frequently, phonological processes apply outside their regular context, if doing so can achieve greater identity between related forms. For example, Hayes (2000) discusses a process in American English (and other varieties) in which coda /l/ becomes dark and creates a diphthongized allophone of the preceding vowel (CHAPTER 31: LATERAL CONSONANTS): /fiːl/ → [fiəɫ] *feel*, /faɪl/ → [faɪəɫ] *file*,

/bɔɪl/ → [bɔɪəɫ] *boil*. This process does not normally occur intervocalically: [siːlɪŋ]/*[siəɫɪŋ] *ceiling*, [paɪlət]/*[paɪəɫət] *pilot*, [tɔɪlət]/*[tɔɪəɫət] *toilet*. However, at morpheme boundaries, coda velarization may apply. Hayes presents survey data documenting a variable or gradient pattern in which /l/ may be realized as [ɫ] before morpheme boundaries: *mai[ɫ]er*, *(touchy-)fee[ɫ]y*. The pattern that Hayes describes for his own speech involves overapplication of diphthongization and coda velarization before both inflectional and derivational affixes: [hiəɫɪŋ] *healing*, [meəɫə] *mailer*. By contrast, in at least the present author's idiolect, coda velarization always overapplies, but diphthongization overapplies (optionally) only in inflected forms. The difference is most clearly heard (and intuited) after underlying diphthongs, where "diphthongization" yields so-called sesquisyllabic outcomes such as [bɔɪəɫ] *boil*. Crucially, the overapplication of diphthongization seems to occur only in inflected forms: *boiling* [bɔɪəɫɪŋ] *patiently vs. boiler* [bɔɪɫɚ], *[bɔɪəɫɚ].

(4) *Overapplication of pre-lateral diphthongization in one idiolect of American English*

base		*inflected*	*derived*	
[bɔɪəɫ]	*boil*	[bɔɪɫɪŋ], [bɔɪəɫɪŋ]	[bɔɪɫɚ], *[bɔɪəɫɚ]	*boiler* ('water-heater')
[spɔɪəl]	*spoil*	[spɔɪɫɪŋ], [spɔɪəɫɪŋ]	[spɔɪɫɚ], *[spɔɪəɫɚ]	*spoiler* (of plot)
			[spɔɪɫədʒ], *[spɔɪəɫədʒ]	*spoilage*
[ɔɪəɫ]	*oil*	[ɔɪɫɪŋ], [ɔɪəɫɪŋ]	[ɔɪɫɚ], *[ɔɪəɫɚ]	(*Edmonton*) *Oiler(s)*
[smaɪəɫ]	*smile*	[smaɪɫɪŋ], [smaɪəɫɪŋ]		
[vaɪəɫ]	*vile*	[vaɪɫəst], [vaɪəɫəst]		
[maɪəɫ]	*mile*		[maɪɫədʒ], *[maɪəɫədʒ]	*mileage*
[naɪəɫ]	*Nile*		[naɪɫarɪk], *[naɪəɫarɪk]	*Nilotic*
[kəmpaɪəɫ]	*compile*	[kəmpaɪɫɪŋ], [kəmpaɪəɫɪŋ]	[kəmpaɪɫɚ], *[kəmpaɪəɫɚ]	*compiler*
[staɪəɫ]	*style*	[staɪɫɪŋ], [staɪəɫɪŋ]	[staɪɫɪst], *[staɪəɫɪst]	*stylist*[2]

A similar example comes from Yiddish, which generally avoids [rn] and [rm] codas by schwa epenthesis (Albright 2010): [alarəm] 'alarm', [ʃturəm] 'storm', [ʃirəm] 'umbrella', [turəm] 'tower' (CHAPTER 26: SCHWA; CHAPTER 67: VOWEL EPENTHESIS). When an /rm/ cluster is intervocalic, epenthesis does not normally occur: [alarm-ir-n] 'to alarm', [ʃturm-ɪʃ] 'stormy', [ʃirm-ə] 'screen', [turm-ə] 'prison'. This pattern is disrupted in verbal paradigms, however: (5a) shows that, if epenthesis applies somewhere within the paradigm, it systematically overapplies in the entire paradigm. This can be compared with (5b), which shows that, if

[2] A possible exception is ['staɪəˌtaɪz] *stylize*. It appears that stress clash avoidance may also play a role in conditioning [əɫ] in this form, however, since the derived noun form is [ˌstaɪɫə'zeɪʃən] *[ˌstaɪələ'zeɪʃən] *stylization*.

epenthesis is not conditioned anywhere within the paradigm, /rm/ surfaces uniformly faithfully. Overapplication in the verb [ʃturəm-ən] cannot be attributed to the influence of the related noun [ʃturəm], since the verb [alarm-ir-ən] does not show overapplication on the basis of the related noun [alarəm]. In addition, as the examples above show, epenthesis does not overapply in derived forms; additional examples include [varm-əs] 'warm food' and [refɔrm-ir-n] 'to reform'.

(5) *Overapplication of epenthesis in Yiddish*

a.	'storm'	Expected ≠	Actual	b.	'alarm'	Expected =	Actual
	INF	ʃturm-ən	ʃturəm-ən		INF	alarmir-ən	alarmir-ən
	1SG	ʃturəm	ʃturəm		1SG	alarmir	alarmir
	2SG	ʃturəm-st	ʃturəm-st		2SG	alarmir-st	alarmir-st
	3SG	ʃturəm-t	ʃturəm-t		3SG	alarmir-t	alarmir-t
	1PL	ʃturm-ən	ʃturəm-ən		1PL	alarmir-ən	alarmir-ən
	2PL	ʃturəm-t	ʃturəm-t		2PL	alarmir-t	alarmir-t
	3PL	ʃturm-ən	ʃturəm-ən		3PL	alarmir-ən	alarmir-ən

In both the English pre-lateral diphthongization and Yiddish epenthesis cases, phonology is blind to inflectional affixes – that is, a coda process applies as if the stem-final consonant is a coda, even if the inflectional affix should bleed epenthesis. It is not always the case that overapplication extends a process from an inner constituent to an outer constituent in this way, however. The Latin [honor] analogy discussed in §1 involved the overapplication of rhotacism, which should normally have occurred only before vowel-initial suffixes (e.g. /hono:s-is/ → [hono:ris]), but came to apply before nominative singular /s/ (or /Ø/) as well: /hono:s-s/ → [hono:r].

Another example of application of a process triggered by an inflectional affix is found in certain dialects of Korean (Han 2002; Kang 2003). In Korean, coronal obstruents regularly palatalize (CHAPTER 71: PALATALIZATION) before a suffix beginning with high front [i]: /os-i/ → [oʃi] 'clothing-NOM', /patʰ-i/ → [paʧʰi] 'field-NOM'.[3] Palatalization does not normally occur before mid front vowels ([patʰ-e] 'field-LOC') or high central vowels ([patʰ-ɨl] 'field-ACC'). Han (2002), citing data from Choi (1998), observes that in certain dialects of North Gyeongsang Korean (namely, in Sangju, Geumneung, Cheongdo, and Mungyeong), palatalization overapplies:

(6) *Overapplication of palatalization in North Gyeongsang Korean*

/patʰ/ 'field'	Conservative	North Gyeongsang
UNMARKED	pat̚	pat̚
NOM	paʧʰ-i	paʧʰ-i
ACC	patʰ-ɨl	paʧʰ-ɨl
DAT/LOC	patʰ-e	paʧʰ-e

In North Gyeongsang Korean, as in other dialects, the unmarked form obeys coda restrictions that neutralize continuancy and laryngeal contrasts (CHAPTER 69:

[3] There is some variability regarding the place of articulation of affricates, ranging from alveolar to post-alveolar or perhaps even palatal; see Cho (1967); Kim-Renaud (1974); Ahn (1998); Kim (1999); Sohn (1999).

FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). The change in North Gyeongsang consists of extending palatalization within noun paradigms, so that it overapplies before mid and central vowels. Unlike Latin or Yiddish, this case involves the overapplication of a process that is triggered by an inflectional affix.

As with Latin, we must ask whether the change could be interpreted instead as a change in the context of palatalization, or as a lexical re-analysis (CHAPTER 1: UNDERLYING REPRESENTATIONS). It is easy to demonstrate that the change is not a general expansion of the context for palatalization to include mid and central vowels. Han (2002) and Kang (2005) point out that, within verbal and adjectival paradigms, palatalization never affects /t^h-i/ sequences: /kat^h-in/ → [kat^{h}in], *[kaʧ^{h}ɨn] 'same-MODIFIER'. Thus, there is no evidence that palatalization in the accusative or locative reflects a broader change in the palatalization process itself. It is also important to ask whether the change is simply a re-analysis of the lexical entry to /paʧ^h/ – indeed, this is precisely what Kim (2005) claims. There are several reasons to think that a purely lexical account is not sufficient, however. First, assuming that the leveling in (6) has affected all /t^h/-final nouns in the relevant dialects and that speakers would reject new words with [ʧ^h] ~ [t^h] alternations, we need to account not only for the change to the specific lexical items in question, but also for the knowledge that there could be no lexical items of this type – i.e. a morpheme structure condition. McCarthy (1998) proposes to analyze such knowledge as a paradigm effect: specifically, speakers know that palatalization must apply before *-i*, and that it must overapply in the remainder of the paradigm by virtue of output–output faithfulness; therefore, no morpheme with /t^h/ could ever surface as such within its inflectional paradigm. Even more telling, it appears that overapplication of palatalization may be observed even in cases of partial leveling, where speakers continue to treat the noun as /t^h/-final in other contexts. Han (2002) and Kang (2005) discuss a related pattern found in the dialects of Gyeonggi, parts of Chungcheong, and North Jeolla, in which palatalization overapplies only before the accusative marker /-ɨl/, while the expected [t^h] surfaces faithfully before locative /-e/ and directive /-ɨlo/: unmarked [pat̚], nominative [paʧ^h-i], accusative [paʧ^h-ɨl], directive [pat^h-iro], locative [pat^h-e]. For these dialects, the use of [ʧ^h] in the accusative cannot be straightforwardly attributed to a re-analysis of the final consonant of the noun stem, since it surfaces as [t^h] elsewhere. Instead, Han (2002) claims that we are observing a more limited identity effect, in which the accusative comes to match the nominative, but the locative and directional remain identical and distinct from the other case forms. Kang (2005) attributes the identity of the locative and directive to the fact that they share locational meanings.

The Korean example shows that overapplication may extend processes that apply in inflected forms (i.e. overapplication is not "blind" to inflectional material). It also illustrates some of the difficulties in establishing a synchronic paradigm effect as opposed to a diachronic re-analysis. Many cases of overapplication that have been documented in the literature are potentially reinterpretable as lexical re-analyses or broadening of phonological processes. In order to demonstrate that a given case truly involves synchronic paradigmatically motivated overapplication, we must be able to show that the process continues to apply as expected elsewhere in the language, that the relevant lexical items have not been re-analyzed, and that the process overapplies just in those cases where paradigmatic identity would result. Similar considerations hold for identity through underapplication and misapplication, to which we now turn.

2.2 *Underapplication*

Just as in cases of underapplication of phonological processes to maintain reduplicative identity (McCarthy and Prince 1995; CHAPTER 100: REDUPLICATION; CHAPTER 119: REDUPLICATION IN SANSKRIT), it appears that phonological processes may be suppressed within inflectional paradigms in order achieve identity. An example can be seen in the behavior of Yiddish [rm] clusters. As shown in (5) above, schwa epenthesis overapplies in Yiddish verb paradigms, just in case epenthesis is expected in a suffixed form such as the 1st or 3rd singular. Interestingly, the converse pattern holds in noun paradigms. As shown in (7), noun plurals are sometimes formed by adding a consonant (/-s/), and sometimes by adding a vowel-initial suffix (usually /-ən/). In contrast to verbal paradigms, schwa epenthesis optionally *underapplies* in noun paradigms if it is not licensed in the plural.

(7) *Underapplication of epenthesis in Yiddish nouns*

	singular	*plural*	
a.	ʃturəm ~ *ʃturm	ʃturəms	'storm'
	ʃirəm ~ *ʃirm	ʃirəms	'umbrella'
	ʃvɔrəm ~ *ʃvɔrm	ʃvɔrəms	'swarm'
	vɔrəm ~ *vɔrm	vɔrəms	'worm'
	ɔrəm ~ *ɔrm	ɔrəms	'arm'
	turəm ~ *turm	turəms	'tower'
	alarəm ~ *alarm	alarəms	'alarm'
b.	fɔrəm ~ fɔrm	fɔrmən	'form'
	?unifɔrəm ~ unifɔrm	unifɔrmən	'uniform'
	*farəm ~ farm	farmən	'farm'

A similar example of parallel over- and underapplication comes from Polish diminutives, as discussed by Kraska-Szlenk (1995: 108–114) and Kenstowicz (1996). Many Polish nouns show an alternation between [u] and [ɔ] before word-final voiced non-nasal consonants, with [u] occurring in closed syllables and [ɔ] occurring in open syllables. The morphological context for raising differs depending on the gender of the noun, since different affixes create closed syllables in masculine *vs.* feminine nouns:

(8) *Regular application of o-raising in Polish nouns* (Kraska-Szlenk 1995: 108)

a. /dɔw/ 'ditch (MASC)'	*singular*	*plural*	b. /krɔv/ 'COW (FEM)'	*singular*	*plural*
NOM	duw	dɔwɨ	NOM	krɔva	krɔvɨ
GEN	dɔwu	dɔwuf	GEN	krɔvɨ	kruf
DAT	dɔwovi	dɔwom	DAT	krɔvje	krɔvom
ACC	duw	dɔwɨ	ACC	krɔvẽ	krɔvɨ
INSTR	dɔwem	dɔwami	INSTR	krɔvã	krɔvami
LOC	dɔle	dɔwax	LOC	krɔvje	krɔvax

In diminutive paradigms, raising alternations are suspended. Instead, the vowel that is expected in the nominative singular occurs everywhere. The diminutive suffix is (on the surface) [-ek] word-finally and [-k-] before a vowel, due to the presence of a fleeting yer vowel, indicated in (9) as /E/ (CHAPTER 122: SLAVIC YERS). As a result, diminutives of masculine nouns end in [-ek-Ø] in the nominative singular. Since the suffix begins with a vowel, raising does not apply in the nominative singular, and it correspondingly underapplies in the remainder of the paradigm (9a). Diminutives of feminine nouns, on the other hand, end in [-k-a]. This suffix does condition raising, and raising overapplies in the remainder of the paradigm (9b).

(9) *Under- and overapplication of raising in Polish diminutives*

a. /dɔw-Ek/ 'ditch (MASC)'	*singular*	*plural*	b. /krɔv-Ek-a/ 'cow (FEM)'	*singular*	*plural*
NOM	dɔwek	dɔ̲wkɨ	NOM	krufka	krufki
GEN	dɔ̲wka	dɔ̲wkuf	GEN	krufki	kru̲vek
DAT	dɔ̲wkovi	dɔ̲wkom	DAT	kruftse	krufkom
ACC	dɔwek	dɔ̲wki	ACC	krufkē	krufki
INSTR	dɔ̲wkjem	dɔ̲wkami	INSTR	krufká	krufkami
LOC	dɔ̲wku	dɔ̲wkax	LOC	kruftse	krufkax

As McCarthy (2005) points out (following Buckley 2001; Sanders 2003), it is not surprising that the distribution of *o*-raising has been disrupted in modern Polish: raising interacts opaquely with final devoicing, it is conditioned by an unusual set of following segments (non-nasal voiced consonants), and it is not generally applied to loanwords or nonce words. Nonetheless, the fact that raising applies normally in the non-diminutive forms of /dɔw/ and /krɔv/ (8) shows that the process is learned in some (perhaps lexically or morphologically restricted) form, and that it continues to apply in the expected contexts. Furthermore, speakers have clearly learned that these particular stems undergo raising, even in the nominative singular of the diminutive. Kraska-Szlenk (1995) and Kenstowicz (1996) argue that the lack of raising alternations in the paradigms in (9) is best attributed to a paradigm identity constraint, which holds specifically within diminutive paradigms.

2.3 *Misapplication*

In some cases, a regular phonological process applies in an unexpected fashion (misapplication). Harris (1973) argues that certain exceptions to regular stress placement in Spanish can be explained as identity effects, and that recognizing paradigmatic identity as a grammatical force could avoid the need to complicate the statement of stress rules. An example comes from the imperfect indicative forms of Spanish (discussed also by Burzio 2005: 65 and Oltra-Massuet and Arregi 2005):

(10) *Avoidance of stress alternations in the Spanish imperfect:* [terminar] 'finish'

	expected	*actual*
1SG IMP INDIC	termi'naba	termi'naba
2SG IMP INDIC	termi'nabas	termi'nabas
3SG IMP INDIC	termi'naba	termi'naba
1PL IMP INDIC	termina'bamos	termi'nabamos
2PL IMP INDIC	termina'bais	termi'nabais
3PL IMP INDIC	termi'naban	termi'naban

The stress pattern as it would have been inherited from Latin is shown in the left column of (10); this expected pattern also reflects a general pattern of penultimate stress that still holds in modern Spanish. The actual stress pattern in [termi'nabamos] is unusual in Spanish, since words with final closed syllables rarely have antepenultimate stress, the sole exceptions being learned words such as *Sócrates* and *Júpiter*, which are felt by speakers to be unusual (Hochberg 1988). This unexpected pattern is immediately explained if we look at the stress pattern of the other cells of the imperfect, where stress is penultimate according to the regular principles of Spanish stress assignment.

Harris (1987) provides another example from certain dialects of Spanish, including Chicano Spanish. As (11a) shows, stress in verbs regularly falls on the penult, both in conservative varieties and in Chicano Spanish. Within the present subjunctive paradigm, however, stress alternations have been eliminated, with stress falling on the antepenult in the 1st plural (11b) (see also Reyes 1974). Once again, a possible explanation for this subversion of the regular stress pattern is that it maintains identity among inflected forms. Note that this effect holds locally within the present subjunctive, but does not affect the present indicative, nor does it affect non-verbal forms such as ['termino] 'end'.

(11) *Avoidance of stress alternations in Chicano Spanish subjunctive:*
[terminar] 'finish'

		Conservative	*Chicano*
a.	1SG INDIC	ter'mino	ter'mino
	2SG INDIC	ter'minas	ter'minas
	3SG INDIC	ter'mina	ter'mina
	1PL INDIC	termi'namos	termi'namos
	2PL INDIC	termi'nais	termi'nais
	3PL INDIC	ter'minan	ter'minan
b.	1SG SUBJ	ter'mine	ter'mine
	2SG SUBJ	ter'mines	ter'mines
	3SG SUBJ	ter'mine	ter'mine
	1PL SUBJ	termi'nemos	ter'minemos
	2PL SUBJ	termi'neis	—
	3PL SUBJ	ter'minen	ter'minen

Kenstowicz (1996) observes a similar effect for certain Russian nouns, in which stress unexpectedly avoids falling on a fleeting yer vowel in the genitive plural ("double retraction"). Like the Spanish cases, this irregular placement of stress serves to avoid stress alternations within the plural paradigm, and may likewise be attributed to a constraint demanding uniform stress.

2.4 *Anti-homophony within paradigms*

In addition to serving as the domain for identity effects, inflectional paradigms also appear to be the locus of anti-homophony effects, in which phonology applies unexpectedly in order to avoid surface identity of morphologically distinct forms. The scope and treatment of anti-homophony effects is controversial, but we mention here two examples.

The first example comes from Kenstowicz (2005), who discusses the distribution of schwa in medial open syllables in Damascus Arabic (CHAPTER 124: WORD STRESS IN ARABIC). Normally in this dialect, unstressed schwas in open syllables are deleted: /səməʕ-ət/ → ['səmʕ-ət] 'heard-3SG FEM SUBJ'. This deletion also occurs in forms that have been cliticized with object markers: both /dˤarb-ət-o/ → ['dˤarbto] 'hit-3SG FEM SUBJ-3SG MASC OBJ' and /ʃaːf-ət-o/ → ['ʃaːfto] 'saw-3SG FEM SUBJ-3SG MASC OBJ' show deletion of the underlying /ə/ in the 3rd singular feminine subject marker. However, this deletion is systematically blocked in cases where the form in question would become identical with the corresponding masculine form. For example, /ʕallam-t-o/ 'taught-3SG MASC SUBJ-3SG MASC OBJ' and /ʕallam-ət-o/ 'taught-3SG FEM SUBJ-3SG MASC OBJ' are expected to yield surface [ʕal'lamto], with syncope of the stressless schwa. Instead, syncope is blocked in the feminine form 'she taught him', and the schwa exceptionally remains and attracts stress: [ʕalla'mәto]. Kenstowicz (2005) attributes this unusual stress pattern, which violates the normal principles of stress placement in Damascus Arabic, to an anti-homophony condition that holds between the 3rd singular masculine and feminine forms. Importantly, this condition does not penalize deletion in cases such as ['dˤarbto] 'she hit him' or ['ʃaːfto] 'she saw him', because for these verbs, the 3rd singular masculine forms happen to follow different vocalic templates: [da'rab-t-o] 'he hit him', ['ʃuf-t-o] 'he saw him'.

Hall and Scott (2007) discuss another example, involving underapplication in Swabian German. In this dialect, /s/ becomes [ʃ] before a coronal: /pɔst/ → [pɔʃt] 'mail'. This process underapplies in inflectional paradigms, however: /griəs-t/→ [griəst], *[griəʃt] 'greet-3SG'. Hall and Scott attribute this underapplication to the influence of inflectionally related forms with [s], such as 1st singular [griəs]. In the 2nd singular, however, the inflectional affix is /-ʃ/. Here, underapplication of /s/ → [ʃ] would yield the illegal form [griəsʃ], which would be expected to assimilate and degeminate. The candidate [griəs], which resolves the /sʃ/ cluster in favor of [s], would create paradigmatic identity with 1st singular [griəs], 3rd singular [griəst], etc., but it would be homophonous with the 1st singular. Instead, the actual outcome is [griəʃ], in violation of paradigm uniformity. Hall and Scott argue that this is due to an anti-homophony condition, in which the 2nd singular is required to be distinct from other forms.

In both the Damascus Arabic and the Swabian German examples, the paradigm is the locus of anti-homophony effects, in the sense that homophony between forms that are not paradigmatically related (as in accidentally homophonous verbs) does not trigger unexpected phonology. If this restriction turns out to hold in a broader range of cases, it could serve as an additional source of evidence that paradigms may serve as the domain of grammatical effects.

3 Grammatical mechanisms for deriving paradigm effects

The fact that the regular application of phonological processes may be disrupted in order to achieve paradigmatic regularity was often commented on in the earlier literature on generative phonology (see, e.g. Kiparsky 1972; Harris 1973; Kenstowicz and Kisseberth 1977: 69–74). However, no formal mechanism was provided for accomplishing this effect within the rule-based approach of Chomsky and Halle (1968; henceforth *SPE*). Indeed, it is hard to see what a mechanism that explicitly enforces paradigmatic identity would look like, since the decision about whether to apply a rule or not (or to reverse the order of two rules) can be made only by "looking ahead" and seeing whether the results would create alternations within the paradigm. Instead, inflectional identity has often been analyzed as a by-product of how morphological and phonological operations are ordered, under the hypothesis that inflectional affixes are added after many phonological processes have already applied. (See Downing *et al.* 2005a for a review of alternative rule-based mechanisms, and §3.3 below for discussion of cyclic approaches employing phase-based spell-out; see also CHAPTER 85: CYCLICITY and CHAPTER 74: RULE ORDERING.)

For example, the American English pre-lateral diphthongization data in (4) ([spɔɪəɫ] *spoil*, [spɔɪəɫɪŋ] *spoiling* vs. [spɔɪɫəʤ] *spoilage*) can be seen as the result of diphthongization applying after derivational affixes like *-age* have been added, but before inflectional affixes have been added. A challenge for such approaches is to establish an internally consistent ordering of morphological and phonological operations. This is not always trivial; for example, in the author's idiolect, the fact that /l/ is at least optionally dark in words like [spɔɪɫəʤ] *spoilage* and [maɪɫəʤ] *mileage* suggests that velarization is ordered before affixation of *-age*, yet there is no option for dark [ɫ] in words like [saɪləʤ] *[saɪɫəʤ] *silage*, which arguably also contains an /l/ before the *-age* suffix. Accounts based on cyclic ordering have no direct way of referring to the fact that for *spoilage* and *mileage*, related forms such as [spɔɪəl] *spoil* and [maɪəl] *mile* contain dark [ɫ], while for *silenge*, the related form [saɪloʊ] *silo* has light [l].

Paradigm effects of this sort are readily accommodated within Optimality Theory (OT), since evaluation in this framework is carried out over surface forms, which is precisely where identity and contrast must be enforced. In this section, we review the main approaches to enforcing identity among paradigms of inflected forms in OT.

Most approaches to paradigm effects in OT employ correspondence constraints (McCarthy and Prince 1995), which place morphologically related forms into correspondence with one another, so that identity can be evaluated with the standard machinery of faithfulness constraints (IDENT, MAX, DEP; see also CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). Under this approach, paradigmatic identity effects are closely related to cyclicity effects, in which derived forms show unexpected similarity to their bases of derivation. Correspondence relations are intrinsically pairwise, so paradigm uniformity is typically enforced by requiring that each individual pair of forms be identical. Thus, in a sense, paradigm constraints are not truly "paradigmatic," since they evaluate sets of pairs rather than the entire distribution at once. A cost of this approach, however, is that the number of pairwise relations that must be considered grows factorially with the size of the paradigm – e.g. $6 \times 5 = 30$ pairwise relations for a six-member paradigm.

Since their introduction, correspondence constraints have been used to analyze the relation between derived forms and their bases in producing cyclicity effects (Burzio 1994; Kenstowicz 1996; Benua 1997; Kager 2000; Steriade 2000). Like input–output faithfulness conditions, base–derivative correspondence is usually assumed to be intrinsically asymmetrical: derived forms must be faithful to their bases, but bases are not constrained to resemble their derivatives (see also

CHAPTER 85: CYCLICITY). There are two main issues that must be resolved in applying faithfulness conditions to paradigm regularity. First, we must determine whether correspondence relations between inflectionally related forms are intrinsically symmetrical, as is generally assumed for base–reduplicant faithfulness (McCarthy and Prince 1995), or asymmetrical, as is usually assumed for other cases of base-derivative faithfulness (Benua 1997). Second, if inflected forms are faithful to a base form, we must determine which form acts as the base in an inflectional paradigm.

In this section, the predictions of a symmetrical approach to paradigm uniformity (McCarthy's 2005 Optimal Paradigms approach) are compared to those of an asymmetrical/base-prioritizing approach (Albright 2002; Kenstowicz 1996; Benua 1997), as well as to a stratal approach in which outputs are evaluated cyclically with interleaved levels of affixation and phonological evaluation (Kiparsky 1982, 2000).

3.1 Symmetrical output–output faithfulness: Optimal paradigms

One approach to enforcing paradigmatic regularity is through symmetrical correspondence relations, which simply demand that all inflected forms have identical forms of the root. This symmetrical identity requirement has gone under various names in the literature, including *consistency* (Burzio 1994, 2005; see also CHAPTER 88: DERIVED ENVIRONMENT EFFECTS), *uniform exponence* (Kenstowicz 1996), *paradigm uniformity* (Steriade 2000), and, most recently, *optimal paradigms* (McCarthy 2005). McCarthy (2005: 172) succinctly sums up the motivation for adopting a symmetrical approach to identity in paradigms: "Inflectional paradigms have no base . . . : Latin *amat* 'he loves' is not derived from *amō* 'I love' or vice versa; rather, both are derived from the lexeme /am-/." The claim that inflected forms are based on a common lexeme but not on each other rests on a morphological notion of *base of affixation*, in which affixation realizes (or marks) morphological features, and affixed forms contain a superset of the morphological information that their inner constituents contain. Under this definition, the base must necessarily contain a subset of the morphological features of derived forms. Depending on the morphological features that one assumes, it is implausible to suppose that the 1st singular "contains" the 3rd singular (though for feature systems that make use of underspecification, see Harley and Ritter 2002; McGinnis 2005). A similar point is made by Kager (1999: 282), who refrains from positing a relation between the 1st singular and the 2nd plural, since neither one appears to compositionally contain the other.

Because no individual form has priority in a symmetrical approach, paradigms must be evaluated as a set, in case a high-ranking constraint demands a modification in one form that would then overapply in the rest of the paradigm. In McCarthy's (2005) Optimal Paradigms (OP) formulation, candidates consist of entire paradigms. The stem in each cell in the paradigm stands in correspondence to the stem in every other cell, and output-output faithfulness constraints (OP-IDENT, OP-MAX, OP-DEP) evaluate every pair of surface forms in the paradigm. Paradigm uniformity effects arise when OP-faithfulness outranks the relevant IO-faithfulness constraint. This is illustrated in (12) for the overapplication of schwa epenthesis in Yiddish, discussed in (5) above.

(12) *Optimal Paradigms analysis of overapplication of schwa epenthesis in Yiddish*

/ʃturm, ʃturm-t, ʃturm-ən/	OP-Dep	*rm]	IO-Dep
a. ʃturm, ʃturmt, ʃturmən		*!*	
b. ʃturəm, ʃturəmt, ʃturmən	*!*		**
☞ c. ʃturəm, ʃturəmt, ʃturəmən			***

The tableau in (12) illustrates several important features of the OP approach. The ranking of *rm] over IO-Dep reflects the fact that there is a general process of epenthesis to repair rm] codas. If this ranking did not hold, there would be no possibility of alternations. With this ranking in place, it is possible to rank OP-Dep high, in order to favor an invariant paradigm with overapplication (candidate (12c)), or low, in order to favor normal application (candidate (12b)). Crucially, the only rankings that favor candidate (12a) (underapplication) are those in which IO-Dep outranks *rm] – but these are incompatible with the fact that the language generally has a process of epenthesis. Thus, there is no ranking in which epenthesis is found in general, but underapplies in order to maintain paradigm uniformity. McCarthy dubs this prediction *overapplication-only*. Furthermore, the relative number of intervocalic *vs.* coda /rm/ sequences in the paradigm cannot affect the outcome, since the logic of strict domination dictates that even one instance of coda /rm/ may be fixed by any number of IO-Dep violations. Thus even if we had considered a paradigm with only a single input like /ʃturm/ and many inputs like /ʃturm-ən/, overapplication of epenthesis would still be the optimal way to satisfy OP-Dep.

On the face of it, the overapplication-only prediction of OP is too strong, since, as we saw in §2.2, paradigm identity may also be achieved through underapplication – in fact, the very same process of epenthesis underapplies in Yiddish nouns. The only mechanism available to handle such cases in the OP framework is to find some higher-ranked constraint that would be violated by overapplication. As a hypothetical example of what such a constraint might be, suppose Yiddish were to borrow the English word *platform*, retaining initial stress and creating an associated plural ['platfɔrmən]. In this case, overapplication of schwa epenthesis would yield a plural form with pre-antepenultimate stress and a long stress lapse: ['platfɔrəm] ~ ['platfɔrəmən]. As it turns out, Yiddish stress, though somewhat unpredictable, must fall within the last three syllables of the word (see Jacobs 2005: 135–140), so the output ['platfɔrəmən] would be illegal. In this hypothetical example, we could analyze underapplication of epenthesis (['platfɔrm] ~ ['platfɔrmən]) as overapplication of lapse avoidance (['platfɔrəm] ~ *['platfɔrəmən]), consistent with the overapplication-only prediction.

(13) *Underapplication as overapplication of an orthogonal process (hypothetical)*

/platfɔrm, platfɔrm-ən/	OP-Dep	OP-Ident (stress)	Stress	IO-Ident (stress)	*rm|	IO-Dep
☞ a. 'platfɔrm, 'platfɔrmən					*	
b. 'platfɔrəm, 'platfɔrmən	*!*					**
c. 'platfɔrəm, 'platfɔrəmən			*!			***
d. 'platfɔrəm, plat'fɔrəmən		*!*		*		
e. plat'fɔrəm, plat'fɔrəmən				*!*		

Unfortunately, in the actual cases of underapplication in Yiddish nouns in (7b), it is not clear what such a constraint would be. In particular, the contrasting pair [vɔrəm] 'worm' *vs.* [fɔrm] 'form' do not appear to differ in any relevant way that could be capitalized on in order to block epenthesis in the latter case by means of a higher-ranked markedness constraint.

A similar point can be made about the Polish diminutive example in (9). Here, the cases in which raising underapplies in the masculine find exact counterparts in which raising applies normally in the feminine: e.g. accusative plural [dɔwki] 'little ditches' *vs.* [krufki] 'little cows'. There is no form in the paradigm of [dɔwek] in which raising would not violate a constraint that is also violated in the paradigm of [krufka]. In particular, hypothetical overapplication of raising in the nominative and accusative singular (*[duwek]) would violate the same constraints that are violated by the attested overapplication of raising in the feminine genitive plural ([kruvek]). As Kenstowicz (1996) points out, symmetric faithfulness constraints are equally well satisfied by over- and underapplication, and it is not obvious what additional constraint would break the tie. This problem leads McCarthy (2005) to suggest that raising alternations must be lexicalized in this case, reflecting the more generally unproductive nature of *o*-raising in Polish. As noted in §2, it is often difficult to determine whether an attested case of underapplication represents a synchronic paradigm effect, or whether it is the outcome of a diachronic re-analysis in which the relevant alternation simply no longer applies, even when paradigmatic identity is not at stake. Although it appears that at least some cases of underapplication demand a synchronic analysis, more careful investigation is needed to determine whether they are more appropriately interpreted as diachronic effects rather than as synchronic exceptions to the overapplication-only prediction.

Misapplication can also pose a challenge to the symmetrical faithfulness model. The stress patterns of the Spanish imperfect (10) and (dialectal) present subjective (11b) show the otherwise practically non-occurring pattern of antepenultimate stress with a closed final syllable ([termi'nabamos], [ter'minemos]). The candidate in which antepenultimate stress is avoided by placing stress too far to the right in other forms (hypothetical [termina'ba], [termina'bas], [termina'ba], [termina'bamos], [termina'bais], [termina'ban]) would fare much better by the constraint that favors penultimate stress with final heavy syllables. This candidate involves a different dispreferred pattern – namely, forms with stress on final open syllables ([termina'ba]). This stress pattern is somewhat rare, but it occurs in native monomorphemic words ([a'ki] 'here', [so'fa] 'sofa', [xo'se] 'José', etc.) to a far greater extent than antepenultimate stress in words like *Sócrates*. Moreover, invariant suffixal stress is found in other tenses, including the future ([termina're], [termina'ras], [termina'ra], [termina'remos], [termina'reis], [termina'ran]). Thus it appears that the OP approach would favor a paradigm that avoids the stress violation in (actual) [termi'nabamos] by placing stress one syllable to the right (*[termina'ba], *[termina'bamos]).

3.2 Base-prioritizing output–output faithfulness

An alternative possibility is that faithfulness among inflected forms is exactly parallel to the asymmetrical structure of faithfulness among derived forms. Asymmetrical, base-prioritizing faithfulness has been given various names in the literature, including *base identity* (Kenstowicz 1996) and *transderivational correspondence* (Benua 1997). In an asymmetric approach, dependent forms are constrained to be faithful to a designated base form, but the base form is not constrained to match the rest of the paradigm. Thus, we expect the base form to exhibit normal application of regular phonological processes, while the remaining forms may show over- or underapplication in order to resemble the base. Since surface properties of the base form must be known ahead of time in order to evaluate base-faithfulness violations, base-prioritizing faithfulness requires a form of cyclic, or recursive, evaluation. In the discussion that follows, I follow Benua (1997) in assuming that evaluation proceeds in two steps: first the base is evaluated, and then its dependent forms are evaluated. Although recursive evaluation involves more steps than a single, parallel evaluation, the evaluation of output–output faithfulness requires a fraction of the comparisons that are involved in the full pairwise evaluation of the symmetric approach.

An asymmetric approach provides a natural solution to many of the problems pointed out in the preceding section. In the case of stress in the Spanish imperfect and (dialectal) present subjunctive, the unexpected pattern in [termi'nabamos] 'we ended' can be attributed to faithfulness to other forms in the paradigm, in which stress obeys the usual pattern of penultimate stress. Deferring for a moment the question of what form should be designated as the base in general, let us assume for now that the base of Spanish verb paradigms (or, at least, of the imperfect and subjunctive) is the 3rd singular: [termi'naba]. In the base-prioritizing approach, properties of the base are determined "first" according to the general rankings of markedness and IO-faithfulness, and this pattern is then transferred to related form via base faithfulness, as shown in (14). We assume that Spanish has constraints against final stressed vowels (*'V]) and against antepenultimate stress with final closed syllables (*'σσσ̄]). We also assume that, although antepenultimate stress is sometimes found on vowel-final words, the default is penultimate stress (Aske 1990; Eddington 2000; Oltra-Massuet and Arregi 2005), and that this is favored by a constraint against final lapse (*'σσσ], not shown). These are overruled, however, by a constraint demanding faithfulness to the stress pattern of the 3rd singular base. (Following Benua 1997, the tableau for the non-basic 1st plural form is indented to show that the evaluation of BASE-IDENT requires reference to the output of the 3rd singular evaluation.)

(14) *Misapplication in the Spanish imperfect*

/termina-ba-Ø/ 'end-3SG'	*BASE-IDENT(stress)	*'σσσ]	*'V]
☞ a. termi'naba			
b. termina'ba			*

/termina-ba-mos/ 'end-1PL'	*BASE-IDENT(stress)	*'σσσ]	*'V]
☞ a. termi'nabamos		*	
b. termina'bamos	*!		

Base identity also addresses the problem of *o*-raising in Polish diminutives. Again deferring the discussion of what form serves as the base in general, let us assume that the base of Polish noun paradigms is the nominative singular. The evaluations of the masculine and feminine nominative diminutives are shown in parallel for /dɔw-(E)k/ and /krɔv-(E)k-a/ in (15). Since these forms are basic, they vacuously satisfy BASE-IDENT. The constraint RAISE is shorthand for whatever constraint motivates [u] in closed syllables; recall that the paradigms of simple (non-diminutive) [duw] 'ditch' and [krɔva] 'cow' in (8) show that both of these stems participate in raising. For brevity, the constraints that are responsible for the vocalism of the diminutive suffix are not shown.

(15) *Normal application of Polish raising in nominative singular base forms*

/dɔw-Ek-Ø/ 'ditch-DIM-NOM SG'	BASE-IDENT[high]	RAISE	IO-IDENT[high]
☞ a. dɔwek			
b. duwek			*

/krɔv-Ek-a/ 'COW-DIM-NOM SG'	BASE-IDENT[high]	RAISE	IO-IDENT[high]
☞ a. krɔfka		*!	
b. krufka			*

The remaining inflected forms are then constrained by BASE-IDENT to preserve the vowel height of the nominative singular, as in (16).

(16) *Over- and underapplication of Polish raising in non-basic forms*

/dɔw-Ek-a/ 'ditch-DIM-GEN SG'	BASE-IDENT[high]	RAISE	IO-IDENT[high]
☞ a. dɔwka		*	
b. duwka	*!		*

/krɔv-Ek-Ø/ 'COW-DIM-GEN PL'	BASE-IDENT[high]	RAISE	IO-IDENT[high]
a. krɔvek	*!		
☞ b. kruvek			*

By hypothesis, BASE-IDENT for inflected forms is a distinct constraint from whatever base–derivative faithfulness constraints may hold of derived forms. This predicts the possibility of different degrees of faithfulness for inflected *vs.* derived forms. As we saw above for Latin, inflected forms may indeed exhibit greater faithfulness than derived forms. A similar point can be made for Yiddish, in which derivationally related forms show schwa alternations: cf. [alarəm] 'alarm' ~ [alarm-ir-n] 'to alarm', [ʃirəm] 'shade' ~ [ʃirm-ə] 'screen', [varəm] 'warm' ~ [varm-əs] 'warm food'. What is not explained by this approach, however, is the cross-linguistic tendency for inflected forms to show greater uniformity than derived forms. If inflectional BASE-IDENT and base-derivative faithfulness are separate and independently re-rankable constraints, we predict that some languages may show greater uniformity in derived forms than in infected forms.

The analysis in (15) and (16) requires that the nominative singular be designated as the base, even though other inflected forms are not necessarily built compositionally from it. For the masculine diminutive [dɔwek], this coincides with the fact that nominative singular is in a substring of the remaining forms. In particular, although the nominative singular differs from the remaining infected forms in having an [e] in the diminutive suffix, its suffix is -Ø. As a result, the underlying form /dɔw-Ek-Ø/ is a phonological substring of the underlying form of inflected forms like /dɔw-Ek-a/. In many cases, it is sufficient to posit that null-affixed forms (sometimes termed *isolation forms*) serve as the base for overtly affixed inflected forms; see, e.g. Kurylowicz (1947); Kenstowicz (1996); Hayes (1999); Hall and Scott (2007).

The comparison with the Polish feminine diminutive form [krufka] shows that base forms need not always be isolation forms, however. In this noun class, there is an overt nominative singular affix /-a/ that is not contained in other case/number forms. A common intuition is that even if the feminine nominative singular is phonologically marked (in the sense of having an overt affix), it is nonetheless a plausible base form because it represents a morphologically "unmarked" category (in the sense of serving as a default or unmarked member of an opposition). Analyses that make use of this more general notion of morphosyntactic markedness are often somewhat vague as to what the criteria are for identifying the morphologically least marked member of a paradigm, but the general consensus appears to be that it is the nominative singular in a Latin or

Polish-like nominal paradigm (Kenstowicz 1996; Sturgeon 2003), and the 3rd singular in a verbal paradigm (Kuryłowicz 1947; Mańczak 1958).

The idea that the base of a paradigm is a morphologically unmarked form (null affix, or unmarked feature values) is attractive from the point of view of grounding the structure of paradigmatic correspondence in a universal, morphosyntactically motivated representation. There are cases that are difficult to reconcile with this hypothesis, however. The Latin [honor] example discussed in §1.2 is one potential counterexample, since it shows overapplication in the nominative singular on the basis of morphosyntactically more marked case forms (oblique and plural forms). The underapplication of schwa epenthesis in Yiddish noun paradigms is also a counterexample. Recall from (7b) that epenthesis may fail to apply in final rm] clusters just in case the plural has a vowel-initial suffix: [vɔram] ~ [vɔram-s] 'worm-SG/PL' (with epenthesis) *vs.* [fɔrm] ~ [fɔrmən] 'form-SG/PL' (epenthesis optionally underapplies). This case is readily accounted for if we posit faithfulness to the plural form. This is illustrated in (17) for the variant [fɔrm] 'form-SG'.

(17) *Underapplication of schwa epenthesis in Yiddish nouns*

/fɔrm-ən/ 'form-PL'	Base-Dep	*rm]	IO-Dep
☞ a. fɔrmən			
b. fɔrəmən			*

/fɔrm/ 'form-SG' Base: [fɔrmən]	Base-Dep	*rm]	IO-Dep
☞ a. fɔrm		*	
b. fɔrəm	*!		*

Under this analysis, the fact that epenthesis underapplies in noun paradigms is attributed to faithfulness to the plural, while the fact that epenthesis overapplies in verb paradigms must be attributed to faithfulness to a form in which normal application would favor epenthesis (e.g. 1st singular: Albright 2010). If this is correct, it suggests that the choice of privileged base form may differ by language and by part of speech (see also CHAPTER 102: CATEGORY-SPECIFIC EFFECTS). This conclusion naturally raises a number of questions: how do learners identify bases? Is it possible to predict which form will act as the base in a given part of speech in a given language, or must it be inferred post hoc from unexpected application of regular phonological processes? Are there limits to which form acts as the base? And if the choice of base can vary freely from language to language, why do certain forms such as the "unmarked" nominative singular and 3rd singular so often act as bases?

Albright (2002) proposes that the choice of base in a given language is not arbitrary, but follows from the distribution of contrasts within paradigms (CHAPTER 2: CONTRAST). Specifically, it is hypothesized that learners identify the form that offers the most phonological and morphological information about lexical items. For example, in Yiddish verbs, some inflectional affixes trigger neutralizing phonological processes such as regressive voicing assimilation (/red-st/ → [retst] 'speak-2SG'), degemination (/red-t/ → [ret] 'speak-3SG'; /heɪs-st/ → [heɪst] 'call-2SG') or coalescence of schwas (/blɔnkə-ən/ → [blɔnkən] 'meander-1PL') (CHAPTER 80: MERGERS AND NEUTRALIZATION). As it turns out, since the 1st singular suffix is null and Yiddish has relatively few processes that affect final codas, the 1st singular frequently preserves phonological contrasts that are neutralized elsewhere. One process that does occur word-finally (including in the 1st singular) is epenthesis in [rm] clusters, however – and as we have seen, this process overapplies in verb paradigms. For nouns, on the other hand, pluralization involves a high degree of morphological unpredictability, with several competing plural markers and irregular vowel changes to many stems. Thus, the plural often contains morphological information that is neutralized in the singular (Albright 2008b). This correlates with the fact that schwa epenthesis optionally underapplies in the singular if it does not apply in the plural.

Informativeness effects that run counter to markedness can be observed in a number of languages. In Latin, the nominative singular underwent many phonological and morphological neutralizations, including cluster simplification and morphological syncretism, which were not found in oblique forms.[4] Albright (2005) provides a quantitative comparison of neutralizations affecting Latin noun forms. This analysis reveals that the nominative singular was the least informative form, which correlates with the fact that nominative singular forms were rebuilt in Latin. An even more striking example comes from Korean verbal inflection. In a survey of dialects and acquisition studies, Kang (2006) shows that a large number of phonologically independent alternations have been eliminated across different dialects and varieties of Korean. However, all of these changes have in common the property that they extend the stem form found before a certain suffix, the informal form /-ə ~ a/. Albright and Kang (2009) show that this form is also the most informative, revealing stem-final consonant and vowel contrasts more clearly than other suffixed forms.

The typological predictions of an information-based approach are less certain. One factor that appears to encourage base status is token frequency (Kuryłowicz 1947; Mańczak 1958). Plausibly, cells in the paradigm with high-token frequency (CHAPTER 90: FREQUENCY EFFECTS) provide more information to learners, since (on average) more lexical items have been encountered in these forms. Albright (2008a) argues that attestation is by itself also an important type of information, which may bias learners to choose more frequent paradigm members as bases. This, in turn, may explain the fact that morphologically unmarked forms tend to be bases, since they often have the highest token frequency (Bybee 1985: ch. 3).

An important prediction of the base-prioritizing faithfulness account is that the choice between over- and underapplication is a straightforward consequence of whether the process applies in the base form or not. The comparison of masculine and feminine diminutives in Polish shows that even within the same language, this may have different consequences for different words. Another prediction is that the same member of the paradigm should act as basic for multiple dimensions of faithfulness. Albright (2008b) shows that for Yiddish nouns, this is true: the plural acts as a privileged base in conditioning underapplication of

[4] The importance of oblique forms in identifying phonological and inflectional properties of Latin nouns can be seen from the fact that dictionaries typically list an oblique form (the genitive singular) alongside the nominative.

/rm/-epenthesis, underapplication of final devoicing, and overapplication of open syllable lengthening. Similarly, Latin showed leveling of not only rhotacism alternations, but also vowel length and nominative singular marking (*-s* > *-is*: Kiparsky 1998; Albright 2005). Finally, the changes to Korean verb paradigms discussed by Kang (2006) show parallel changes for a large number of logically independent alternations. If this prediction proves true in general, it could provide strong support for an asymmetric approach in which paradigm uniformity extends an independently designated base form.

3.3 *Stratal and cyclic approaches*

The approaches discussed above rely on correspondence relations between related surface forms to enforce identity. An alternative approach, which denies the existence of paradigms altogether, attributes the phonological similarity of related forms to the fact that they share morphological and syntactic structure, and are thus (by hypothesis) identical at a certain stage in their derivation. For example, the framework of Lexical Phonology (Kiparsky 1982) and its successor, Stratal OT (Kiparsky 2000), assumes that phonological evaluation and affixation are interleaved, so that phonological processes may apply to inner constituents of the word prior to the addition of later affixes. Work in Lexical Phonology (LPM) has explored the implications of this approach for cyclicity effects in derivational morphology; see especially Giegerich (1999) and CHAPTER 85: CYCLICITY for overview and discussion (also CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME). Importantly, it is assumed that inflectional affixes are added in the last level of affixation, and are therefore ignored by all earlier cycles of phonological evaluation. This suggests an analysis of paradigm uniformity effects that is parallel to the stratal analysis of cyclicity effects in derivational morphology: processes apply incorrectly when phonology evaluates a morphological subconstituent that does not yet contain the relevant infectional affixes. This is sketched in (18) for the overapplication of epenthesis in Yiddish verbs.

(18) *LPM approach to overapplication of schwa epenthesis in Yiddish verbs*

a. *Stem level: Inflectional affixes not present*

/ʃturm-/	IO-MAX	*rm]	IO-DEP
a. ʃturm		*!	
☞ b. ʃturəm			*

b. *Word level: Inflectional affixes added*

/[ʃturəm]-ən/	IO-MAX	*rm]	IO-DEP
a. ʃturmən	*!		
☞ b. ʃturəmən			

Under this approach, the lack of epenthesis in paradigms of verbs such as *alarm-ir-n* 'to alarm' would be attributed to the fact that verbalizing *-ir-* is a stem-level affix, attaching to bound stems and occurring inside other affixes such as *-ist*.

(19) *LPM approach to normal application in* alarmirn

a. *Stem level: Derivational -ir- is present*

/alarm-ir/	IO-Max	*rm]	IO-Dep
☞ a. alarmir			
b. alarəmir			*!

b. *Word level: Inflectional affixes added*

/[alarmir]-ən/	IO-Max	*rm]	IO-Dep
☞ a. alarmirən			
b. alarəmirən			*!

A closely related approach relies on the syntactic mechanism of derivation by phase (Chomsky 2001) to evaluate successively larger portions of the word (Marvin 2002; Arad 2003; Piggott and Newell 2006; Marantz 2007; Skinner 2008). Under this approach, it is likewise hypothesized that words are built cyclically, and that certain syntactic heads trigger spell-out and phonological evaluation of the structure that has been built thus far. We assume that verbs like [alarmirn] 'alarm' have a structure [[[alarm]$_{\sqrt{}}$-ir]$_v$-n] and verbs like [ʃturəmən] 'storm' have a structure [[[ʃturm]$_{\sqrt{}}$-Ø]$_v$-ən], with a null verbalizing head. Under the assumption that verbalizing heads trigger spell-out (Marantz 2007) and that the "little v" head is spelled out with (or is at least visible to) its complement, the innermost spell-out domain of [alarmirn] consists of /alarm-ir/ (no epenthesis necessary), whereas the inner spell-out domain of [ʃturəmən] consists of /ʃturm-Ø/ (epenthesis expected). Thus derivations similar to those in (18) and (19) would hold in this approach as well.

Stratal and cyclic approaches evaluate morphologically complex words strictly from the inside outwards. Crucially, there is no way for affixes that are not present to influence the outcome of phonology. For this reason, effects such as the underapplication of epenthesis in Yiddish nouns are unexpected, since the outcome for singular forms like [vɔrəm] 'worm' *vs.* [fɔrm] 'form' requires reference to the shape of the plural ([vɔrəm-s] *vs.* [fɔrm-ən]). More generally, any case in which an output-output correspondence account requires reference to a base form that is not an isolation form or substring of the remaining forms poses a potential challenge to cyclic approaches. It remains a question for future research to determine whether such cases have a different synchronic status from the more common cases of "inside → out" directionality of cyclic influence.

4 Defining and constraining paradigms

Up to this point, we have been intentionally vague as to what forms count as belonging to a single paradigm, relying on an intuition notion of sets of forms that share a root, and in some cases also a set of inflectional features (e.g. imperfect indicative or present subjunctive). One possibility that cannot be immediately discarded is that output-output correspondence relations are established along many dimensions of shared inflectional features (imperfect forms, 3rd singular forms, subjunctive forms, etc.). This unstructured notion of paradigmatic relations appears to miss some general tendencies, however. In point of fact, paradigmatic similarity appears to hold much more strongly within certain sets of forms, such as the set of person/number forms in a given tense and mood. In this section, we briefly review some of the major tendencies that have been identified in the literature, and consider where these tendencies may come from.

The examples in the preceding sections show several recurring patterns. For example, person and number inflections of a verb are often constrained to resemble each other within a given tense, while alternations are tolerated across tenses – e.g. Spanish 1st singular present [ter'mino] ~ 1st singular imperfect [termi'naba]; 1st plural [termi'naban]. Although I know of no comprehensive survey of synchronic paradigmatic effects across different inflectional categories, similar tendencies can be observed in the degree of tolerance for morphologized stem changes across different categories. In a survey of 50 genetically unrelated languages, Bybee (1985) confirms that morphophonological stem changes are virtually never associated with person distinctions, and rarely with number distinctions. Tense and mood, on the other hand, are sometimes accompanied by stem changes, and aspect often is (see also Veselinova 2006 for discussion). These tendencies diagnose a paradigm structure in which certain inflectional features (aspect, and secondarily tense and mood) define rather cohesive paradigms, whereas person and number do not. We must interpret Bybee's findings cautiously, since they admit two possible explanations: perhaps sharing an aspect or tense feature is particularly important in conditioning paradigmatic cohesion, or perhaps differences in tense or aspect are preferentially marked with phonologically salient differences (Wurzel 1989). To the extent that tendencies in morphophonological stem changes are correlated with tendencies in phonological over- and underapplication, we find support for an interpretation in terms of paradigmatic cohesion.

Among nouns, it appears that the domain of phonological misapplication is more likely to be a number subparadigm (e.g. singular or plural) than a case subparadigm (e.g. all accusative forms). This appears to be mirrored by the fact that number suppletion is more common than case suppletion (though see Corbett 2007 for discussion). Similar tendencies are also found in cases of diachronic paradigm leveling.

Bybee (1985: 13) interprets these differences in terms of what she calls the *relevance* of different features to the stem. Relevance is defined as the degree to which a given inflectional category changes the meaning of the lexical root – for example, changes in valence, particularly between intransitive and transitive, tend to change the verbal action in a way that changes in person do not. Bybee argues that morphological features that are more relevant are more likely to be marked with overt morphological markers, and are more likely to be accompanied by stem changes. Restated in terms of paradigmatic identity, it appears that the less salient the meaning difference between two forms, the less likely speakers are to tolerate alternations.

Another tendency that can be observed in the examples above is that identity is more strongly enforced among some paradigms than others. For example, Spanish verbs have stress alternations within the present indicative (1sg [ter'mino] ~ 1pl [termi'namos]), but more "remote" tenses, such as the imperfect, lack alternations (1sg [termi'naba]; 1pl [termi'nabamos]). In Polish, raising alternations are observed within the case/number forms of simple nouns (8), but are suspended within the diminutive forms (9). This suggests that paradigmatic identity is not only enforced more strongly for some dimensions (e.g. person/number) than others (e.g. tense), but it is also enforced more strongly in some morphosyntactic contexts (e.g. the imperfect or the subjunctive) than others.

One attempt to derive these tendencies in a formal system is Burzio's (2005) Representational Entailments Hypothesis (see also CHAPTER 88: DERIVED ENVIRONMENT EFFECTS). Burzio proposes that the strength of the identity condition that holds between two related forms should depend on the degree to which they already have shared meaning, morphology, and phonology. According to this hypothesis, linguistic representations of different items are lined up, and the more material they share, the greater the expectation that they are alike in other respects as well. For example, suppose a given item has feature values [+F, +G, +H, +I]. This can be restated as a set of associations between co-occurring features: [+F] ⇒ [+G], [+F] ⇒ [+H], [+F] ⇒ [+I], and so on. Now suppose we are given another item with feature values [+F, +G, −H, −I]. This item shares the association of [+F] ⇒ [+G], but differs in its other associations – for example, [+F] ⇒ [+I] and [+G] ⇒ [+I] are not met in this form. Finally, compare a third item [+F, +G, +H, −I]. In this form, three entailments concerning [±I] are not met: [+F] ⇒ [+I], [+G] ⇒ [+I], [+H] ⇒ [+I]. In other words, the amount of overlap between [+F, +G, +H, +I] and [+F, +G, +H, −I] makes their difference in [±I] more salient or surprising. Burzio proposes that families of output–output faithfulness constraints are ranked to reflect such differences in overlap: OO-IDENT([±I]) (+F,+G,+H ⇒ +I) >> OO-IDENT([±I]) (+F,+G ⇒ +I). Entailments may be stated in terms of shared semantic or morphosyntactic features, or in terms of shared phonological features.

The Representational Entailments Hypothesis is useful in accounting for why some paradigms are more cohesive than others. For example, in Spanish, the imperfect is marked with an overt marker (*-ba-*) whereas the present tense has no overt tense marker. Therefore, imperfect forms share more properties in common (namely, the property of containing *-ba-*), which may in turn beget additional identity. The present subjunctive does not have an overt marker that makes subjunctive forms more similar to each other than indicative forms, but it is not unreasonable to suppose that the semantic or morphosyntactic representation of subjunctive involves more structure than the indicative does. If such explanations are on the right track, then it should be possible to correlate the degree of structural overlap between two forms and the pressure for paradigmatic identity between them. Crucially, it is possible to infer the structure of the representation of a given inflectional category through independent means (observing overt marking, finding implicational relations and default values, etc.; Bybee 1985; Harley and Ritter 2002; McGinnis 2005). Therefore, if the Representational Entailments Hypothesis is correct, it should be possible to predict the strength of paradigmatic identity effects.

5 Conclusion

Kenstowicz and Kisseberth (1977: 74) proclaimed that ". . . the notion 'paradigm' will have to be much more rigorously defined in order for the appeal to paradigm regularity to have much explanatory force." As this discussion has made clear, more than three decades later we are still in the early stages of understanding what determines the strength of paradigmatic effects, and what this might tell us about the underlying structure of correspondence among inflected forms. The growing literature on paradigm uniformity effects in the past 10 years has made progress on a number of issues, however. First, it has demonstrated that identity among inflectionally related forms is not just a diachronic phenomenon, but can be seen as a synchronic effect in the "wrong application" of productive phonological processes. Furthermore, formalized grammatical approaches to paradigm uniformity make testably different predictions about possible uniformity effects, pointing the way to those cases which deserve the closest empirical scrutiny. Finally, comparison of the cases discussed so far in the literature suggests a number of cross-linguistic trends that must be accounted for in a theory of how phonology refers to morphological structure. A deeper understanding of these tendencies will require a more comprehensive survey of synchronic paradigmatic effects, in order to understand how best to represent – and perhaps also derive – the observed tendencies concerning when paradigmatic identity is enforced.

ACKNOWLEDGMENTS

This chapter has benefited greatly from the comments, suggestions, and patient guidance of the editors and an anonymous reviewer. All remaining errors are, of course, my own.

REFERENCES

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.) *Analogy in grammar: Form and acquisition*, 54–81. Oxford: Oxford University Press.

Ahn, Sang-Cheol. 1998. *An introduction to Korean phonology*. Seoul: Hanshin.

Albright, Adam. 2002. The identification of bases in morphological paradigms. Ph.D. dissertation, University of California, Los Angeles.

Albright, Adam. 2005. The morphological basis of paradigm leveling. In Downing *et al.* (2005b), 17–43.

Albright, Adam. 2008a. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.) *Linguistics universals and language change*, 144–181. Oxford: Oxford University Press.

Albright, Adam. 2008b. Inflectional paradigms have bases too: Evidence from Yiddish. In Bachrach & Nevins (2008), 271–312.

Albright, Adam. 2010. Base-driven leveling in Yiddish verb paradigms. *Natural Language and Linguistic Theory* 28. 475–537.

Albright, Adam & Yoonjung Kang. 2009. Predicting innovative alternations in Korean verb paradigms. In *Current issues in unity and diversity of languages: Collection of the papers selected from CIL 18*, 893–913. Seoul: Linguistic Society of Korea.

Anderson, Stephen R. 1986. Disjunctive ordering in inflectional morphology. *Natural Language and Linguistic Theory* 4. 1–32.

Arad, Maya. 2003. Locality constraints on the interpretation of roots: The case of Hebrew denominal verbs. *Natural Language and Linguistic Theory* 21. 737–778.

Aske, Jon. 1990. Disembodied rules versus patterns in the lexicon: Testing the psychological reality of Spanish stress rules. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 16. 30–45.

Bachrach, Asaf & Andrew Nevins (eds.) 2008. *Inflectional identity*. Oxford: Oxford University Press.

Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2005. *The syntax–morphology interface: A study of syncretism*. Cambridge: Cambridge University Press.

Baldi, Philip. 1999. *The foundations of Latin*. Berlin & New York: Mouton de Gruyter.

Barr, Robin. 1994. A lexical model of morphological change. Ph.D. dissertation, Harvard University.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Ph.D. dissertation, University of Massachusetts, Amherst.

Blevins, James P. 2003. Stems and paradigms. *Language* 79. 737–767.

Bobaljik, Jonathan D. 2002. Syncretism without paradigms: Remarks on Williams 1981, 1994. *Yearbook of Morphology 2001*. 53–85.

Bobaljik, Jonathan D. 2008. Paradigms (Optimal and otherwise): A case for skepticism. In Bachrach & Nevins (2008), 29–54.

Buckley, Eugene. 2001. Polish o-raising and phonological explanation. Paper presented at the 75th Annual Meeting of the Linguistic Society of America, Washington, DC.

Burzio, Luigi. 1994. Metrical consistency. In Eric Sven Ristad (ed.) *Language computations*, 93–125. Providence, RI: American Mathematical Society.

Burzio, Luigi. 2005. Sources of paradigm uniformity. In Downing *et al.* (2005b), 65–106.

Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Cho, Seung-Bog. 1967. *A phonological study of Korean with a historical analysis*. Uppsala: Almqvist & Wiksell.

Choi, Myung-Ok. 1998. *Gukeoeumunlongwa jalyo* [*Korean phonology and data*]. Seoul: Taehaksa.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.) *Ken Hale: A life in language*, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Corbett, Greville G. 2007. Canonical typology, suppletion, and possible words. *Language* 83. 8–42.

Dammers, Ulf, Walter Hoffmann & Hans-Joachim Solms. 1988. *Grammatik des Frühneuhochdeutschen*, vol. 4: *Flexion der starken und schwachen Verben*. Heidelberg: Carl Winter Universitätsverlag.

Dekkers, Joost, Frank van der Leeuw & Jeroen van de Weijer (eds.) 2000. *Optimality Theory: Phonology, syntax, and acquisition*. Oxford: Oxford University Press.

Downing, Laura J., T. A. Hall & Renate Raffelsiefen. 2005a. Introduction: The role of paradigms in phonological theory. In Downing *et al.* (2005b), 1–16.

Downing, Laura J., T. A. Hall & Renate Raffelsiefen (eds.) 2005b. *Paradigms in phonological theory*. Oxford: Oxford University Press.

Eddington, David. 2000. Spanish stress assignment within the analogical modeling of language. *Language* 76. 92–109.

Giegerich, Heinz J. 1999. *Lexical strata in English: Morphological causes, phonological effects*. Cambridge: Cambridge University Press.

Hale, Mark, Madelyn Kissock & Charles Reiss. 1997. Output-output correspondence in Optimality Theory. *Proceedings of the West Coast Conference on Formal Linguistics* 16. 223–236.

Hall, T. A. & John H. G. Scott. 2007. Inflectional paradigms have a base: Evidence from s-dissimilation in Southern German dialects. *Morphology* 17. 151–178.

Han, Eunjoo. 2002. Optimal paradigms in Korean nominal inflections. *Studies in Phonetics, Phonology and Morphology* 8. 303–322.

Harley, Heidi. 2008. When is a syncretism more than a syncretism? In Daniel Harbour, David Adger & Susana Béjar (eds.) *Phi theory: Phi-features across interfaces and modules*, 251–294. Oxford: Oxford University Press.

Harley, Heidi & Elizabeth Ritter. 2002. Person and number in pronouns: A feature geometric analysis. *Language* 78. 482–526.

Harris, James W. 1973. On the ordering of certain phonological rules in Spanish. In Stephen R. Anderson & Paul Kiparsky (eds.) *A Festschrift for Morris Halle*, 59–76. New York: Holt, Rinehart & Winston.

Harris, James W. 1987. The accentual patterns of verb paradigms in Spanish. *Natural Language and Linguistic Theory* 5. 61–90.

Hartweg, Frédéric & Klaus-Peter Wegera. 1989. *Frühneuhochdeutsch: Eine Einführung in die deutsche Sprache des Spätmittelalters und der frühen Neuzeit*. Tübingen: Max Niemeyer Verlag.

Hayes, Bruce. 1999. Phonological restructuring in Yidɪɲ and its theoretical consequences. In Ben Hermans & Marc van Oostendorp (eds.) *The derivational residue in phonological Optimality Theory*, 175–205. Amsterdam & Philadelphia: John Benjamins.

Hayes, Bruce. 2000. Gradient well-formedness in Optimality Theory. In Dekkers *et al*. (2000), 88-120.

Hermans, Ben & Marc van Oostendorp (eds.) 1999. *The derivational residue in phonological Optimality Theory*. Amsterdam & Philadelphia: John Benjamins.

Hochberg, Judith G. 1988. Learning Spanish stress: Developmental and theoretical perspectives. *Language* 64. 683–706.

Hock, Hans Henrich. 1991. *Principles of historical linguistics*, 2nd edn. Berlin & New York: Mouton de Gruyter.

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10. 210–231.

Jacobs, Neil G. 2005. *Yiddish: A linguistic introduction*. Cambridge: Cambridge University Press.

Joesten, Maria. 1931. *Untersuchungen zu ahd. (as.)* ë, i, *vor* u *der Folgesilbe und zur 1. Pers. Sg. Präs. Ind. der starken eVerben (Kl. IIIb, IV, V)*. Giessen: Wilhem Schmitz Verlag. Reprinted 1968, Amsterdam: Swets & Zeitlinger N.V.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kager, René. 2000. Stem stress and peak correspondence in Dutch. In Dekkers *et al*. (2000), 121–150.

Kang, Yoonjung. 2003. Sound changes affecting noun-final coronal obstruents in Korean. In William McClure (ed.) *Japanese/Korean Linguistics* 12, 128–139. Stanford: CSLI.

Kang, Yoonjung. 2005. The emergence of the unmarked in an analogical change. Paper presented at the Seoul Forum in Phonology, December, 2005, Seoul National University. Available (August 2010) at http://individual.utoronto.ca/yjkang/files/SeoulForum2005.pdf.

Kang, Yoonjung. 2006. Neutralization and variations in Korean verbal paradigms. *Harvard Studies in Korean Linguistics* 11. 183–196.

Kenstowicz, Michael. 1996. Base identity and uniform exponence: Alternatives to cyclicity. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, 365–394. Salford: ESRI.

Kenstowicz, Michael. 2005. Paradigmatic uniformity and contrast. In Downing *et al*. (2005b), 145–169.

Kenstowicz, Michael & Charles W. Kisseberth. 1977. *Topics in phonological theory*. New York: Academic Press.

Kim, Hyunsoon. 1999. The place of articulation of Korean affricates revisited. *Journal of East Asian Linguistics* 8. 313–347.

Kim, Jinhyung. 2005. A reconsideration of phonological leveling: A case of noun inflection in Korean. *Studies in Phonetics, Phonology and Morphology* 11. 259–274.

Kim-Renaud, Young-Key. 1974. Korean consonantal phonology. Ph.D. dissertation, University of Hawaii. Seoul: Hanshin.

Kiparsky, Paul. 1972. Explanation in phonology. In Stanley Peters (ed.) *Goals in linguistic theory*, 189–227. Englewood Cliffs, NJ: PrenticeHall.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.

Kiparsky, Paul. 1998. Covert generalization. In Geert Booij, Angela Ralli & Sergio Scalise (eds.) *Proceedings of the 1st Mediterranean Conference of Morphology*, 65–76. Patras: University of Patras.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–365.

Kraska-Szlenk, Iwona. 1995. The phonology of stress in Polish. Ph.D. dissertation, University of Illinois.

Kurylowicz, Jerzy. 1947. The nature of the so-called analogical processes. *Diachronica* 12. 113–145.

Maiden, Martin. 2005. Morphological autonomy and diachrony. *Yearbook of Morphology* 2004. 137–175.

Mańczak, Witold. 1958. Tendences générales des changements analogiques. *Lingua* 7. 298–325, 387–420.

Marantz, Alec. 2007. Phases and words. In Sook-Hee Choe (ed.) *Phases in the theory of grammar*, 191–222. Seoul: Dong In.

Marvin, Tatjana. 2002. Topics in stress and the syntax of words. Ph.D. dissertation, MIT.

Matthews, Peter H. 1965. Some concepts in Word and Paradigm Morphology. *Foundations of Language* 1. 268–289.

McCarthy, John J. 1998. Morpheme structure constraints and paradigm occultation. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 32(2). 123–150.

McCarthy, John J. 2005. Optimal paradigms. In Downing *et al.* (2005b), 170–210.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.

McGinnis, Martha. 2005. On markedness asymmetries in person and number. *Language* 81. 699–718.

Oltra-Massuet, Isabel & Karlos Arregi. 2005. Stress-by-structure in Spanish. *Linguistic Inquiry* 36. 43–84.

Paul, Hermann, Peter Wiehl & Siegfried Grosse. 1989. *Mittelhochdeutsche Grammatik*, 23rd edn. Tübingen: Niemeyer.

Piggott, Glyne L. & Heather Newell. 2006. Syllabification and the spell-out of phases in Ojibwa words. *McGill Working Papers in Linguistics* 20. 39–64.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Reyes, Rogelio. 1974. Studies in Chicano Spanish. Ph.D. dissertation, Harvard University.

Sanders, Nathan. 2003. Opacity and sound change in the Polish lexicon. Ph.D. dissertation, University of California, Santa Cruz.

Skinner, Tobin R. 2008. The consequences of cyclic spell-out for verbal stem-allomorphy. In Susie Jones (ed.) *Proceedings of the 2008 Annual Conference of the Canadian Linguistic Association*. Available (August 2010) at http://homes.chass.utoronto.ca/~cla-acl/actes2008/CLA2008_Skinner.pdf.

Sohn, Ho-Min. 1999. *The Korean language*. Cambridge: Cambridge University Press.

Steriade, Donca. 2000. Paradigm uniformity and the phonetics–phonology boundary. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, 313–334. Cambridge: Cambridge University Press.

Stump, Gregory T. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.

Sturgeon, Anne. 2003. Paradigm uniformity: Evidence for inflectional bases. *Proceedings of the West Coast Conference on Formal Linguistics* 22. 464–476.

Veselinova, Ljuba N. 2006. *Suppletion in verb paradigms: Bits and pieces of the puzzle*. Amsterdam & Philadelphia: John Benjamins.

Vincent, Nigel. 1974. Analogy reconsidered. In John M. Anderson & Charles Jones (eds.) *Historical linguistics II: Theory and description in phonology*, 426–445. Amsterdam: North Holland Publishing Company.

Williams, Edwin S. 1994. Remarks on lexical knowledge. *Lingua* 92. 7–34.

Wurzel, Wolfgang Ullrich. 1989. *Inflectional morphology and naturalness*. Dordrecht: Kluwer.

Zwicky, Arnold M. 1985. How to describe inflection. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 11. 372–386.

84 Clitics

Stephen R. Anderson

1 Introduction

The notion of "clitic" derives from one of the oldest problems in the study of language: how to define the "word." Grammarians have long noted that a difficulty is posed in this area by the fact that certain elements in many languages seem to play an independent role in the grammatical structure of sentences, and thus to warrant the status of "grammatical words," but in terms of their sound structure form parts of unitary "words" (in a distinct, phonological sense) with other "grammatical words." Examples such as those in (1) from Homeric Greek are typical of the phenomenon.[1]

(1) a. hē de kai autōs **m'** aiei
she-N PTC even so me-A always
[PP ***en=*** *athanatoisi theoisi*] *neikei*
among immortal-D gods-D upbraids
even so she always upbraids me among the immortal gods
(*Iliad* 1.520, *apud* Taylor 1996: 480)

b. *theios* ***=moi*** *enupnion ēlthen Oneiros*
divine me-D dream came Oneiros
divine Oneiros came to me in a dream
(*Iliad* 2.56, *apud* Taylor 1990: 35)

In the first of these sentences (*proclitic*) **en=** is grammatically an independent preposition, but forms a word together with the following *athanatoisi* 'immortal'. In the second (*enclitic*) **=moi** is a pronominal adjunct 'to me' of the verb *ēlthen* 'came' but forms a word together with the preceding adjective *theios* 'divine'. In both cases the independent status of the pro- or enclitic seems assured by the grammar of the sentence, but the unitary status of its combination with a host is confirmed by its phonological (especially accentual) behavior. It is this conflict between two

[1] Clitics are identified in boldface type, with the symbol "=" indicating the direction of their relation to a host.

equally well-grounded notions of "word" that brought clitics to the attention of traditional grammarians, and subsequently that of linguists.

The problem as just presented is essentially a phonological one (how to get the phonology to treat two or more elements that appear distinct from the point of view of grammatical structure as one unit). The study of clitics was quickly complicated, however, by the suggestion that the same elements that displayed this anomalous phonological behavior also had specific, idiosyncratic syntactic properties. Jakob Wackernagel (1892) proposed, following Delbrück (1878), that the unstressed clitics of the oldest Indo-European languages (and thus, Proto-Indo-European) occurred systematically after the first word of the sentence, regardless of their grammatical function. This notion of a special syntax for clitics later became part of the very definition of "clitic" for some linguists, and much of the literature presumes that designating something as a clitic entails special behavior both in the phonology and in the syntax.

It is nonetheless useful to disentangle two distinct dimensions of "clitic" behavior, the phonological and the morphosyntactic, which turn out to be logically (and empirically) orthogonal (see Anderson 2005 for elaboration of this point, as well as related discussion in surveys such as that of Halpern 1998, the papers in Dixon and Aikhenvald 2002, and much other literature both traditional and modern). In the context of the present book, this chapter will focus almost exclusively on the phonological aspects of clitic behavior, and references to "clitics" will be to elements that display the relevant phonological properties (without regard to whether they display unusual syntactic distribution).

2 What is a (phonological) clitic?

As a starting point, we can ask which elements we ought to consider as clitics from such a perspective. The notion of clitic in traditional grammar is that of a "little" word, and in particular one that does not bear an independent accent but rather leans accentually on an adjacent word.[2] The proposal that clitics are always unaccented, however, is problematic.

For instance, in Modern Greek, enclitics do not usually receive stress; thus, [ˈðose] 'give!', [ˈðose=**mu**] 'give me!' with no stress on the clitic =**mu**. But when two such enclitics are attached to the same host, a stress appears on the penultimate one, as in [ˌðose=ˈ**mu=to**] 'give it to me!'.[3] This is a consequence of a general rule of Modern Greek that builds a trochaic foot over two otherwise unstressed syllables at the right edge of a word, provided the result does not involve a stress clash. Thus, when a clitic is added to antepenultimate-stressed [triaˈðafilo] 'rose', the result is [triaˌðafiˈlo=**mu**] 'my rose'. It is not the sequence of clitics *per se* that results in the penultimate stress in [ˌðose=ˈ**mu=to**], but rather the application of this rule: cf. [ˈpes=**mu=to**] 'say it to me!', with no stress on =**mu** in an otherwise identical sequence, because such a stress would clash with that on the monosyllabic stem. On the traditional understanding, the claim that =**mu** '1SG' is a clitic seems to be

[2] The word *clitic* derives from Greek *klitikos* 'leaning', from *klīnein* 'to lean'.

[3] Arvaniti (1992) provides experimental evidence that, contrary to the proposals of some previous authors, the added stress in such cases is primary with the original word stress being reduced to secondary. This result has been confirmed by the judgments of several native speakers.

compromised by the fact that it sometimes bears an accent, but we can see that this accent is due to the regular phonology of the language, and not to properties of =**mu** '1SG' itself.

Similarly, in the Papuan language Bilua (Obata 2003), stress is generally initial, but proclitic pronouns do not bear stress, as in (2a). Proclitics can also appear in constructions in which there is no adjacent stressed element, however. This occurs under two conditions: first, a vowel-final clitic does not form part of a word with a following vowel-initial stem, as in (2b); and, second, under some circumstances a cluster of clitics arises which is not associated with any non-clitic host, as in (2c). In these circumstances, the clitic receives stress if initial, as illustrated below.

(2) a. [o= ˈβouβaɛ =k =a]
3SG.MASC kill 3SG.FEM.OBJ TNS
'He killed it.'

b. ˈo ˈodiɛ =k =a]
3SG.MASC call 3SG.FEM.OBJ TNS
'He called her.'

c. ˈo =k =a ˈzari=a ˈrae=ng=o
3SG.MASC 3SG.FEM PRT want-TNS marry-2SG.OBJ-NOM
'He wants to marry you.'

Designating as "clit ics" exactly those elements that do not bear stress, then, does not appear to give us the results we desire. Instead, it is proposed in Anderson (2005) that the right way to pick out clitics phonologically is as *prosodically deficient* elements. Let us assume that full words in general have a lexical representation that organizes their phonological content into syllables, feet, and ultimately one or more *Phonological Words* (PWords; see CHAPTER 51: THE PHONOLOGICAL WORD). We can then say that a phonological form realizing some grammatical element, whose segmental content may be organized into syllables and possibly feet but which is not lexically assigned the status of a PWord, is a clitic in the desired phonological sense. This characterization is not compromised by the fact that such a clitic will typically become part of a PWord (perhaps together with other clitics, as in Modern Greek [ˌðose=ˈ**mu=t●**] or in the Bilua sentence (2c) above) as a consequence of the principles of prosodic organization of the language in question.

The property of being a clitic in this sense, then, is not necessarily a characteristic of a lexical item, but rather of a phonological form that can realize that lexical item. The same item may well have both clitic and non-clitic forms. The classic example of this is the case of the auxiliary verbs in English: many of these have both full, non-clitic forms (*is, has, had, would, will*, etc.) and clitic forms (*'s, 'd, 'll*, etc.). From the point of view of the grammar, these are essentially free variants. If a reduced (clitic) form is chosen to lexicalize the auxiliary in a given sentence, however, this may result in prosodic ill-formedness, as a consequence of the impossibility of incorporating the prosodically deficient item into the overall sound structure of the sentence in a well-formed way (see Anderson 2008 for discussion and analysis). Apart from these differential phonological effects, however, the reduced and unreduced auxiliaries are instantiations of the same grammatical element.

In order to be pronounced, such prosodically deficient material must be incorporated into the larger prosodic structure in some way: thus, the penultimate stress in Modern Greek [ˌðose='mu=to] results from incorporating both enclitics into the same phonological word as the host verb, and then building a trochaic foot over the resultant sequence of unstressed word-final syllables. The Bilua examples result from assigning PWord status to clusters of material that cannot be incorporated into any independent PWord, and then assigning initial stress to this word. It is the characterization of this sort of integration (which we will refer to as *Stray Adjunction*) that constitutes the phonology of cliticization, and this area of phonology will be central to the discussion in later sections of the present chapter.

3 How do clitics differ from affixes?

Although the characterization of clitics as prosodically deficient grammatical elements appears to capture the phonological dimension of their behavior, it does not pick them out uniquely in grammatical structure. With relatively few exceptions, the affixes found within words as formal markers of derivational and inflectional structure also lack an autonomous organization into prosodic constituents at or above the level of the PWord (see CHAPTER 104: ROOT–AFFIX ASYMMETRIES), and the question naturally arises of how clitics and affixes are to be distinguished.

The classic characterization of the issues involved is provided by the widely cited work of Zwicky and Pullum (1983), who enumerate a number of differences between clitics and affixes in defense of their analysis of English *-n't* as the realization of an inflectional category of modals and other auxiliary verbs rather than as a clitic. These include the points in (3).

(3) a. Clitics have a low degree of selection with respect to their hosts; affixes a high degree of selection.
b. Affixed words are more likely to have accidental or paradigmatic gaps than host + clitic combinations.
c. Affixed words are more likely to have idiosyncratic shapes than host + clitic combinations.
d. Affixed words are more likely to have idiosyncratic semantics than host + clitic combinations.
e. Syntactic rules can affect affixed words, but not groups of host + clitic.
f. Clitics, but not affixes, can be attached to material already containing clitics.

These points can be illustrated, following Zwicky and Pullum (1983), by the contrasts in (4) between English clitic auxiliaries (e.g. *'s* 'is, has', *'d* 'would') and the element they argue is an inflectional affix, *n't* 'NEG'.

(4) a. The clitic auxiliaries can attach to words of any class that happen to fall at the right edge of the preceding constituent; *n't* can only be added to finite forms of auxiliary and modal verbs.
b. Combinations of clitic auxiliaries with preceding material are limited only by the possibilities of the syntax; some combinations of auxiliary + *n't* do not exist (e.g. **mayn't*; in most dialects also **amn't*) while one (*ain't*) does not correspond to a specific non-negative form.

c. Combinations of host + clitic auxiliary are governed by the regular phonology of English, as seen for instance in regular plurals and past tense forms with the endings /z/ and /d/; forms such as *don't*, *won't*, *can't*, and *shan't* bear idiosyncratic relations to their non-negative counterparts.
d. Clitic auxiliaries make the same syntactic and semantic contribution to a sentence as full forms; auxiliaries in *n't* can have idiosyncratic semantics (thus, in *you mustn't go* the negation is within the scope of the modal, while in *you can't go* the modal is in the scope of negation).
e. Clitic auxiliaries do not move together with their host (thus, a question corresponding to *I think John's at the door* is *Who do you think's at the door?* and not **Who's do you think at the door?*) while the negated auxiliaries move as a unit (the question corresponding to *I haven't any more bananas* is *Haven't you any more bananas?* and not **Have youn't any more bananas?*).
f. While clitics can be added to other clitics (*I'd've done better if I could've*), *n't* cannot (thus, *I wouldn't do that if I were you* cannot be expressed as **I'dn't do that if I were you*).

Zwicky and Pullum present these differences as descriptive observations, supported by comparisons between uncontroversial instances of clitics and affixes. They can be argued to follow, however, from the proposal that clitics are introduced into syntactic structure as prosodically deficient, but morphosyntactically independent, elements, while affixed words are formed by lexical operations and appear as units in the syntax.[4] Although Zwicky and Pullum formulate some of the principles in (3) only as tendencies, the present account of the nature of clitics suggests that nearly all of them should be construed categorically.

The exception to this generally absolute nature of the differences between clitics and affixes is (3d): syntactically compositional idioms can be semantically idiosyncratic (e.g. *build castles in the air* 'make unrealistic plans or proposals'), and there is no reason to exclude host + clitic combinations from this same possibility. Indeed, many languages assign special meanings to verbs in the presence of particular clitics, such as French *il y a* 'there is' or the Italian *verbi procomplementari* studied by Russi (2008). These latter are combinations of a verb with a specific clitic or cluster of clitics, which take on a conventionalized meaning that is not compositionally related to that of the basic verb. Examples are provided in (5).

(5) a. far=la 'deceive, prevail on someone cunningly'
from *fare* 'make, do' + *la* '3sg DO'
b. voler=ne 'resent, have hard feelings for someone'
from *volere* 'want' + *ne* 'PARTITIVE'

[4] The specific framework I presume is roughly that of Anderson (1992). Within that theory, productive inflection results from the operation of Word Formation Rules that take a lexical stem and the morphosyntactic representation of a syntactic position as their input and yield inflected words as their output; while derivation and lexically idiosyncratic inflection result from Word Formation Rules that construct and relate lexical stems. The details of this position are not essential: what matters is the claim that fully inflected words, structured as PWords, appear in the prosodic structure projected from the syntax. Clitics appear in this structure either as prosodically deficient lexical items (e.g. the contracted forms of English auxiliaries: see Anderson 2008) or as "special clitics" introduced (as phrasal morphology) into that structure at a point where non-clitic material is already present, as described in Anderson 2005.

c. prender=se=la 'take offense, be upset'
from *prendere* 'take' + *si* 'REFL' + *la* '3SG DO'

The constructions in which such combinations are found are syntactically normal (e.g. *Me l'ha fatta di nuovo* 'he tricked me again'), but their interpretation cannot be directly derived from those of their parts.

The rest of the properties in (3) follow from the proposed architecture of grammar. Clitics *per se* are not selective with respect to their hosts (3a), because they are placed by principles that do not make direct reference to the host (although in the case of "special clitics," the phrasal environment for their introduction may be such that only a restricted range of hosts will be present in the appropriate position). They do not display gaps (3b), because individual host + clitic combinations are not listed as such in the lexicon and so are not subject to omission. Similarly, such combinations are not available for lexical listing of idiosyncratic form (3c). Host+ clitic combinations are not affected in a unitary way by the syntax (3e), because the fact of being a clitic entails only a phonological, not a syntactic, relation to the host.

The only way a clitic could appear "inside of" an affix (3f) would be if some special circumstances caused it to be introduced in that way as an "endoclitic." Most of the putative instances of this situation that have been adduced, such as the pronominal clitics that appear between the verbal stem and a future or conditional ending in Portuguese (e.g. *mostrár-no-los-á* 's/he will show them to us'), appear to have alternative analyses that do not involve "endocliticization" (see Anderson 2005: 152ff.).

One exception is the case of Udi as discussed by Harris (2002), which does appear to be a real example. In an Udi form like that in (6), the clitic =**ne** '3SG' comes between one affix and another.

(6) nana-n äjel-ax ak'-es=**ne**-d-e k'uʧan
mother-ERG child-DAT see-INF-3SG-CAUS-AORII puppy.ABS
'The mother showed a puppy to the child.'

In a form such as *a=**z**-q'-e* 'I received', indeed, the clitic element =**z** '1SG' appears within the monomorphemic root *aq'* 'receive'. The analysis of such cases is extremely interesting, but, as argued in Anderson (2005: 161–165), the principles involved are still consistent with the claim that clitics are added to affixed words, and not the reverse.

4 How are clitics prosodically related to their hosts?

Let us assume, then, that lexical elements appear in the input to the phonology with a certain amount of prosodic organization, and that non-clitics differ from clitics in that only the former are lexically organized into PWords. Clitics and non-clitics alike must be organized into *Phonological Phrases* (PPhrases) and perhaps higher levels of prosodic structure, although that is of less importance for present concerns. This phrasing can be regarded as being projected at least in part from syntactic structure, but the question remains of how prosodically deficient material is related to adjacent PWords within this overall organization.

The categories of prosodic structure are generally assumed to be related in a hierarchical fashion, with syllables constituting Feet, which are parts of PWords, which are in turn grouped into PPhrases, etc.

(7) *The Prosodic Hierarchy*
σ < Foot < PWord < PPhrase < Intonational Phrase . . .

A particularly restrictive view of this hierarchy known as the Strict Layering Hypothesis was defended by Nespor and Vogel (1986), for whom the relation between category types was seen as exhaustive at all levels: that is, PPhrases consisted exclusively of PWords, which in turn consisted exclusively of Feet, etc.[5] In a paper which is fundamental to the study of clitic phonology, however, Selkirk (1995), following arguments of Inkelas (1989), proposed that the principles of the Prosodic Hierarchy ought to be regarded as a set of individually ranked, violable constraints, and this view has dominated subsequent research (see CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE, CHAPTER 40: THE FOOT, CHAPTER 51: THE PHONOLOGICAL WORD, and CHAPTER 57: QUANTITY-SENSITIVITY for more discussion of the prosodic hierarchy).

Associating positions on the ordering in (7) with consecutive integers, we could express the basic nature of the Prosodic Hierarchy as involving two fundamental requirements.

(8) a. *Layeredness*
No C_i dominates a C_j, where $j > i$ (e.g. no Foot contains a PWord).

b. *Headedness* (first approximation)
Every C_i directly dominates some C_{i-1} (e.g. every PWord contains a Foot).

The Strict Layering Hypothesis can be expressed as the claim that representations also meet two other requirements.

(9) a. *Exhaustivity*
No C_i directly dominates a C_j, where $j < i-1$ (e.g. no PWord directly dominates a σ).

b. *Non-Recursivity*
No C_i directly dominates another C_i (e.g. no PWord contains another PWord; adjunction structures do not exist).

In order to maintain its logical independence from Non-Recursivity, the formulation of Headedness in (8) can be replaced by the following.

(10) *Headedness*
Every C_i directly dominates some C_j, where $j \geq i-1$.

[5] Nespor and Vogel also posited a category of *Clitic Group* between the PPhrase and the PWord. Subsequent work, such as Booij (1988) and Zec and Inkelas (1991), has generally concluded that no such distinct prosodic category need be introduced, and it is disregarded here. For some discussion, see Anderson (2005: 42ff.).

As noted already by Selkirk (1995), Layeredness and Headedness are inherent in the nature of the Prosodic Hierarchy; since these notions are in some sense definitional, they are not violable, and if construed as constraints should be treated as always undominated. Another undominated requirement, which we could call that of *Full Interpretation*, mandates that all phonological material to be pronounced be integrated into the overall prosodic structure, which means in effect that there must be a path from it to the root of the prosodic tree. It is this constraint that enforces the application of some process of Stray Adjunction in the case of material which is otherwise prosodically unaffiliated.

The requirements in (9), however, make substantive claims about the range of prosodic structures found in the languages of the world and, as such, are subject to empirical confirmation. Evidence suggests, in fact, that they are violated in some instances, and this is the basis for interpreting them not as definitional of prosodic structure, but as potentially violable constraints.

Of these the conditions in (11), formulated now as constraints, are apparently never violated and so can be regarded as undominated along with Full Interpretation.

(11) a. Layeredness
No category dominates a higher-level category.

b. Headedness
Every category directly dominates (at least) one element no more than one level below it on the hierarchy.

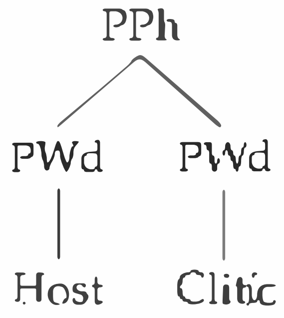
The additional conditions of the Strict Layering condition can, as we have seen, be violated. Furthermore, violation may be "local" in the sense that a language violating, say, Exhaustivity at the PPhrase level may nonetheless conform to this constraint at other levels, such as the PWord. The relevant principles thus need to be formulated as families of constraints, varying over the categories of the hierarchy as in (12).

(12) a. Exhaustivity(C_i)
Every element of category C_i is exhaustively composed of elements of category C_{i-1}.

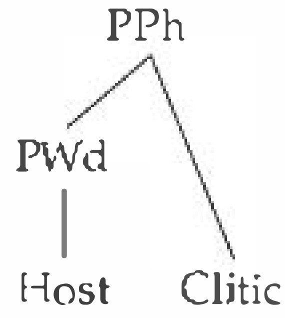
b. Non-Recursivity(C_i)
No element of category C_i directly dominates another instance of C_i.

Adherence to the Strict Layering Condition led Nespor and Vogel to require that clitics always constitute PWords in their own right, sisters of their host within a constituent of the next highest level of the hierarchy. This is somewhat problematic, given that clitics do not generally manifest the properties of independent PWords, such as autonomous stress. If we construe the conditions characterizing the Prosodic Hierarchy in (12) as constraints that can be violated under the pressure of other constraints, however, there are a variety of possible relations that might obtain between a clitic and its host, and Selkirk (1995) justifies the claim that all of these are in fact instantiated. The typology of clitic–host relations that she proposes is as in (13).

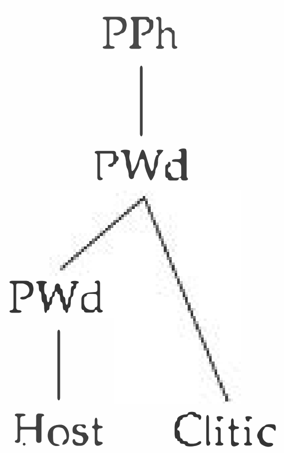
(13) a. *PWord Clitic*[6]



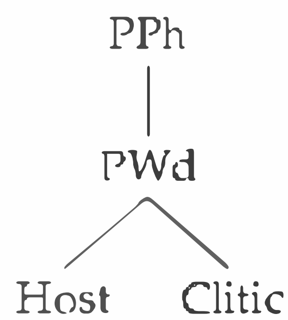
b. *Free Clitic*



c. *Affixal Clitic*



d. *Internal Clitic*



PWord clitics, of course, are structures that result when all of the constraints in (12) are satisfied, so that Strict Layering obtains. Free clitics, in contrast, result when some other constraint forces violations of Exhaustivity(PPhrase): the PPhrase thus contains a constituent lower in the hierarchy than a PWord, such as a stray syllable or foot. Affixal clitics result when Exhaustivity(PPhrase) is satisfied, but Non-recursivity(PWord) is not (and Exhaustivity(PWord) is also violated, in case the stray material constituting the clitic is a syllable and not a foot).

Internal clitics, like PWord clitics, involve no violations of any of the constraints. Differentiating these two possibilities requires us to invoke another constraint:

(14) Prosodic Faithfulness
Prosodic structure in the input should be preserved in the output.

If we assume that the prosodic structure of the host up to the level of the PWord is present in the input to that part of the phonology enforcing Stray Adjunction, we can see that this structure is preserved intact if the stray material is incorporated as a PWord clitic, but altered if it is incorporated as an internal clitic. The choice between the two, then, depends on the relative importance of Prosodic Faithfulness and some constraint disfavoring the creation of additional PWord structure (say, *Struct).

For an example of PWord clitics, we can appeal to Bilua examples such as (2b) and (2c), where other aspects of the structure prevent the incorporation of the clitic into an adjacent PWord, but the prohibition against building new PWords is not highly enough ranked to prevent a new PWord from being formed. The remaining possibilities can be demonstrated through a set of closely related systems

[6] Since clitics have been defined precisely as elements lacking PWord structure, the notion of a "PWord Clitic" may seem paradoxical. The point is that while clitics do not have such structure underlyingly, the subsequent operation of the language's broader principles of prosodic organization may give rise to such a structure, as we saw in the case of Bilua above.

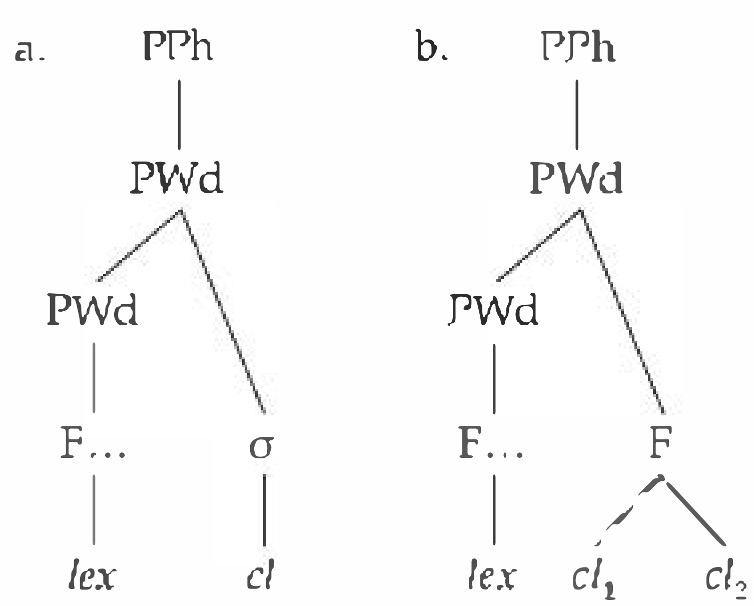
analyzed elegantly by Peperkamp (1997). As reviewed below, she argues that Standard Italian post-verbal pronominal clitics have the structure of free clitics, while the corresponding elements in Neapolitan are affixal clitics and those of Lucanian are internal clitics. The three dialects provide a nice contrasting set, differing minimally in the way clitics are incorporated into prosodic structure as described in terms of varied rankings of the constraints introduced above.

The first system to be considered is that of Neapolitan, as illustrated in (15).

(15) *Neapolitan*

imperative	*imperative* + 'it'	*imperative* + 'you' + 'it'	
'fa	'fallə	fat'tillə	'do'
'conta	'contalə	'conta'tillə	'tell'
'pettina	'pettinalə	'pettina'tillə	'comb'

I assume that PWords are built lexically over the host verbs, and then prosodically deficient clitics are added post-lexically. Note that when clitics are added, the first stress does not change except in one case ([fat'tillə]), where we can say that the new stress appearing on the clitic sequence has the effect of suppressing the original stem stress to avoid violating *Clash (which penalizes a sequence of two adjacent stresses). Peperkamp shows that we can describe this system by saying that the clitic material is adjoined to the existing prosodic word, without modifying its structure, as in (16).

(16)


A single clitic constitutes a single syllable, and not a Foot; two clitics, however, provide enough material to constitute a Foot, and thus introduce an additional stress. Peperkamp's discussion suggests that there are aspects of formal suppletion that require the treatment of the two-clitic sequence as a single unit, which is eligible to be a Foot. Alternatively, we could assume simply that the two monosyllabic units are introduced together, and subsequently organized into a Foot.

We can describe this system as follows. Full Interpretation, Headedness, and Layeredness are all undominated well-formedness conditions on the candidates that are to be compared, so they play no part in the ranking. It is also the case that prosodic structure assigned lexically is generally preserved, so Prosodic Faithfulness (14) is also ranked high.

In the case of a monosyllabic stem followed by two clitics, however, the need to avoid successive stressed syllables is more important than the preservation of input prosody, so the stress on the stem is lost as a result of the domination of Prosodic Faithfulness by another constraint (17).[7]

(17) *Clash
Sequences of two consecutive stressed syllables are disallowed.

To satisfy Full Interpretation, prosodically deficient material (i.e. the clitics) must be incorporated into the structure somewhere, and the choices are limited. Incorporation into a foot would violate well-formedness conditions on feet, as well as faithfulness to existing prosodic structure. Incorporation into the existing PWord would also violate faithfulness. Incorporation at the PPhrase level would violate Exhaustivity(PPhrase). The Affixal clitic structures that are actually found indicate that Exhaustivity(PPhrase) outranks Non-Recursivity(PWord): that is, building a recursive PWord preserves the existing prosodic structure, and avoids having lower-level constituents (syllables, feet) directly dominated by a PPhrase. The overall constraint ranking for Neapolitan is as in (18).

(18) *Clash >> Prosodic Faithfulness >> Exhaustivity(PPhrase) >> Non-Recursivity(PWord)

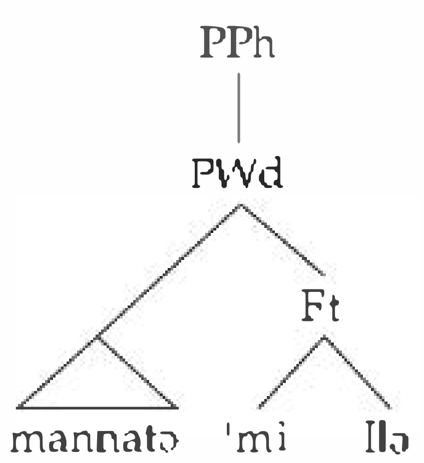
Now compare the Neapolitan approach to Stray Adjunction with that employed in another dialect, Lucanian.

(19) *Lucanian*

a.	'vinnə	'sell'	vən'nillə	'sell it'
b.	ram'millə	'give me it'	mannatə'millə	'send me it'

We see in (19a) that the addition of a clitic in this language causes stress to shift rightward.[8] Apparently, a binary trochaic foot is constructed over the last two syllables of the form, including both stem and any following clitics. The forms in (19b), with two clitics, have this foot constructed entirely over clitic material. In this language, Stray Adjunction produces Internal clitics, sacrificing Faithfulness to maintain the Strict Layering constraints. The resulting structure for a form with two clitics is as in (20).

(20)



[7] The fact that it is the first, rather than the second, of two adjacent stresses that is lost must be resolved by other aspects of the prosodic phonology of Neapolitan not considered here.

[8] Stress shift is responsible for the vowel alternation in these forms, with stressed [i] corresponding to unstressed [ə].

The constraint ranking necessary to obtain this result is (21).

(21) Non-Recursivity(PWord), Exhaustivity(PWord) >>
Prosodic Faithfulness

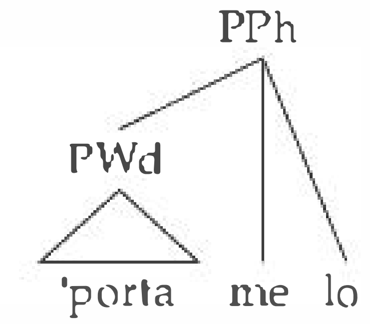
Let us finally compare the situation in (standard) Italian, illustrated in (22).

(22) *Standard Italian*

a.	'porta	'bring'	'portami	'bring me'
b.	'portamelo	'bring me it'	te'lefonamelo	'telephone it to me'

Here the addition of a clitic does not alter the lexically assigned stress, suggesting that Prosodic Faithfulness is highly ranked. Even when two clitics are added, as in (22b), the stress is not altered, and apparently no new stress is assigned even though two syllables of additional material would support the construction of a new Foot if this material were within the PWord. Apparently, then, Stray Adjunction in Standard Italian produces free clitics by attachment to the PPhrase, as in (23).

(23)



The required ranking is that in (24).

(24) Non-Recursivity, Exhaustivity(PWord), Prosodic Faithfulness >>
Exhaustivity(PPhrase)

Stray Adjunction in these three Italian dialects is thus based on different rankings of the prosodic constraints, yielding three different structural types of clitic as a reflection of these differences in their post-lexical phonology.

Peperkamp argues for the structural differences amongst Italian dialects on the basis of the distribution of stress alone, but sometimes this is insufficient to provide an unambiguous analysis. For example, in the case of a language with stress oriented to the left of the word (or simply preserved by high-ranking faithfulness constraints) and a set of unstressed enclitics, stress alone will not allow us to differentiate among the structures of free, affixal, and internal clitics. To do so, we must establish the location of PWord boundaries in the resulting form. The three possibilities can be distinguished in this way, as in (25).

(25) a. Free clitic: (. . .)host]$_{PWd}$ clitic]$_{PPh}$
b. Affixal clitic: (. . .)host]$_{PWd}$ clitic]$_{PWd}$]$_{PPh}$
c. Internal clitic: (. . .)host clitic]$_{PWd}$]$_{PPh}$

Determining which of these structures is present in a given instance is certainly not trivial, but it can often be done by looking for phonological phenomena which occur at the edges of PWords or across PWord boundaries. Revithiadou (2008) provides a detailed study of a range of dialects of Modern Greek of exactly this sort, showing that phonological regularities characteristic of prosodic boundaries identify different host–clitic relationships in different dialects. Similar arguments provided by Booij and Rubach (1987) for Polish can be interpreted as showing that proclitic prepositions in that language (e.g. *bez* 'without' in *bez namysłu* 'without thinking') are related to a following host as affixal clitics, as reviewed in Anderson (2005: 40f.).

It appears that, in general, the attachment of a clitic to a host on one side or the other can be derived from the overall prosodic organization of a language. Typically, prosodic structure above the level of the PWord is projected from the syntax, and the commonest tendency is for this structure to be respected: that is, a clitic attaches phonologically to the host (on its right or on its left) with which it is most closely affiliated grammatically. In some instances, though, this direction of attachment is directly contravened.

In Kwakw'ala, for instance, as discussed at length in Anderson (2005), DP-initial determiner clitics associate phonologically not with the following word, which is part of the same DP, but rather with the preceding word, which is not. An example is provided by the sentence in (26).

(26) jəlkʷəmas[=**ida** bəgʷanəma]$_{DP}$[=**χ-a** 'watsi]$_{DP}$[=**s-a** gʷaχt͡ɬuχʷ]$_{DP}$
cause hurt-DEM man-OBJ-DEM dog-INST-DEM stick
'The man hurt the dog with the stick.'

Here the square brackets indicate syntactic constituents, while inter-word spaces delineate PWords: thus, /bəgʷanəma[=**χ-a**/ is a single PWord, while [=**ida** bəgʷanəma]$_{DP}$ is a single DP.

This situation can be related to the fact that Kwakw'ala is a language in which virtually all morphological marking is suffixal, and thus the lexical root is always (with the exception of reduplicated forms) word initial. A preference to maintain this same situation at the level of prosodic structure can be expressed as a constraint such as (27).

(27) ALIGN(PWord, L; LexWord, L) (>> ALIGN(XP, L; PPhrase, L))

That is, it is important that the left edge of a PWord coincide with the left edge of a lexical word (and not e.g. a clitic determiner). This constraint is more highly ranked than the requirement that the left edges of syntactic phrases coincide with the left edges of PPhrases, and forces the clitics to associate anti-syntactically to their left.

The claim that the direction of attachment of clitics can be derived from the prosodic organization of the language as a whole (including constraints such as the one in (27)) is a strong one. It is at variance with proposals such as that of Klavans (1985), where it is claimed that among the dimensions defining individual clitics in a language is a parameter of direction of attachment. Subsequent research has suggested, however, that once grammatical structure and its relation to prosody are taken into account, a unitary analysis can be offered for the way clitics attach in any individual language. Counterexamples to this claim would have to involve pairs of clitics that were entirely comparable in their grammar, but where one attached to a host on its left and the other to a host on its right (under otherwise identical prosodic conditions). Such examples do not appear to exist, and it seems reasonable to propose that the direction of attachment of clitics is a function of the overall grammar of a language, rather than a property of individual clitics.

In summary, clitics can be characterized from a phonological point of view as linguistic elements lacking in prosodic structure at (or below) the level of the PWord. Linguistic units that are called "clitics" on the basis of unusual syntactic behavior may or may not be clitics in this sense: for example, Italian *loro* '3PL DAT' behaves in a way which is partially similar to the other Italian pronominal clitics, but *loro* is not prosodically deficient, and thus does not constitute a clitic from the phonological point of view. Similarly, Hungarian verbal prefixes such as *oda* in *oda-ment-em* 'I went over there' constitute PWords in their own right (as shown by stress and vowel harmony), and thus are not phonological clitics, even though they bear a special grammatical relation to an associated verb.

Material that is not fully integrated into prosodic structure (at the PWord level) in the input can be called "stray," and the phonology of cliticization is fundamentally a matter of how this stray material is incorporated into the overall prosodic structure of the sentence: how "Stray Adjunction" is enforced. The basic mechanics of this can be described by an ordering of the constraints characterizing prosodic layering with respect to one another and to other constraints within the grammar of the language in question. Arguments for this ranking can be provided either directly from properties of the resulting prosodic structure (such as the location of stress) or from other phonological phenomena that are sensitive to it.

5 How is the segmental phonology of a clitic related to that of its host?

A consequence of the grammatical architecture proposed here concerns the phonology applicable to clitic + host combinations. Since the formation of these presupposes the forms of lexical words, it would appear that in terms of classical Lexical Phonology (e.g. Kiparsky 1985; see also CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME), any adjustments to their shape must follow from principles of the post-lexical phonology, not the lexical phonology *sensu stricto*. Bermúdez-Otero and Payne (forthcoming) note this, but assert that examples exist which controvert it: cases in which host + clitic combinations are affected by rules that are lexical, not post-lexical in character.

The one such example cited by Bermúdez-Otero and Payne concerns laryngeal neutralization in Catalan. They argue, following the descriptive literature (e.g. Wheeler 2005) that voicing is neutralized in coda obstruents in this language (see CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). When these are closely followed by an onset consonant (in the same or a following word), they show the same voicing as that consonant, and it is plausible to attribute this to assimilation. Word-finally, however, coda obstruents are devoiced; and this devoicing persists even if the consonant in question is resyllabified post-lexically with a following vowel. These facts are illustrated in (28) for the stem /ʎob/ 'wolf', which ends in underlying voiced /b/.

(28)	a.	*llop*	[ʎop]	'wolf'
	b.	*llop lliure*	[ʎob.ʎiw.re]	'free wolf'
	c.	*llop trist*	[ʎop.trist]	'sad wolf'
	d.	*lloba*	[ʎo.βə]	'she-wolf'
	e.	*llop amic*	[ʎo.pə.mik]	'friendly wolf'

They suggest that there must be a "word-level" principle of laryngeal neutralization which is counterbled by post-lexical resyllabification in forms like (28e). In forms where a stem-final voiced obstruent is followed by a vowel-initial clitic, however, the pattern is subtly different: resyllabification bleeds laryngeal neutralization, as illustrated in (29) for the stem /rɛb/ 'receive'.

(29) a. *rebre* [rɛ.βrə] 'receive (INF)'
b. *rep això!* [rɛ.pə.ʃɔ] 'receive (2SG.IMP) that!'
c. *rep=ho!* [rɛ.βu] 'receive (2SG.IMP)-3SG.ACC.N!'

Why should there be a difference in voicing between the stem-final /b/ as it appears in (29b) and in (c)? Bermúdez-Otero and Payne conclude that this must be because the clitic in (29c) must already be present (and have triggered resyllabification) at the point where Laryngeal Neutralization takes place:

> These data show unequivocally that enclitic =ho belongs in the same grammatical word as the verb stem, since it causes the stem-final consonant to be syllabified as an onset already at the word level [. . .] Therefore, enclitic =ho cannot be a phrasal affix.

This conclusion does not follow, however. It results from Bermúdez-Otero and Payne's equation of "word-level" phonology with "lexical" phonology, and the assumption that the "post-lexical" phonology is monolithic. In fact, however, we can take the "word-level" character of laryngeal neutralization to refer to the PWord, not (as Bermúdez-Otero and Payne do) to the grammatical word. If we assume that post-verbal pronominal clitics in Catalan are affixal clitics, the result of stray adjunction in (29c) will be [[[rɛb]$_{PWd}$u]$_{PWd}$]$_{PPh}$. This entire construction is a PWord, and it is plausible to assume that resyllabification of this PWord yields a structure like [[rɛ]$_{PWd}$.bu]$_{PWd}$, bleeding Laryngeal Neutralization. In (29b), however, the structure is [[rɛb]$_{PWd}$[ə.ʃɔ]$_{PWd}$]$_{PPh}$. Laryngeal Neutralization, a rule whose scope is the PWord, converts this to [[rɛp]$_{PWd}$[ə.ʃɔ]$_{PWd}$]$_{PPh}$, which is subsequently resyllabified at the PPhrase level to [[rɛ]$_{PWd}$[pə.ʃɔ]$_{PWd}$]$_{PPh}$.

Resyllabification at the PPhrase level does not bleed Laryngeal Neutralization, but Resyllabification at the PWord level does. Since Bermúdez-Otero and Payne do not show that Laryngeal Neutralization has other characteristics of a "lexical" rather than "post-lexical" process (e.g. lexical exceptions), it follows only that the post-lexical phonology displays a sort of cyclic structure, with a round of phonological adjustment induced by each of the categories of the Prosodic Hierarchy, and not that clitics like Catalan =ho are not phrasal affixes. The notion that phonological regularities enforced at different levels of the Prosodic Hierarchy (such as the PWord *vs.* the PPhrase) can be at least partially distinct is a cornerstone of prosodic theory, and a basic way in which one argues that a given prosodic constituent is of one type rather than another (see Nespor and Vogel 1986).

I conclude, then, that the phonology relating clitics to their hosts is in general of the "post-lexical" type, with the specifics depending on the regularities governing various prosodic constituent types within a given language. Given the current state of instability that governs the architecture of phonological theory, with classical rule-based Lexical Phonology and its most direct constraint-based dependent, Stratal OT (Bermúdez-Otero, forthcoming), in conflict both with the "standard" monolithic model of OT and also with various alternatives such as

OT-CC (McCarthy 2007) and Optimal Interleaving (Wolf 2008), Phase-based Phonology, as represented by various papers in Grohmann (2009), and others, it is difficult to see the facts above from Catalan as decisively incompatible with the view of clitics as phonologically integrated with their hosts at the syntactic level, rather than in the lexicon.

6 Conclusion

There is very little to the phonology of clitics, then, that is unique to these elements. In terms of their representation, they have the character of being incompletely organized in prosodic terms: they are deficient in not constituting PWords, as opposed to normal lexical items. Once that is taken into account, the rest of their behavior follows from the prosodic phonology of the language. Aspects of prosodic well-formedness require that they undergo Stray Adjunction, or incorporation into adjacent prosodic units at some level, in ways that depend on the language's particular ranking of constraints governing prosodic structure. The language's "post-lexical" phonology (in some appropriate, architecture-dependent sense) then governs adjustments in the phonological shape of the resulting combination of clitic and host. Neither the prosodic organization nor the phonological adjustments involved are uniquely identified with clitics, although clitics may well provide essential clues in the determination of how the phonology (including prosody) of a language works.

A full treatment of the linguistic category of "clitics," of course, would have to deal with more than the phonological characteristics of items so designated. In particular, the principles underlying the distinctive (morpho)syntactic behavior of "Special Clitics" must be elucidated. Linguistic items that show clitic behavior in the morphosyntactic sense are usually, though not always, prosodically deficient and thus phonologically clitic as well. The analysis of this dimension of the (not entirely homogeneous) class of "clitics" would, however, take us much too far afield in the context of this *Companion*, and the interested reader can only be referred to Anderson (2005) for the development of one view.

ACKNOWLEDGMENTS

This work was supported in part by NSF award #BCS 98-76456 to Yale University. I thank two anonymous referees and the editors of the *Companion* for useful comments, and also Argyro Katsika for her assistance with the phonology of Modern Greek. None of these people is to be held accountable for my use of the information they supplied, of course.

REFERENCES

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Anderson, Stephen R. 2005. *Aspects of the theory of clitics*. Oxford: Oxford University Press.

Anderson, Stephen R. 2008. English reduced auxiliaries really are simple clitics. *Lingue e Linguaggio* 7. 169–186.

Arvaniti, Amalia. 1992. Secondary stress: Evidence from Modern Greek. In Gerard J. Docherty & D. Robert Ladd (eds.) *Papers in laboratory phonology II: Gesture, segment, prosody*, 398–423. Cambridge: Cambridge University Press.
Bermúdez-Otero, Ricardo. Forthcoming. *Stratal Optimality Theory*. Oxford: Oxford University Press.
Bermúdez-Otero, Ricardo & John Payne. Forthcoming. There are no special clitics. In Alexandra Galani, Glyn Hicks & George Tsoulas (eds.) *Morphology and its interfaces*. Amsterdam & Philadelphia: John Benjamins.
Booij, Geert. 1988. Review of Nespor & Vogel (1986). *Journal of Linguistics* 24. 515–525.
Booij, Geert & Jerzy Rubach. 1987. Postcyclic versus postlexical rules in Lexical Phonology. *Linguistic Inquiry* 18. 1–44.
Delbrück, Berthold. 1878. *Syntaktische Forschungen*, vol. 3: *Die altindische Wortfolge aus dem çatapathabrâ mana*. Halle: Buchhandlung des Waisenhauses.
Dixon, R. M. W. & Alexandra Y. Aikhenvald (eds.) 2002. *Word: A cross-linguistic typology*. Cambridge: Cambridge University Press.
Grohmann, Kleanthes K. (ed.) 2009. *Interphases: Phase-theoretic investigations of linguistic interfaces*. Oxford: Oxford University Press.
Halpern, Aaron L. 1998. Clitics. In Andrew Spencer & Arnold M. Zwicky (eds.) *The handbook of morphology*, 101–122. Oxford & Malden, MA: Blackwell.
Harris, Alice C. 2002. *Endoclitics and the origins of Udi morphosyntax*. Oxford: Oxford University Press.
Inkelas, Sharon. 1989. Prosodic constituency in the lexicon. Ph.D. dissertation, Stanford University.
Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.
Klavans, Judith L. 1985. The independence of syntax and phonology in cliticization. *Language* 61. 95–120.
McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.
Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
Obata, Kazuko. 2003. *A grammar of Bilua, a Papuan language of the Solomon islands*. (*Pacific Linguistics* 540.) Canberra: Australian National University.
Peperkamp, Sharon. 1997. Prosodic words. Ph.D. dissertation, University of Amsterdam.
Revithiadou, Anthi. 2008. A cross-dialectal study of cliticization in Greek. *Lingua* 118. 1393–1415.
Russi, Cinzia. 2008. *Italian clitics: An empirical study*. Berlin & New York: Mouton de Gruyter.
Selkirk, Elisabeth. 1995. The prosodic structure of function words. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 439–469. Amherst: GLSA.
Taylor, Ann. 1990. Clitics and configurationality in Ancient Greek. Ph.D. dissertation, University of Pennsylvania.
Taylor, Ann. 1996. A prosodic account of clitic position in ancient Greek. In Aaron L. Halpern & Arnold M. Zwicky (eds.) *Approaching second: Second position clitics and related phenomena*, 477–503. Stanford: CSLI.
Wackernagel, Jacob. 1892. Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen* 1. 333–436.
Wheeler, Max W. 2005. *The phonology of Catalan*. Oxford: Oxford University Press.
Wolf, Matthew. 2008. Optimal interleaving: Serial phonology–morphology interaction in a constraint-based model. Ph.D. dissertation, University of Massachusetts, Amherst.
Zec, Draga & Sharon Inkelas. 1991. The place of clitics in the prosodic hierarchy. *Proceedings of the West Coast Conference on Formal Linguistics* 10. 505–519.
Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. Cliticization *vs.* inflection: English *n't*. *Language* 59. 502–513.

85 Cyclicity

Ricardo Bermúdez-Otero

1 Introduction

The phonology of a natural language will often treat the same string differently according to whether it is wholly contained within a single morph, arises through a morphological operation like affixation, or straddles the edges of two adjacent grammatical words. In the generative tradition there is a widespread and long-standing consensus that such morphosyntactic conditioning effects may come about in two ways: representationally or procedurally (Scheer 2008: §3ff.; see Table 85.1). Representational morphosyntactic conditioning occurs when phonological processes are sensitive to the presence or absence of certain phonological objects – boundary symbols in *SPE*, prosodic categories in most later frameworks – which are in turn positioned by reference to the edges of morphosyntactic units. In procedural morphosyntactic conditioning, in contrast, morphosyntax directly controls the amount of structure visible during a given round of phonological computation, either by submitting to the phonology only a morphosyntactic sub-constituent of a complete linguistic expression (as in the theory of the cycle) or

Table 85.1 Two types of morphosyntactic conditioning acknowledged throughout the history of generative phonology

Theory	*Representational effects*	*Procedural effects*	*Sample reference*
SPE	boundary symbols (+, #)	the cycle	Chomsky & Halle (1968)
Lexical Phonology	prosodic units (built by rules)	the cycle (with levels)	Booij & Rubach (1984)
Stratal OT	prosodic units (controlled by Align)	the cycle (with levels)	Bermúdez-Otero & Luís (2009)
Classical OT	prosodic units (controlled by Align)	OO-correspondence	Raffelsiefen (2005)
Lateral Phonology	empty CV units	the cycle (phases)	Scheer (2008)

by allowing the phonology access to the surface representation of some morphosyntactically related expression (as in the theory of transderivational or output–output correspondence, henceforth OO-correspondence).

This chapter addresses current debates about procedural morphosyntactic conditioning in phonology, focusing in particular on the contest between the cycle and OO-correspondence (§5–§9); we shall be concerned with prosody only insofar as it raises the non-trivial problem of demarcating procedural from representational effects (§4). Much of the discussion will be taken up with three instances of morphosyntactically induced misapplication that challenge the basic premises of transderivational theories: in all three cases, the surface bases needed for an analysis relying on OO-correspondence appear to be unavailable for phonological or morphological reasons (§6–§8). As the argument unfolds, however, it will become clear that questions about morphosyntax–phonology interactions are intricately entangled with problems in every other area of phonology, notably including the theory of representations, the phonology–phonetics interface, and the balance between synchronic and diachronic explanation.

2 Two cases of cyclic misapplication in English: Post-nasal plosive deletion and Belfast dentalization

Let us begin with a well-known instance of morphologically induced overapplication. Present-day English tolerates homorganic consonant clusters consisting of a nasal followed by a non-coronal voiced plosive (i.e. [b] or [g]) only if the latter is syllabified in onset position; if the plosive would otherwise surface in the coda, it undergoes deletion (Borowsky 1993: 202).[1]

(1)	a.	*bomb*	[bɒm]	b.	*bombard*$_V$	[ˌbɒm.ˈbɑːd]
		thumb	[θʌm]		*thimble*	[ˈθɪm.bl̩]
		crumb	[kɹʌm]		*crumble*	[ˈkɹʌm.bl̩]
		long	[lɒŋ]		*elongate*$_V$	[ˈiː.lɒŋ.ˌgeɪt]

The forms in (1a) and (1b) display normal application and non-application of deletion, respectively. In (2a), however, the process overapplies: the plosives [b] and [g] fail to surface stem-finally, even though in that position they would be syllabified as onsets; cf. (2b).[2]

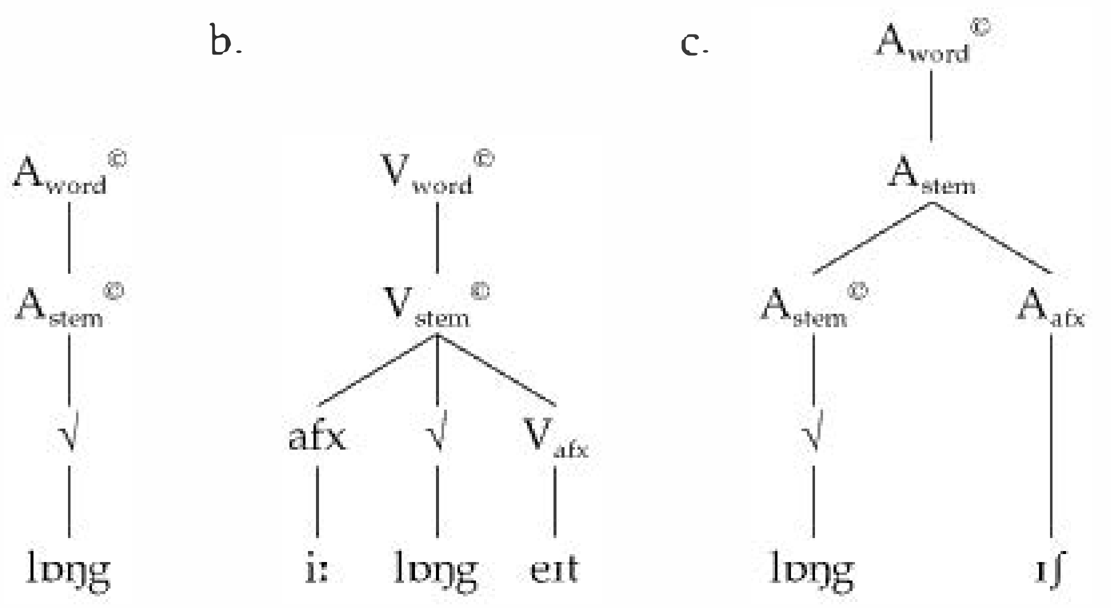
(2)	a.	*bomb-ing*	[ˈbɒ.mɪŋ]	b.	*[ˈbɒm.bɪŋ]
		thumb-ing	[ˈθʌ.mɪŋ]		*[ˈθʌm.bɪŋ]
		crumb-y	[ˈkɹʌ.mɪ]		*[ˈkɹʌm.bɪ]
		long-ish	[ˈlɒ.ŋɪʃ]		*[ˈlɒŋ.gɪʃ]

[1] All varieties of English exhibit post-nasal /b/-deletion; /g/-deletion varies across dialects.

[2] *Thumb*$_V$ and *thimble* are highly unlikely to be synchronically related, and so native speakers probably have no reason to derive the noun *thumb*$_N$ or the converted verb *thumb*$_V$ from a root /θʌmb/. If so, the gerund *thumb-ing* [ˈθʌ.mɪŋ] is in fact transparent. This does not affect our argument, however: the key point is that the grammar of English systematically disallows transparent alternations between infinitives ending in [. . . Vm] and gerunds ending in [. . . Vm.bɪŋ].

According to the theory of the phonological cycle, first formulated by Chomsky *et al.* (1956: 75), the key to such instances of morphosyntactically induced misapplication is to be found in part–whole relationships within the grammatical constituent structure of the relevant linguistic expressions. Consider, for example, the morphological structure of the adjective *long*, the verb *elongate*, and the derived adjective *longish*.

(3) a. b. c.



Let us suppose that some of the morphosyntactic constituents shown in (3) define domains for phonological computation; I shall henceforth refer to these as "cyclic nodes." Assume, at a minimum, that the set of cyclic nodes in (3) includes every stem immediately derived from a root, as well as every fully inflected, syntactically free grammatical word; these are flagged with a superscript ●. Given these premises, one obtains the following nested domain structures:

(4) a. [[lɒŋg]] b. [[iː-lɒŋg-eɪt]] c. [[lɒŋg] ɪʃ]

Now suppose that phonological computation proceeds iteratively, starting with the domains defined by the smallest, most deeply embedded cyclic nodes, and then moving to larger, less deeply embedded cyclic nodes: in other words, suppose that the computation of the phonological form of the parts precedes and feeds the computation of the phonological form of the whole.

(5)	a. [[lɒŋg]]	b. [[iː-lɒŋg-eɪt]]	c. [[lɒŋg] ɪʃ]
inner cycle	lɒŋ	iː.lɒŋ.geɪt	lɒŋ
outer cycle	—	—	lɒ.ŋɪʃ

According to this cyclic analysis, post-nasal plosive deletion overapplies in *longish* because its conditions are met within a morphosyntactic subconstituent, the stem *long-*, which defines a cyclic domain by itself. The environment for deletion disappears in the outer cycle, as the vowel of the suffix *-ish* projects a syllable with an onset capable of sheltering the underlying /g/; but deletion has already applied in the inner cycle. The result is a counterbleeding interaction.

Observe that not all morphosyntactic constituents trigger phonological cycles. In (3b) and (4b), for example, it is absolutely crucial that roots (as opposed to stems) should not count as cyclic nodes (Kiparsky 1982: 32–33; Inkelas 1989: §3.5.5); otherwise, post-nasal plosive deletion would incorrectly overapply in *e-long-ate*.[3] A fully articulated theory of the cycle must of course specify criteria for designating particular morphosyntactic nodes as cyclic or non-cyclic. Scholars working in the tradition of Lexical Phonology (e.g. Kiparsky 1982; Hargus and Kaisse 1993) and Stratal OT (e.g. Bermúdez-Otero 1999, 2003; Kiparsky 2000) have reached a broad consensus on a number of points, sometimes strikingly at variance with phase theory in minimalist syntax (Chomsky 2001): e.g. there appear to be no cyclic nodes between the grammatical word (X^0) and the utterance (see further Scheer 2008: §740ff.). Many other issues remain open: e.g. whether or not certain affixes should be allowed to define cyclic domains by themselves (Baker 2005; see also Mohanan 1982, and cf. McCarthy 2007: 133–134).

To account for morphologically induced underapplication, the theory of the cycle needs to be supplemented with the concept of level or stratum (CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME).[4] Consider, for example, the process of dentalization found in certain varieties of English spoken in and around Belfast: the coronal non-continuants /t d n l/ have alveolar realizations unless immediately followed by /(ə)ɹ/, in which case they become dental (Harris 1985: 58, 211ff.).

(6)	a.	*train*	[t̪ɾ̥ɪən]	*drain*	[d̪ɾ̥ɪən]				
		true	[t̪ɾ̥ʉ:]	*drew*	[d̪ɾʉ:]				
	b.	*Peter*	['pit̪əʴ]	*ladder*	['lad̪əʴ]	*dinner*	['dɛ̈n̪əʴ]	*pillar*	['pɛ̈l̪əʴ]
		matter	['mat̪əʴ]	*rudder*	['ɹɔ̈d̪əʴ]	*spanner*	['spän̪əʴ]		

Dentalization underapplies when its environment is created by adding a suffix like agentive *-er* (7a) or comparative *-er* (7b) to a free stem, although it does apply normally when comparative *-er* is suffixed to a suppletive bound root (7c).

(7)	a.	*hea*[t]*er*	*loa*[d]*er*	*di*[n]*er* 'diner'	*ki*[l]*er*	
		wai[t]*er*		*ru*[n]*er*		
		shou[t]*er*				
	b.	*fa*[t]*er*	*lou*[d]*er*	*fi*[n]*er*	*coo*[l]*er*	
		la[t]*er*				
	c.	*better*	['bæt̪əʴ]	'good.COMPARATIVE'		
		cf. *better*	['bætəʴ]	'one who bets'		

Applying the same principles as in (4), we obtain cyclic domain structures like the following:

(8)	a.	[[*train*]]		b.	[[*heat*] *er*]
		[[*Peter*]]			[[*fat*] *er*]
		[[*bett-er*]] 'good.COMPARATIVE'			[[*bet*] *er*] 'one who bets'

However, it becomes immediately apparent that, if dentalization applies in every cycle, then cyclic derivation will just fail to produce the desired underapplication in (8b): dentalization will simply take place in the outer cycle.

3 For the root-based status of this form, compare verbs like *e-dulcor-ate*, *e-mancip-ate*, *e-viscer-ate*, etc., which are manifestly derived from uninflectable bound bases.

4 Some scholars pursue an alternative approach, based on Chomsky's (2001) Phase Impenetrability Condition: e.g. Marvin (2002), Scheer (2008).

Many cyclic frameworks solve this problem by asserting that phonological domains associated with morphosyntactic constituents of different kinds may be subject to different phonological generalizations: in common usage, such morphosyntactic constituents are said to "belong to different phonological levels." Theories differ as to the number of phonological levels that may be distinguished within the grammar of a single language. Lexical Phonology and Stratal OT often assume that each grammar specifies precisely three levels: the stem, word, and phrase levels. In affixal constructions, the ascription of the construction to the stem-level or word-level phonology is deemed to depend on properties both of the base (Giegerich 1999) and of the affix: the attachment of an affix to a root necessarily produces a stem-level category; the attachment of an affix to a stem may produce a stem-level or word-level category depending on the idiosyncratic affiliation of the affix (e.g. Bermúdez-Otero 2007d: 283). In constrast, full grammatical words trigger word-level cycles and complete utterances trigger phrase-level cycles.

In the case of Belfast English, one must assume that dentalization applies only within stem-level domains, and that agentive *-er* and comparative *-er* are word-level suffixes unless attached to bound roots. This yields the appropriate counterfeeding relationship between stem-level dentalization and word-level suffixation.

(9)		
	Peter	'good.COMPARATIVE'
	$_{WL}$[$_{SL}$[*Peter*]]	$_{WL}$[$_{SL}$[*bett-er*]]
SL (dentalization on)	t̪	t̪
WL (dentalization off)	–	–
	fatt-er	'one who bets'
	$_{WL}$[$_{SL}$[*fat*] *er*]	$_{WL}$[$_{SL}$[*bet*] *er*]
SL (dentalization on)	–	–
WL (dentalization off)	–	–

3 The Russian Doll Theorem and the life cycle of phonological processes

There is no room in this brief chapter to review all the predictions about morphosyntactically induced misapplication that follow from the theory of the cycle. It will therefore be appropriate to concentrate here on one of the most fundamental:

(10) *The Russian Doll Theorem*

Let there be the nested cyclic domains $_{\gamma}[\ldots\ _{\beta}[\ \ldots\ _{\alpha}[\ldots]\ \ldots]\ \ldots]$. If a phonological process p is opaque in β because its domain is α, then p is opaque in γ.

To my knowledge, this entailment of cyclic theory has not been formally enunciated before, probably because it has been considered so obviously true as to be entirely trivial. Later, however, we shall see that OO-correspondence is easily capable of violating the Russian Doll Theorem and captures its effects only by stipulation (§9).

The Russian Doll Theorem has the following corollary:

(11) If a phonological process exhibits cyclic misapplication within a certain phonological configuration created by affixation, then it must also exhibit cyclic misapplication if the same configuration arises by word concatenation.

This follows logically from elementary facts of morphosyntactic layering: a phonological process can cyclically misapply in the presence of an affix only if that affix is excluded from its cyclic domain, which must therefore correspond to a morphosyntactic category smaller than the grammatical word, i.e. a stem; but, by its very nature, a stem cannot straddle the edges of adjacent words. Post-nasal plosive deletion (§2) bears out this prediction: overapplication before word-level suffixes beginning with a vowel, as in *long-ish* [ˈlɒ.ŋɪʃ], entails overapplication in word-final prevocalic environments, as in *long effect* [ˌlɒ.ŋɪ.ˈfɛkt]; cf. *[ˌlɒŋ.gɪ.ˈfɛkt].

Not only does corollary (11) hold true of post-nasal plosive deletion in present-day English, but it also captures key facts in the diachronic evolution of the process. For /ŋg/ clusters, in particular, we can reconstruct the four historical stages shown in (12): Stage 0 represents the situation in Early Modern English; Stage 1 is attested in the formal, relatively conservative register of eighteenth-century orthoepist James Elphinston; Stage 2 corresponds to Elphinston's description of his own casual, more innovative register; and Stage 3 is observed in present-day RP (Garrett and Blevins 2009: 527–528). The symbol ‖ represents pause (i.e. the end of the phonological utterance).

(12)

	Stage 0	Stage 1	Stage 2	Stage 3
elongate	ŋg	ŋg	ŋg	ŋg
prolong-er	ŋg	ŋg	ŋg	ŋ
prolong it	ŋg	ŋg	ŋ	ŋ
prolong ‖	ŋg	ŋ	ŋ	ŋ

In compliance with (11), the diachronic transition from normal application (Stage 1) to word-internal overapplication (Stage 3) is effected through an intermediate phase involving overapplication at word boundaries but normal application word-internally (Stage 2).

More generally, the diachronic pathway shown in (12) provides a clear illustration of the typical life cycle of phonological processes, which stratal-cyclic frameworks capture in a particularly perspicuous way; see e.g. Bermúdez-Otero (1999: 99–103, 239–240; 2007b: 503) and McMahon (2000: ch. 4). First, phonetically driven innovations enter the grammar from below as gradient phonetic rules, which later become stabilized as categorical phonological processes applying across the board at the phrase level (Bermúdez-Otero 2007b: 505; see also CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY; CHAPTER 93: SOUND CHANGE): in (12), this is the transition from Stage 0 to Stage 1. Subsequently, analogical change causes the new phonological process to climb up to progressively higher levels, concomitantly narrowing down its domain of application (Dressler 1985: 149): in (12) we see deletion climbing up from the phrase level (Stage 1) to the word level (Stage 2), and from the word level (Stage 2) to the stem level

(Stage 3). Eventually, senescent processes typically undergo morphologization or lexicalization: see Bermúdez-Otero (2008) for incipient symptoms of this in the case of post-nasal plosive deletion.

The overall sequence of events is represented in greater detail in (13), where g̶ indicates a site of /g/-deletion.

(13)

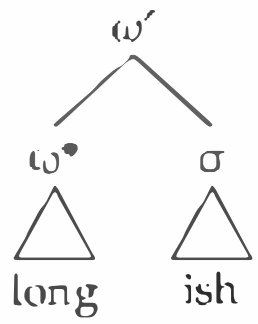
	level	deletion?	*elongate*	*prolonging* ‖	*prolong it*	*prolong* ‖
a.	*Stage 0: Early Modern English*					
	SL	no	[iː.lɒŋ.geɪt]	[pɹə.lɒŋg][ɪŋg]	[pɹə.lɒŋg][ɪt]	[pɹə.lɒŋg]
	WL	no	[iː.lɒŋ.geɪt]	[pɹə.lɒŋ.gɪŋg]	[pɹə.lɒŋg][ɪt]	[pɹə.lɒŋg]
	PL	no	[iː.lɒŋ.geɪt]	[pɹə.lɒŋ.gɪŋg]	[pɹə.lɒŋ.gɪt]	[pɹə.lɒŋg]
b.	*Stage 1: Elphinston's formal register*					
	SL	no	[iː.lɒŋ.geɪt]	[pɹə.lɒŋg][ɪŋg]	[pɹə.lɒŋg][ɪt]	[pɹə.lɒŋg]
	WL	no	[iː.lɒŋ.geɪt]	[pɹə.lɒŋ.gɪŋg]	[pɹə.lɒŋg][ɪt]	[pɹə.lɒŋg]
	PL	yes	[iː.lɒŋ.geɪt]	[pɹə.lɒŋ.gɪŋg̶]	[pɹə.lɒŋ.gɪt]	[pɹə.lɒŋg̶]
c.	*Stage 2: Elphinston's casual register*					
	SL	no	[iː.lɒŋ.geɪt]	[pɹə.lɒŋg][ɪŋg]	[pɹə.lɒŋg][ɪt]	[pɹə.lɒŋg]
	WL	yes	[iː.lɒŋ.geɪt]	[pɹə.lɒŋ.gɪŋg̶]	[pɹə.lɒŋg̶][ɪt]	[pɹə.lɒŋg̶]
	PL	yes (vacuous)	[iː.lɒŋ.geɪt]	[pɹə.lɒŋ.gɪŋ]	[pɹə.lɒ.ŋɪt]	[pɹə.lɒŋ]
d.	*Stage 3: Present-day RP*					
	SL	yes	[iː.lɒŋ.geɪt]	[pɹə.lɒŋg̶][ɪŋg̶]	[pɹə.lɒŋg̶][ɪt]	[pɹə.lɒŋg̶]
	WL	yes (vacuous)	[iː.lɒŋ.geɪt]	[pɹə.lɒ.ŋɪŋ]	[pɹə.lɒŋ][ɪt]	[pɹə.lɒŋ]
	PL	yes (vacuous)	[iː.lɒŋ.geɪt]	[pɹə.lɒ.ŋɪŋ]	[pɹə.lɒ.ŋɪt]	[pɹə.lɒŋ]

The analogical changes involved in the transitions between Stages 1 and 2 and between Stages 2 and 3 were driven by input restructuring (Bermúdez-Otero and Hogg 2003: 105ff.; Bermúdez-Otero 2006: 501ff.). At Stage 1, for example, surface $_{\text{PL}}$[pɹə.lɒŋ‖] was derived unfaithfully from word-level $_{\text{WL}}$[pɹə.lɒŋg] by a phrase-level application of deletion. By Stage 2, however, $_{\text{PL}}$[pɹə.lɒŋ‖] has been re-analyzed as derived faithfully from an identical word-level representation $_{\text{WL}}$[pɹə.lɒŋ]. This has the effect of introducing deletion into the word-level phonology, and gives rise to the innovative opaque surface form $_{\text{PL}}$[pɹə.lɒ.ŋɪt], derived from word-level $_{\text{WL}}$[pɹə.lɒŋ] $_{\text{WL}}$[ɪt]. Bermúdez-Otero (1999: 100–103, 239–240; 2003: 4ff.) outlines an approach to phonological learning that accounts straightforwardly for such patterns of recurrent input restructuring.

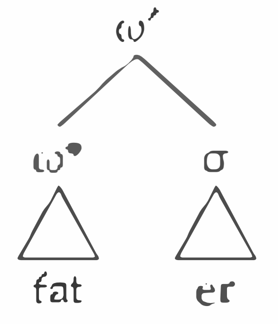
4 Cyclicity *vs.* prosody

In §2 we assumed that the morphosyntactic conditioning effects displayed by post-nasal plosive deletion and Belfast dentalization were procedural, not representational (see §1). However, several scholars have proposed that the behavior of English word-level suffixes should be explained prosodically, rather than cyclically (e.g. Szpyra 1989: 178–200; Hammond 1999: 322–329). In this approach, suffixes like agentive *-er* and adjectival *-ish* are not incorporated into the prosodic word containing the stem, but attach under a second projection of ω (CHAPTER 51: THE PHONOLOGICAL WORD):

(14) a.



b.



If this were true, then the absence of dentalization in Belfast *fatt-er* ['fatɚ], cf. *['fat̪ɚ] (§2), could be described as a case of transparent non-application, rather than as an instance of opaque underapplication: one would just need to stipulate that dentalization does not apply unless its conditions are met within the first projection of ω.

(15) $\begin{bmatrix} \text{coronal} \\ -\text{continuant} \end{bmatrix} \rightarrow$ [+distributed] / $_{\omega^*}$[. . . __ (ə)ɹ . . .]

The uncertainty whether a particular instance of morphosyntactic conditioning in phonology should be analyzed procedurally or representationally is in fact one of the most serious and recurrent obstacles faced by empirical research into the morphosyntax–phonology interface. A great deal of existing work fails to make the cut on explicit, consistent, and principled grounds (Raffelsiefen 2005: 214–215). Typically, the question cannot be settled without analyzing a very substantial fragment of the phonology of the language in question, including both morphosyntax–phonology and phonology–phonetics interactions (e.g. Bermúdez-Otero and Luís 2009).

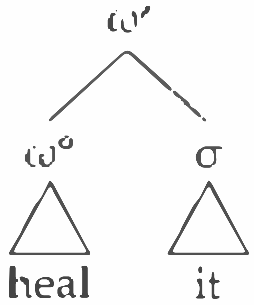
In our case, phonological variation and phonetic gradience in English provide strong evidence against the prosodifications shown in (14). Let us first consider variation. Hayes (2000: 98) shows that, in American English, the application frequency of /l/-darkening (CHAPTER 31: LATERAL CONSONANTS) follows the cline in (16):

(16) higher frequency of [ɫ] lower frequency of [ɫ]

heal ‖	>	*heal it*	>	*heal-ing*	>	*Healey*

On the basis of a comprehensive survey of English function words, however, Selkirk (1996: 204–206) shows that combinations of a verb and a weak object pronoun like *heal it* undergo affixal cliticization (CHAPTER 84: CLITICS):

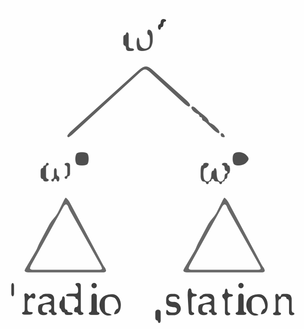
(17)



Therefore, if one adopts the approach to English word-level suffixes shown in (14), *heal-ing* will end up being prosodified in the same way as *heal it*, and so prosody will be unable to explain the fact that /l/-darkening applies with greater frequency in the latter than in the former. One would then have to fall back on a procedural (cyclic) explanation: see §9.

The argument from variable /l/-darkening suggests that the prosodification shown in (14) is descriptively insufficient (though cf. Raffelsiefen 2005: 253–256); the evidence of gradient durational effects confirms that it is incorrect. It is a well-known fact that, in English, each of the members of a transparent compound forms a prosodic word by itself:

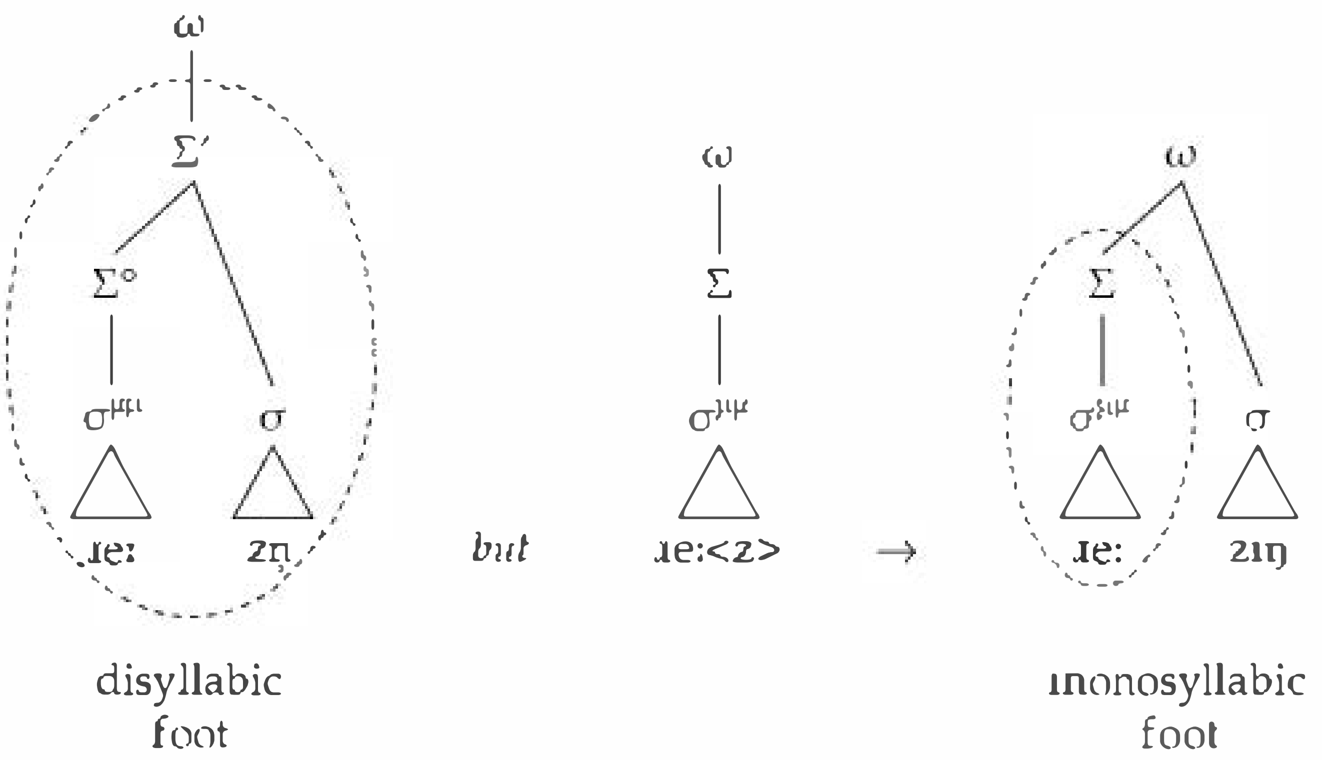
(18)



Given this fact, the approach to word-level suffixes outlined in (14) predicts that stems in word-level suffixal constructions will display the same patterns of gradient durational adjustment (resistance to polysyllabic shortening; liability to pre-boundary lengthening) as the first members of transparent compounds, since both occur in the environment $_{\omega'}[\ _{\omega^\bullet}[\ \_\_\ ]\ldots]$. This prediction proves incorrect. In an experiment with nonce words, Sproat and Fujimura (1993) found no durational effects of stem-level suffixation (e.g. ***beel**-ic*) or word-level suffixation (e.g. ***beel**-ing*) when compared with monomorphemic controls (e.g. ***Beel**ik*), whereas the first members of compounds (e.g. ***beel** equator*) were consistently lengthened; see Sproat (1993: 178). A more recent study of Scottish English has detected a very small effect of word-level suffixation: the phonetic realization of the string /ɹeːz/ appears to be slightly shorter in *raisin* ['ɹeːzn̩] than in *rais-ing* ['ɹeːzɪŋ] (Sugahara and Turk 2009). Nonetheless, this effect falls far below that of compounding: it is not statistically significant at "normal" speech rates (Sugahara and Turk 2009: 496); it manifests itself as a 6.6 percent difference (mean of 23 msecs) at "slow" speech rates; and it reaches only 9.6 percent (mean of 42 msecs) at "extra-slow" speech rates. *Pace* Sugahara and Turk (2009: 488), these findings are best understood as an effect of footing, rather than of recursive prosodic-word structure (CHAPTER 40: THE FOOT):[5]

[5] In (19a), Noun Extrametricality (Hayes 1982: 240) is implemented through the exclusion of the final syllable of the noun stem from the first foot projection (Σ•) and through its attachment under a second foot projection (Σ′) (CHAPTER 43: EXTRAMETRICALITY AND NON-FINALITY).

(19) a. *raisin* b. *raise* c. *rais-ing*



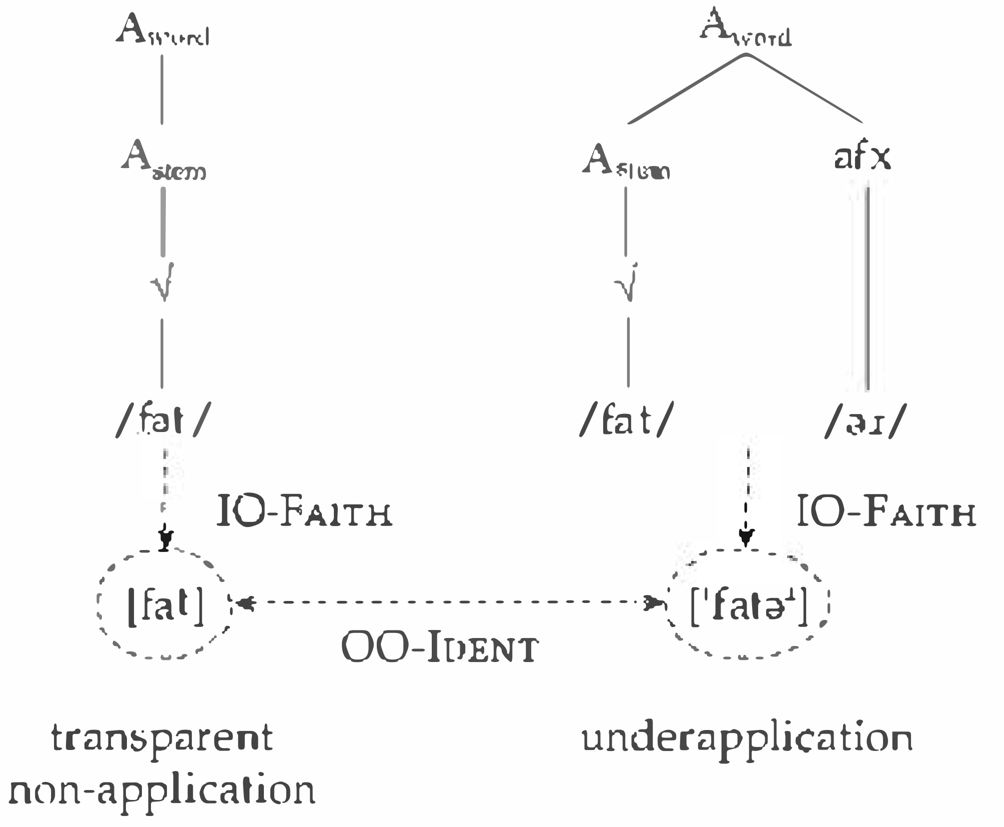
Indeed, the idea that stray syllables affiliated to word-level suffixes attach directly to ω, as in (19c), instead of being footed, makes straightforward sense of the fact that, unless autostressed, word-level suffixes are stress-neutral.

The diagnostics that we have applied so far can be used to demarcate representational morphosyntactic conditioning from procedural morphosyntactic conditioning regardless of one's particular theory of the latter. However, if one commits to a stratal-cyclic analysis of procedural morphosyntactic conditioning, then further demarcation criteria become available. One such criterion is cyclic locality: prosodic structure assigned in an early cycle can persist, and continue to affect the application of phonological processes, throughout later cycles; in contrast, the morphosyntactic structure visible during a phonological cycle ceases to be accessible in the next cycle (by "Bracket Erasure": see e.g. Orgun and Inkelas 2002: 116). Cyclic locality entails, for example, that the contrast between American English ˌ*capi*[ɾ]*a'listic* and ˌ*mili*[t]*a'ristic* must be mediated by prosody, as /t/-flapping is demonstrably phrase-level (see (36b) below) and so cannot access the internal morphological structure of words: see e.g. Davis (2005) and Bermúdez-Otero and McMahon (2006: 403–404); cf. Steriade (2000).

5 Cyclicity *vs.* OO-correspondence

Whilst phonologists generally agree that both representational and procedural morphosyntactic conditioning effects exist, as we saw in §1 and §4, there is currently no consensus on the best way to analyze procedural morphosyntactic conditioning. Within OT, the most popular alternative to the cycle is transderivational correspondence (e.g. Kenstowicz 1996; Benua 1997; Kager 1999; etc.). This theory claims that morphosyntactically induced misapplication arises when high-ranking OO-identity constraints cause a transparently derived surface property of a given expression (the "surface base") to be transmitted to the surface representation of some morphosyntactically related expression, where its presence is opaque. Thus, the underapplication of Belfast dentalization in *fatt-er* ['fatəˑ] (§2) would be analyzed as follows:

(20)



The implementation of this solution poses a number of non-trivial technical challenges, such as motivating the selection of the surface base and preventing the satisfaction of OO-identity by means of overapplication in the base (i.e. transparent /fat-əɹ/ → *['fat̪əɹ] leading to opaque /fat/ → *[fat̪]); I return to these issues in §9 below.

At this point, however, I should like to compare the core predictions of cyclicity and OO-correspondence. The comparison is in fact easy, because the two theories share a fundamental assumption:

(21) *Ultimate transparency*

If a phonological generalization *p* misapplies in the surface representation *s* of some linguistic expression, then *p* must apply transparently in some other representation *r*, with which *s* is in direct or indirect correspondence.

The theory of the cycle predicts that *p* will apply transparently in some cyclic domain defined by some morphosyntactic constituent of the expression: the output of this cycle is connected with the surface representation by relationships of input–output faithfulness. In contrast, OO-correspondence predicts that *p* will apply transparently in the surface representation of some appropriately related linguistic expression; the two surface representations are linked to each other by means of transderivational correspondence. In §6 to §8 I adduce empirical evidence supporting the first prediction and challenging the second.

6 Phonologically masked bases I: Quito Spanish /s/-voicing

Spanish has a voiceless alveolar fricative phoneme /s/. In the dialect spoken in Quito (Robinson 1979; Lipski 1989), /s/ is realized faithfully in the onset (22a), but displays contextual laryngeal allophony in the coda: coda /s/ surfaces as [s] before voiceless segments and utterance-finally (22b), and becomes [z] when followed by a voiced segment either in the same grammatical word or across a word boundary (22c).

(22)	a.	*gasa*	/gasa/	[ˈga.sa]	'gauze'
		ganso	/gaNso/	[ˈgan.so]	'gander'
		da sueño	/da sueɲo/	[ˌd̪a.ˈswe.ɲo]	'makes one sleepy'
		el sueño	/el sueɲo/	[el.ˈswe.ɲo]	'the dream'
	b.	*rasco*	/rasko/	[ˈras.ko]	'I scratch'
		gas caro	/gas kaɾo/	[ˌgas.ˈka.ɾo]	'expensive gas'
		gas ‖	/gas/	[gas ‖]	'gas'
	c.	*rasgo*	/rasgo/	[ˈraz.ɰo]	'feature'
		plasma	/plasma/	[ˈplaz.ma]	'plasma'
		gas blanco	/gas blaNko/	[ˌgaz.ˈβ̞laŋ.ko]	'white gas'
		gas noble	/gas noble/	[ˌgaz.ˈno.β̞le]	'noble gas'

Coda /s/ undergoes voicing not only before voiced obstruents, but also before sonorants: e.g. *plasma* [ˈplaz.ma], *gas noble* [ˌgaz.ˈno.β̞le]. For our purposes, the crucial fact is that voicing overapplies to word-final prevocalic /s/:

(23)	a.		*gas acre*	/gas akɾe/	[ˌga.ˈza.kɾe]	'acrid gas'
		cf.	*gasa*	/gasa/	[ˈga.sa]	'gauze'
	b.		*has ido*	/as ido/	[a.ˈzi.ð̞o]	'hast gone'
		cf.	*ha sido*	/a sido/	[a.ˈsi.ð̞o]	'hath been'

On the surface, expressions like *gas acre* [ˌga.ˈza.kre] fail to meet the conditions for /s/-voicing: in *gas acre*, [z] surfaces in a pre-sonorant environment, but not in the coda, for Spanish has a phrase-level process of resyllabification that moves word-final prevocalic consonants into the onset.[6] In this position, therefore, the transparent realization of /s/ would be voiceless: cf. *gasa* [ˈga.sa].

In a stratal-cyclic framework, the laryngeal allophony of Quito Spanish /s/ submits to the following analysis. First, the stem-level phonology allows output [s], but forbids output [z]: in an optimality-theoretic implementation, therefore, a hypothetical underlying /z/ present in the rich base would be unfaithfully mapped onto [s] in the stem-level output (see Bermúdez-Otero 2007c for an illustration of this strategy). At the word level, in turn, [s] remains unchanged if syllabified in the onset; in the coda, however, [s] loses its laryngeal node, becoming laryngeally unspecified [S]: see (24a). Finally, at the phrase level, input [s] is realized faithfully, whereas delaryngealized [S] acquires voice specifications either by leftward autosegmental spreading from an immediately following obstruent or by default: on the assumption that sonorants are not redundantly specified as [+voice] (CHAPTER 8: SONORANTS), we can just say that [S] becomes voiced before sonorants in order to satisfy a positional constraint designating [+voice] as the unmarked feature in this particular context, whereas utterance-final [S] is assigned the context-free default specification [−voice];

[6] This is confirmed, *inter alia*, by the fact that [ɾ] undergoes optional emphatic trilling in canonical coda positions, but not word-finally before a vowel (Harris 1983: 70–71): e.g. [maɾ] ~ [mar] 'sea', [ˌmaɾ.ˈne.ɰɾo] ~ [ˌmar.ˈne.ɰɾo] 'Black Sea'; but [ˌma.ɾe.ˈxe.o] 'Aegean Sea', not *[ˌma.re.ˈxe.o].

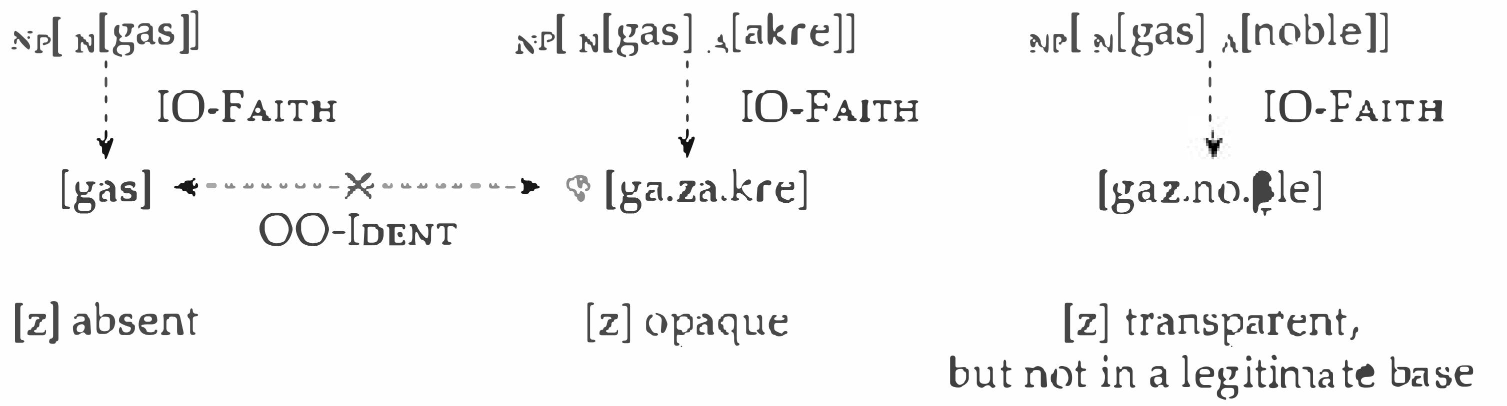
see (24b).[7] In this analysis, underlying /s/ becomes vulnerable to voicing if it finds itself in the coda in a word-level cycle and so loses its laryngeal node; the generalization is rendered opaque by phrase-level resyllabification.[8]

(24)

		$_{PL}$[$_{WL}$[gasa]]	$_{PL}$[$_{WL}$[gas] $_{WL}$[akre]]	$_{PL}$[$_{WL}$[gas]]
a.	WL (coda delaryngealization)	[ga.sa]	[gaS] [a.kre]	[gaS]
b.	PL (defaults)	[ga.sa]	[ga.za.kre]	[gas]

As stated, the facts of Quito Spanish /s/-voicing pose a challenge to OO-correspondence (Colina 2006). This theory can explain the opaque voicing of onset /s/ in *gas acre* [ˌga.ˈza.kre] only by reference to a surface base containing a transparently voiced correspondent [z] in the coda. Many such expressions are found: e.g. *gas blanco* [ˌgaz.ˈβ̞laŋ.ko], *gas noble* [ˌgaz.ˈno.β̞le]. The problem, however, is that none of them bears a non-arbitrary morphosyntactic relationship to *gas acre* [ˌga.ˈza.kre], and so none can straightforwardly qualify as its base. If surface bases are selected by the containment criterion (Benua 1997: 28–29; Kager 1999: 215ff.), the only plausible option is the citation form *gas* [gas], which consists of a subset of the morphs of *gas acre*; but this exhibits [s]. In contrast, *gas noble* [ˌgaz.ˈno.βle], which contains the desired [z], has no better claim to being the base of *gas acre* than, say, *gas caro* [ˌgas.ˈka.ɾo], again showing [s]. Within inflectional paradigms, some versions of OO-correspondence allow surface bases to be designated by arbitrary stipulation (e.g. Kenstowicz 1996: 387, 391), but this option is of no avail here, since expressions like *gas*, *gas acre*, and *gas noble* do not belong in an inflectional paradigm; see the discussion of surface base selection in CHAPTER 83: PARADIGMS.

(25)



[7] Analyzing the pre-sonorant voicing of [S] as driven by a position-sensitive default (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION; CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS), rather than by feature spreading from a following sonorant redundantly specified as [+voice], allows for a closer fit between this categorical phonological operation and the gradient phonetic processes of passive voicing on which it is grounded and from which it diachronically emerges (see below): passive voicing in environments such as that occupied by the /s/ in *plasma* involves lengthening of the voicing tail from the preceding vowel, rather than anticipation of glottal pulsing for the following sonorant (Jansen 2004).

[8] This cyclic derivation accords partly with Mascaró's (1987) reduction-and-spreading model of laryngeal phenomena, though cf. note 7. Bermúdez-Otero (2007c: §31–§34) proposes a similar account for the voicing of word-final prevocalic sibilants in Catalan (cf. Wheeler 2005: 162–164). See also Rubach (1996: 72, 82–85) on the alleged voicing of all word-final obstruents before sonorants (including vowels) in Cracow Polish, but cf. Strycharczuk (2010).

However, Colina (2009: 8–10) shows that OO-correspondence can avoid this problem by shifting part of the burden of description onto the phonetics. Colina suggests that, in Quito Spanish, delaryngealized coda [S] does not acquire categorical voice specifications during the phonological derivation either by autosegmental spreading or by default feature insertion; she claims, rather, that expressions like *gas acre* and *gas noble* merely display the effects of gradient passive voicing in phonetic implementation (Keating 1988). If Colina is right, then the surface phonological representation of *gas acre* is [ˌga.ˈSa.kɾe], with over-application of delaryngealization in the onset; but this can be analyzed without difficulty as involving OO-correspondence with the citation form *ga*[S]: cf. (25) and (26).

(26) $_{NP}$[$_{N}$[gas]] $\qquad$ $_{NP}$[$_{N}$[gas] $_{A}$[akɾe]]

$\qquad$ IO-FAITH $\qquad\qquad$ IO-FAITH

$\qquad$ [gaS] ←-------------→ ☞ [ga.Sa.kɾe]

$\qquad\qquad$ OO-IDENT

Is Colina's re-analysis correct? This question cannot be settled on a priori grounds: in particular, the fact that the environment for /s/-voicing in *gas acre* straddles a word boundary does not by itself warrant the conclusion that the process must be gradient rather than categorical (CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY). Electropalatographic studies have admittedly shown that many instances of assimilatory external sandhi involve gradient co-articulation (i.e. reduction, overlap, and blending of articulatory gestures), rather than categorical assimilation (i.e. delinking and spreading of discrete phonological features): see e.g. Barry (1985), Wright and Kerswill (1989), Nolan (1992), Hardcastle (1995), and Zsiga (1995). However, there is also compelling evidence for the existence of categorical external sandhi. Holst and Nolan (1995) and Nolan *et al.* (1996) argue persuasively that at least some instances of /s#ʃ/ → [ʃː] sandhi in British English do involve discrete feature delinking and spreading; the likelihood of categorical assimilation increases in the absence of the major prosodic boundary associated with a break between clauses. Ladd and Scobbie (2003) report that, in Sardinian, total anticipatory assimilation between singletons across word boundaries yields long consonants that are phonetically equivalent to underlying geminates (CHAPTER 37: GEMINATES). Ellis and Hardcastle (2002) examined inter- and intra-speaker variation in fast-speech /n#k/ sandhi in British English, and found no fewer than four different idiolectal strategies (CHAPTER 92: VARIABILITY): (i) absence of accommodation between the two segments (in two out of ten subjects); (ii) gradient co-articulation (in two out of ten subjects); (iii) categorical assimilation (in four out of ten subjects); (iv) variation between categorical assimilation and absence of accommodation, with avoidance of co-articulation (in two out of ten subjects). Crucially, type (iv) speakers did not produce residual coronal gestures, but realized the nasal either without any tongue-tip raising at all or with full mid-sagittal linguo-alveolar closure; this behavior is inconsistent with gradient gestural reduction, but reflects the variable application of discrete feature delinking and spreading across word boundaries. Kochetov and Pouplier's (2008: 414) Korean subjects exhibited the same behavior in /t#p/ and /t#k/ sandhi. These findings clearly indicate that a process of external sandhi may apply gradiently for a speaker in some tokens, and still be categorical for other speakers, or for the same speaker in other tokens. It is therefore unsafe to relegate external sandhi to the phonetics without further argument.

Although instrumental evidence is lacking, Robinson (1979) and Lipski (1989) provide strong indirect evidence that, in many instances, word-final prevocalic /s/ does undergo categorical voicing in Quito Spanish. First, the process applies regularly in all registers independently of speech rate: it "may be frequently observed even in slow, disconnected or interrupted speech" (Lipski 1989: 53–54). Secondly, native speakers of the dialect rely on the difference between [s] and [z] to discriminate between minimal pairs like (23b): *ha sido* [a.'si.ð̞o] 'hath been' *vs.* *has ido* [a.'zi.ð̞o] 'hast gone' (Robinson 1979: 136, 140–141; Lipski 1989: 55). Thirdly, word-final /s/ voicing can be used as a turn-holding device before hesitation pauses (Robinson 1979: 141). Robinson records the following example, where he describes the realization of the /s/ of *es* as "strongly voiced":

(27) *es . . . tres . . .*
[ezːː ˈᵊtresː]
'it's . . . uh . . . three . . .' (Robinson 1979: 141)

It appears that the speaker intentionally produced a sandhi form of *es* to signal the fact that he or she had not reached the end of the utterance. Lipski (1989: 54) adduces further cases. For these reasons, Bradley and Delforge (2006: 39) conclude that the voicing of word-final prevocalic /s/ in Quito Spanish "reflects a phonological [+voice] specification," as opposed to "gradient interpolation of glottal activity through the constriction period of phonetically targetless [S]." This conclusion is incompatible with Colina's (2009) answer to the questions that Quito Spanish /s/-voicing raises for OO-correspondence.

The cyclic derivations proposed in (24) can moreover be seen as the synchronic outcome of a simple series of commonplace diachronic innovations (§3). We may assume that, in an initial round of phonologization and stabilization, the lack of robust phonetic cues for laryngeal features in codas was reinterpreted as phrase-level coda delaryngealization. Analogical change then caused this process of coda delaryngealization to percolate up to the word level. Finally, a second round of phonologization and stabilization caused the gradient passive voicing of delaryngealized sibilants in pre-sonorant contexts to be re-analyzed as a categorical phrase-level process of context-specific default feature insertion.

7 Phonologically masked bases II: English linking and intrusive *r*

Quito Spanish /s/-voicing is not an isolated case: it is not unusual for word-final prevocalic consonants to exhibit properties that are opaque in prevocalic position, but which nonetheless fail to match those of utterance-final consonants in citation forms. Linking and intrusive *r* in non-rhotic dialects of English provides another instance of this phenomenon. Again, a stratal-cyclic approach to the morphosyntax–phonology interface can easily deal with the facts, whereas OO-correspondence must shift some of the descriptive burden to a different component of the grammar: in this case, the theory of representations.

Most non-rhotic dialects of English (Wells 1982: 75–76, 218ff.) allow [ɹ] in onset positions, such as word-initially or word-medially before a stressed or unstressed vowel (28a),[9] but forbid [ɹ] in coda positions, such as word-medially or word-finally before a consonant or pause (28b). This is formalized as (28c).

(28) a. *rack* [ɹæk]
raccoon [ɹə.ˈkʰuːn]
carouse [kʰə.ˈɹaːʊz]
caramel [ˈkʰæ.ɹə.ˌmɛɫ]
b. *cart* [kʰɑːt] *[kʰɑːɹt]
car ‖ [kʰɑː‖] *[kʰɑːɹ‖]
the car came]ðə.ˌkʰɑː.ˈkʰeːɪm] *[ðə.ˌkʰɑːɹ.ˈkʰeːɪm]
c. *CODA[ɹ]
*Coda
|
ɹ

Crucially, most non-rhotic dialects tolerate [ɹ] word-finally before a vowel, whether the consonant was present etymologically ("linking *r*") or not ("intrusive *r*").

(29) a. *the car is new* [ðə.ˌkʰɑː.ɹ̞ɪz.ˈnjuː] linking *r*
b. *the spa is new* [ðə.ˌspɑː.ɹ̞ɪz.ˈnjuː] intrusive *r*

The fact that linking and intrusive *r* escapes the phonotactic ban in (28c) indicates that it surfaces in the onset (CHAPTER 55: ONSETS). In English, however, word-final prevocalic *r* (including linking and intrusive *r* in non-rhotic dialects) exhibits lenition in comparison with canonical onset *r*; the transcriptions above reflected this phenomenon by distinguishing between unlenited [ɹ] in (28a) and lenited [ɹ̞] in (29). Compared with word-initial [ɹ], word-final prevocalic [ɹ̞] displays the following properties: (i) shorter duration (Cruttenden 2001: 289; Tuinman *et al.* 2007: 1905–1906); (ii) earlier timing of the tongue-root gesture (Campbell *et al.* 2010: 62); (iii) smaller magnitude of the lip gesture (Wells 1990; Campbell *et al.* 2010: 63–64); (iv) smaller magnitude of the tongue-tip gesture (Gick 1999: 47–49; Campbell *et al.* 2010: 63–64); (v) greater magnitude of the tongue-root gesture (Campbell *et al.* 2010: 63–64); (vi) greater intensity (McCarthy 1993: 179; Tuinman *et al.* 2007: 1905–1906); and (vii) higher F3 (Hay and Maclagan 2010). Thus dialects with intrusive *r* afford minimal pairs such as the following (McCarthy 1993: 179):

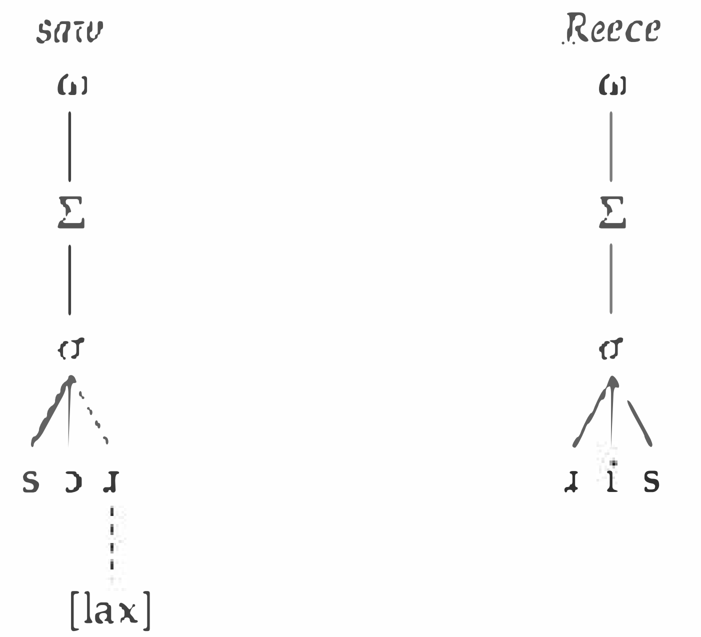
(30) a. *saw eels* [sɔː.ɹ̞iːɫz]
b. *saw reels* [sɔː.ɹiːɫz]

If, as I have suggested, linking and intrusive [ɹ̞] escapes the phonotactic restriction in (28c) because it surfaces in onset position, just like word-initial [ɹ], then the reasons why the former undergoes lenition and the latter does not are not apparent on the surface: thus *r*-lenition overapplies (CHAPTER 74: RULE ORDERING).

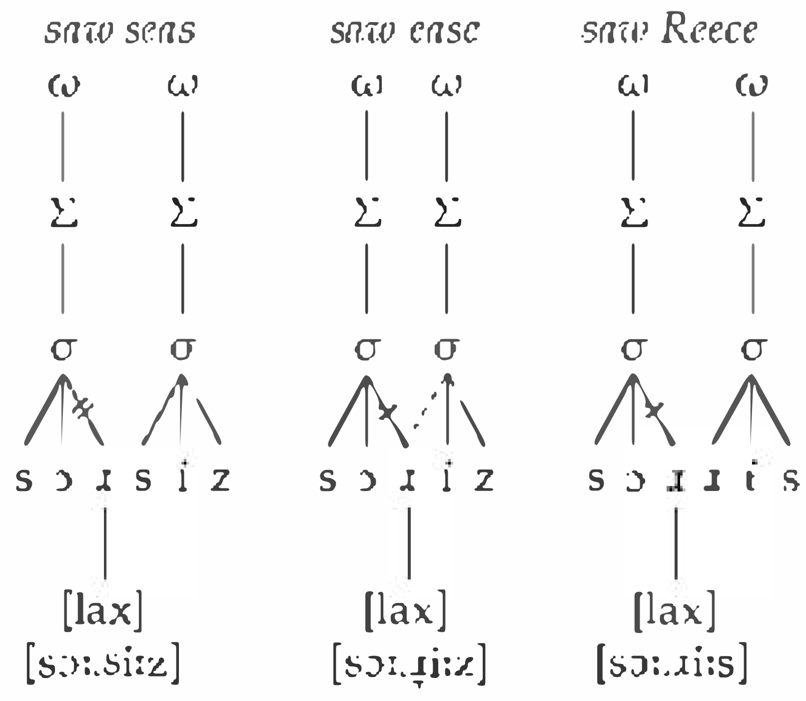
[9] Harris (2006) reports that in some Southern US dialects [ɹ] is banned outside foot-initial onsets: e.g. *ˈve~~r~~y, ˈshe~~r~~iff, ˌCa~~r~~oˈlina*.

However, this opaque pattern is easy to describe in stratal-cyclic terms (Kiparsky 1979: 437ff.; McCarthy 1991: 203–204). Intrusive *r* is inserted at the word level after ω-final non-high vowels in order to satisfy the constraint FINALC, i.e. *V]ω (McCarthy 1991: 203, 1993: 176), which outranks *CODA[ɹ] at the word level.[10] In the same cycle, the inserted *r* is targeted by coda lenition and undergoes a corresponding featural change: it acquires the feature [lax], say. At the phrase level, however, the relative ranking of FINALC and *CODA[ɹ] is reversed: in consequence, word-final *r* undergoes deletion in preconsonantal and prepausal environments, but in prevocalic position it escapes into the onset, carrying with it the feature [lax].

(31) a. *Word level:* FINALC >> *CODA[ɹ]



b. *Phrase level:* *CODA[ɹ] >> FINALC



10 The idea that *r*-intrusion is driven by FINALC receives independent support from the absence of intrusive *r* after reduced function words (which do not project an ω-node) in the non-rhotic dialect of Eastern Massachusetts (McCarthy 1991: 200ff., 1993: 173ff.). In the case of words ending with high vowels or closing diphthongs, we assume that *r*-intrusion is blocked by the final offglide, which suffices to satisfy FINALC; alternatively, FINALC can be replaced with *V$_{[-high]}$]$_{\omega}$. If *r*-intrusion applies ω-finally, then stem-level applications may be needed to generate forms like *draw*[ɹ]-*ing*, as word-level suffixes like -*ing* are incorporated into the prosodic word of the stem, and not adjoined: i.e. $_{\omega}$[*draw*] → $_{\omega}$[*drawing*], not $_{\omega}$[*draw*] → *$_{wd}$[$_{\omega}$[*draw*]*ing*]; see §4 above. If so, we may assume that listed allomorphy pre-empts stem-level *r*-intrusion in cases like *algebr*[ə] ~ *algebr*[eɪ]*ic* (McCarthy 1991: 196).

As in the case of Quito Spanish /s/-voicing (§6), this synchronic system can easily be understood as the product of a straightforward series of ordinary phonological changes:

(32)	level	processes	*manner*	*manner is*	*Anna*	*Anna is*
a.	*Initial stage*					
	WL		[mæ.nəɹ]	[mæ.nəɹ][ɪz]	[æ.nə]	[æ.nə][ɪz]
	PL		[mæ.nəɹ]	[mæ.nə.ɹɪz]	[æ.nə]	[æ.nə.ɪz]
b.	*Phonologization and stabilization (I): Lenition of* [ɹ] *in codas enters the phrase level*					
	WL		[mæ.nəɹ]	[mæ.nəɹ][ɪz]	[æ.nə]	[æ.nə][ɪz]
	PL	lenition	[mæ.nəɹ̞]	[mæ.nə.ɹɪz]	[æ.nə]	[æ.nə.ɪz]
c.	*Analogical input restructuring (I): Lenition of* [ɹ] *in codas climbs up to the word level*					
	WL	lenition	[mæ.nəɹ̞]	[mæ.nəɹ̞][ɪz]	[æ.nə]	[æ.nə][ɪz]
	PL	lenition (vacuous)	[mæ.nəɹ̞]	[mæ.nə.ɹ̞ɪz]	[æ.nə]	[æ.nə.ɪz]
d.	*Phonologization and stabilization (II): Deletion of* [ɹ̞] *in codas enters the phrase level*					
	WL	lenition	[mæ.nəɹ̞]	[mæ.nəɹ̞][ɪz]	[æ.nə]	[æ.nə][ɪz]
	PL	deletion	[mæ.nə]	[mæ.nə.ɹ̞ɪz]	[æ.nə]	[æ.nə.ɪz]
e.	*Analogical input restructuring (II): Analogical extension of word-level final* [ɹ̞]					
	WL	insertion, lenition	[mæ.nəɹ̞]	[mæ.nəɹ̞][ɪz]	[æ.nəɹ̞]	[æ.nəɹ̞][ɪz]
	PL	deletion	[mæ.nə]	[mæ.nə.ɹ̞ɪz]	[æ.nə]	[æ.nə.ɹ̞ɪz]

The path for (32e) was smoothed by a general process of schwa apocope (CHAPTER 26: SCHWA) in Middle English (Minkova 1991). As a result of this, Early Modern English had relatively few words like *Anna*, with an underlying final /ə/. Thus, the rise of phrase-level *r*-deletion in codas brought about a situation in which most tokens of preconsonantal or prepausal [ə] alternated with prevocalic [əɹ̞]. In these circumstances, learners re-analyzed phrase-level representations like [æ.nə||] as derived by *r*-deletion from word-level [æ.nəɹ̞].[11]

(33) WL [mæ.nəɹ̞] [æ.nəɹ̞]

PL [mæ.nə||] [mæ.nə.ɹ̞ɪz] → [æ.nə||] [æ.nə.ɹ̞ɪz]

In turn, this analogical extension of final [ɹ̞] across word-level outputs eventually resulted in a word-level ban of ω-final [ə], enforced where necessary by [ɹ̞]-insertion.

This stratal-cyclic account of the diachronic rise and synchronic operation of *r*-intrusion avoids many of the pitfalls incurred by its best-known competitors. Rule-inversion scenarios resulting in a phrase-level hiatus-breaking rule of [ɹ]-epenthesis in onsets (e.g. Vennemann 1972: 216; McMahon 2000: ch. 6; Bermúdez-Otero and Hogg 2003: 99ff.) do not account for the lenited realization of intrusive *r* (CHAPTER 66: LENITION). In turn, restructuring scenarios in which /ə/ is replaced

[11] In this view, the analogical extension of final [ɹ̞] across word-level outputs can only have begun after variable *r*-deletion entered the phrase level, but it may well have been in progress before the application rate of *r*-deletion approached 100 percent (see Hay and Sudbury 2005).

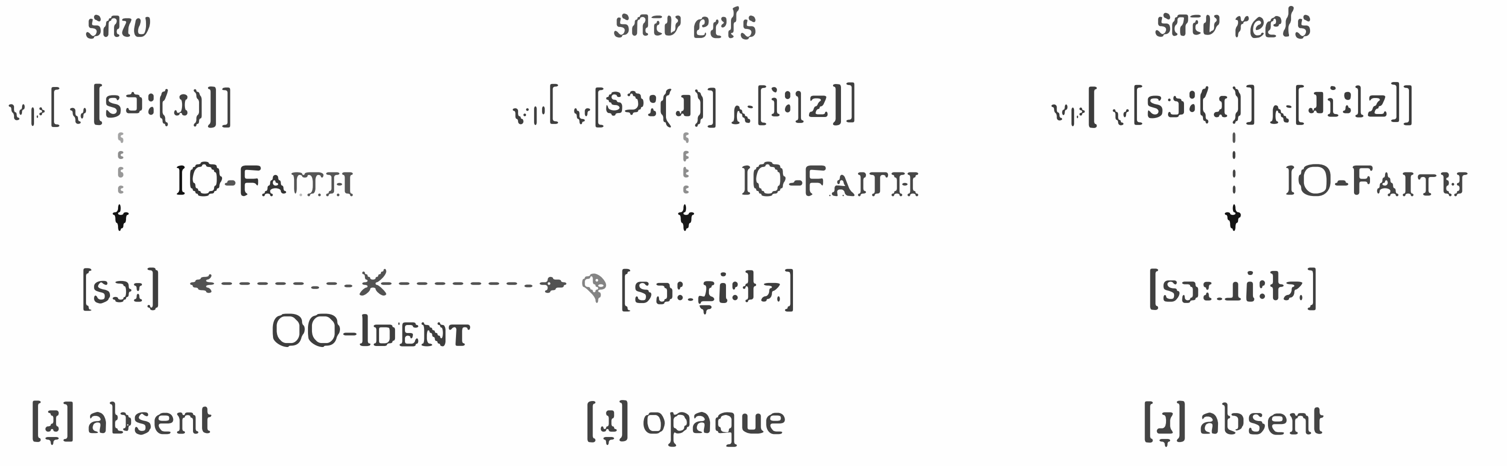
by /əɹ/ in underlying representations (e.g. Donegan 1993) fail to account for the regular and productive nature of *r*-intrusion (see the references in Heselwood 2009: 86). A regular process of [ɹ]-epenthesis in ω-final position at the word level incurs neither problem.

Furthermore, the diachronic scenario outlined in (32) accords perfectly with the normal life cycle of phonological processes (§3). Both *r*-lenition ([ɹ] → [ɹ̞]) and *r*-deletion ([ɹ̞] → Ø) first entered the categorical phonology from below, as phrase-level processes applying across the board ((32b) and (32d)). The analogical change causing lenition to climb up from the phrase to the word level (32c) proceeds by input restructuring: the lenited [ɹ̞] in surface $_{PL}$[mæ.nəɹ̞] is re-analyzed as present in the output of the word level.[12] Moreover, *r*-lenition entered the grammar earlier than *r*-deletion (as must be the case, since the former is a precondition for the latter), and so has been exposed to analogical pressures for longer: it is therefore unsurprising that *r*-lenition should be more advanced in its life cycle than *r*-deletion, the former having reached the word level, the latter remaining at the phrase level.

In this light, the synchronic markedness reversal illustrated in (31) can be seen as arising from a clash between disparate diachronic forces: the high ranking of *Coda[ɹ] at the phrase level reflects the phonologization of phonetic effects; in contrast, the high ranking of FinalC at the word level reflects the analogical restructuring of phrase-level inputs. If so, McCarthy's (1993: 181–182) complaint of arbitrariness against his own previous stratal analysis (McCarthy 1991: 203–204) arguably betrays a failure to strike a proper balance between synchronic and diachronic explanation (cf. Bermúdez-Otero 1999: 98–107).

If this account is correct, then English linking and intrusive *r* raises difficulties for OO-correspondence. The segment's lenited realization is opaque because there is no *r*-lenition in onsets. To explain the facts, OO-correspondence would need to find a surface base in which [ɹ̞] occurred transparently, i.e. in the coda. Yet this is impossible, as the defining property of non-rhotic dialects is precisely that they do not allow *r* to surface outside the onset.

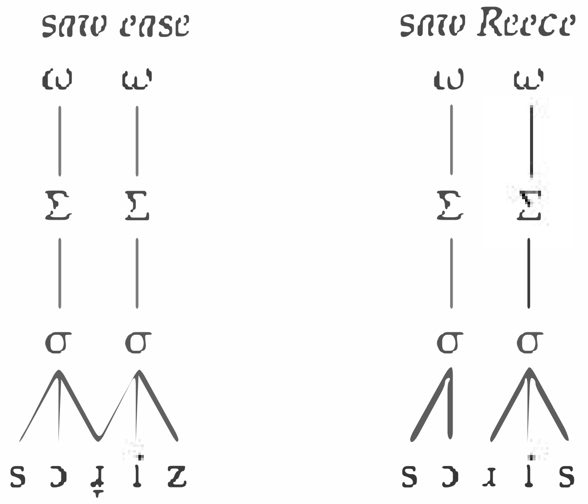
(34)



Yet, as in the case of Quito Spanish /s/-voicing, the proponents of OO-correspondence may deflect this argument by putting forward a transparent analysis of linking and intrusive [ɹ̞]. McCarthy (1993: 178–181) does so by invoking ambisyllabicity (Kahn 1976). In this approach, linking and intrusive [ɹ̞] is

12 This progression from lower to higher levels correctly predicts that, diachronically, word-internal *r*-intrusion, as in *draw-ing* (see note 10), starts later than *r*-intrusion at word boundaries, as in *draw in* (see Hay and Sudbury 2005: 816–818, 820).

permitted to surface because it has an onset attachment, but it is lenited because it has a link to the coda too: cf. (31b) and (35).

(35)


In this sense, ambisyllabicity enables McCarthy (1993: 178–181) to conflate two stages of a cyclic derivation into a single representation – at the cost of adopting a less restrictive theory of syllable structure.

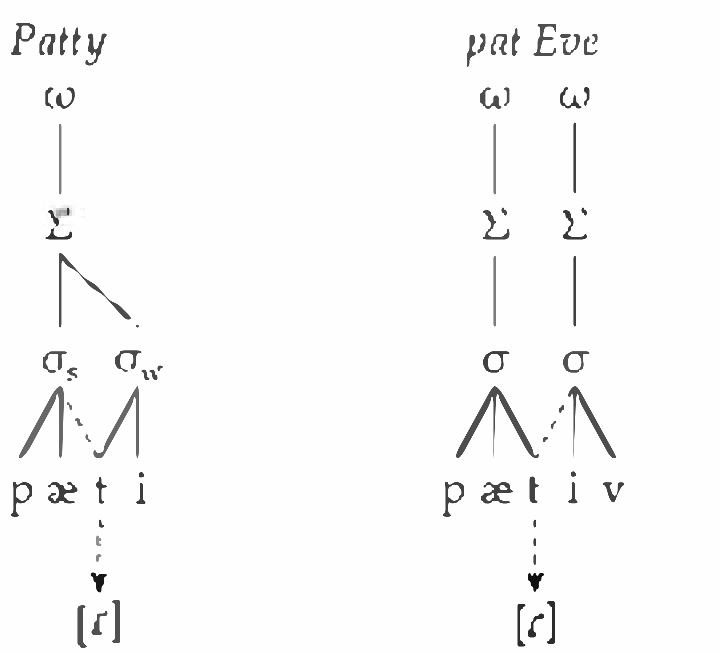
However, ambisyllabicity incurs problems of its own, and has been argued to provide an inconsistent account of English segmental allophony (e.g. Kiparsky 1979: 437ff.; Jensen 2000; Harris 2003). Bermúdez-Otero (2007a: §14–§24) notes two ambisyllabicity paradoxes. Since Kahn (1976), the standard diagnostic for ambisyllabification in English has been /t/-flapping. In most North American dialects, /t/ undergoes flapping in two environments: foot-medially between a vowel or /ɹ/ and another vowel (36a), and word-finally between a vowel or /ɹ/ and another vowel (36b).

(36) a. $_{\Sigma}[\ldots \begin{Bmatrix} \text{V} \\ \text{ɹ} \end{Bmatrix} \_\_ \text{V} \ldots]$ e.g. [ɾ] in *Patty, party, parity*

b. $\begin{Bmatrix} \text{V} \\ \text{ɹ} \end{Bmatrix} \_\_]_{\text{GWord}}\ \text{V}$ e.g. [ɾ] in *pat it, pat Eve, at it, at ease*

Since the segmental conditions in these two environments are exactly identical, formulating two separate rules of flapping would miss a generalization. Accordingly, Kahn proposed that the two environments could be unified prosodically: in (36a) /t/ becomes ambisyllabic by Coda Capture, and in (37b) /t/ becomes ambisyllabic by Onset Capture.

(37) a. *Coda Capture* b. *Onset Capture*



Thus Kahn's strategy was to use syllabification to channel the allophonic effects of both stress and word boundaries. Yet this solution does not generalize to other English consonants. Consider, for example, /l/-darkening in the Midwestern American dialect studied by Sproat and Fujimura (1993). This dialect exhibits Kahn's canonical pattern of /t/-flapping. By implication, /l/ too should display the same allophone, either clear [l] or dark [ɫ], in foot-medial intervocalic position (e.g. *Bee/l/ik*) and in word-final intervocalic position (e.g. *Bee/l/ equates*): /l/ should be ambisyllabic in the former by Coda Capture and in the latter by Onset Capture. As Sproat and Fujimura (1993: 308) themselves note in passing, however, this prediction proves false: X-ray microbeam cinematography revealed that their subjects produced clear [l], with the coronal gesture phased before the dorsal gesture, in *Bee/l/ik*, whereas they produced dark [ɫ], with the dorsal gesture phased before the coronal gesture, in *Bee/l/ equates*.

(38)	*form*	*/l/ allegedly ambisyllabic by . . .*	*/l/ allophone*	*the coronal gesture . . .*
	Beelik	Coda Capture	clear [l]	leads
	Beel equates	Onset Capture	dark [ɫ]	lags

In this dialect, therefore, Kahn's ambisyllabification rules work for /t/, but not for /l/. This is Bermúdez-Otero's (2007a) first ambisyllabicity paradox. A second paradox arises from a conflict between /t/-flapping and pre-fortis clipping (Bermúdez-Otero 2007a: §21–§24), and Kiparsky (1979: 440) observes a third paradox, further discussed by Nespor and Vogel (1986: 93–94). By casting doubt on the existence of ambisyllabicity, these paradoxes challenge McCarthy's (1993) transparent re-analysis of linking and intrusive *r* in (35).

In contrast, the English dialect described by Sproat and Fujimura poses no difficulties for a stratal-cyclic model with onset-maximal stem-level syllabification and resyllabification of prevocalic consonants in word-level and phrase-level cycles (Bermúdez-Otero 2007a: §18–§20). The right results follow from the operation of two word-level processes: one laxes /t/ in non-foot-initial position (Kiparsky 1979: 437ff.; Jensen 2000; Harris 2003); the other darkens /l/ in the coda. A full typology of English dialects supports the need to allow individual allophonic processes to target either weak positions in the syllable (i.e. the coda) or weak positions in the foot (i.e. in a trochaic system, anywhere outside foot-initial onsets). Notably, an innovative pattern of foot-based /l/-darkening (e.g. *ye[ɫ]ow*, *vi[ɫ]age*) is attested alongside the conservative syllable-based pattern: see Olive *et al.* (1993: 366) and Hayes (2000: 95–96) for American dialects, and Carter and Local (2007) for British dialects.[13]

In sum, English linking and intrusive *r* raises the same problem for OO-correspondence as Quito Spanish /s/-voicing: both are patterns of external sandhi in which word-final prevocalic consonants display opaquely derived properties that are absent from citation forms. In both cases, OO-correspondence declines responsibility for the facts, and shifts the burden of explanation either to phonetic implementation or to the theory of representations.

[13] Similarly, alongside the conservative pattern of syllable-based *r*-deletion in non-rhotic dialects, an innovative foot-based pattern has been detected in the south of the USA: see note 9 above.

8 Non-surfacing bases in non-canonical paradigms: Albanian stress

In the examples of morphosyntactically induced misapplication discussed in §6 and §7, the surface bases required by OO-correspondence are unavailable for phonological reasons: a phonological process applies normally in a non-final cycle C, but the output of C never surfaces transparently, because it is always altered by the operation of subsequent phonological processes in later cycles. However, the output of C may also fail to surface unchanged, for purely morphological reasons. This effect stands out with particular clarity in non-canonical inflectional paradigms, i.e. paradigms exhibiting phenomena such as deponency, defectiveness, suppletion, or heteroclisis (Corbett 2007). In such circumstances, the predictions of cyclicity and OO-correspondence diverge dramatically. Let two words *a* and *b* have identical syntagmatic structures in all relevant respects, but belong to paradigms with different sets of cells: one canonical, the other non-canonical. The theory of the cycle predicts that, in the phonology, *a* and *b* must exhibit the same effects of procedural morphosyntactic conditioning (§1), since the course of cyclic derivations depends on syntagmatic structure alone (Bobaljik 2008: 32; Bailyn and Nevins 2008: 242). In contrast, OO-correspondence predicts the opposite, as transderivational identity effects depend on the availability of surface bases. On the basis of evidence from Albanian, Trommer (2006, 2009) argues that the first prediction is true, the second false. In this section I briefly summarize Trommer's argument, omitting his detailed motivation of the morphological segmentations underpinning the analysis.

Trommer (2004) found that Albanian polysyllabic words bearing no overt inflection display final stress in either of two cases: (i) if the final syllable is headed by a non-mid vowel (i.e. by /i/, /u/, or /a/), as in (39a) and (39b), or (ii) if the final syllable is both headed by a full vowel (i.e. by a vowel other than /ə/) and closed by a consonant, as in (39b) and (39c)). Otherwise, stress falls on the penultima, as in (39d) and (39e).

(39)	a.	[ɟu.hɔ.'si]	'linguistics'
		[a.kə.'ku]	'here and there'
		[ɾi.'ʤa]	'prayer'
	b.	[aɾ.'mik]	'enemy'
		[ʧi.'fut]	'gipsy'
		[ɾe.zul.'tat]	'result'
	c.	[a.'det]	'habit'
		[pa.'tok]	'gander'
	d.	['ho.le]	'swing'
		['ba.bo]	'midwife'
		['hɔ.nɔ]	'moon'
	e.	['a.fəɾ]	'near'

In word-forms containing overt inflectional markers, however, stress assignment often misapplies. Consider, for example, the present indicative of a verb with a canonical paradigm (Table 85.2).

According to Trommer, metrical opacity arises as a consequence of the fact that the domain of stress assignment is the stem, not the word: stress is assigned transparently in stem-level cycles, but is rendered opaque at the word level

Table 85.2 The present indicative of the Albanian verb *formoj* 'form' (NACT denotes 'non-active')

			UR	*SR*	*opaque stress?*
ACT	SG	1	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-j]]	[for.'moj]	no
		2	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-n]]	[for.'mon]	no
		3	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-n]]	[for.'mon]	no
	PL	1	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-j] $_{\text{Affix}}$[mə]]	[for.'moj.mə]	no
		2	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-n] $_{\text{Affix}}$[ni]]	[for.'mo.ni]	yes: *[for.mo.'ni]
		3	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-j] $_{\text{Affix}}$[nə]]	[for.'moj.nə]	no
NACT	SG	1	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-j] $_{\text{Affix}}$[he-m]]	[for.'mo.hem]	yes: *[for.mo.'hem]
		2	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-j] $_{\text{Affix}}$[he-ʃ]]	[for.'mo.heʃ]	yes: *[for.mo.'heʃ]
		3	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-j] $_{\text{Affix}}$[he-t]]	[for.'mo.het]	yes: *[for.mo.'het]
	PL	1	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-j] $_{\text{Affix}}$[he-mi]]	[for.'mo.he.mi]	yes: *[for.mo.he.'mi]
		2	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-j] $_{\text{Affix}}$[he-ni]]	[for.'mo.he.ni]	yes: *[for.mo.he.'ni]
		3	$_{\text{GWord}}$[$_{\text{Stem}}$[formo-j] $_{\text{Affix}}$[he-n]]	[for.'mo.hen]	yes: *[for.mo.'hen]

by the addition of inflectional suffixes and by regular internal sandhi at the stem–suffix juncture.

(40) a. *Internal sandhi processes*
nn → n
j → Ø / __ h

b. *Sample derivations*

	$_{\text{WL}}$[$_{\text{SL}}$[formo-j]]	$_{\text{WL}}$[$_{\text{SL}}$[formo-j] $_{\text{SL}}$[he-m]]
SL (stress assignment)	[for.'moj]	[for.'moj] [hem]
WL (internal sandhi)	—	[for.'mo.hem]
	'form (ACT 1SG)'	'form (NACT 1SG)'

Let us now turn to verbs with non-canonical paradigms. The verb *pendohem* 'regret', for example, exhibits deponency: it lacks a voice alternation, and its fixed lexical meaning is expressed by a series of non-active forms (Table 85.3). Crucially, the absence of non-active forms entails that the location of stress is opaque throughout the present indicative.

This fits with the predictions of cyclicity: since the single series of forms of a deponent verb has the same syntagmatic structure as the non-active series of a canonical verb, both must display the same pattern of metrical opacity; compare (40b) and (41).

(41)

	$_{\text{WL}}$[$_{\text{SL}}$[pendo-j] $_{\text{SL}}$[he-m]]
SL (stress assignment)	[pen.'doj] [hem]
WL (internal sandhi)	[pen.'do.hem]
	'regret (1SG)'

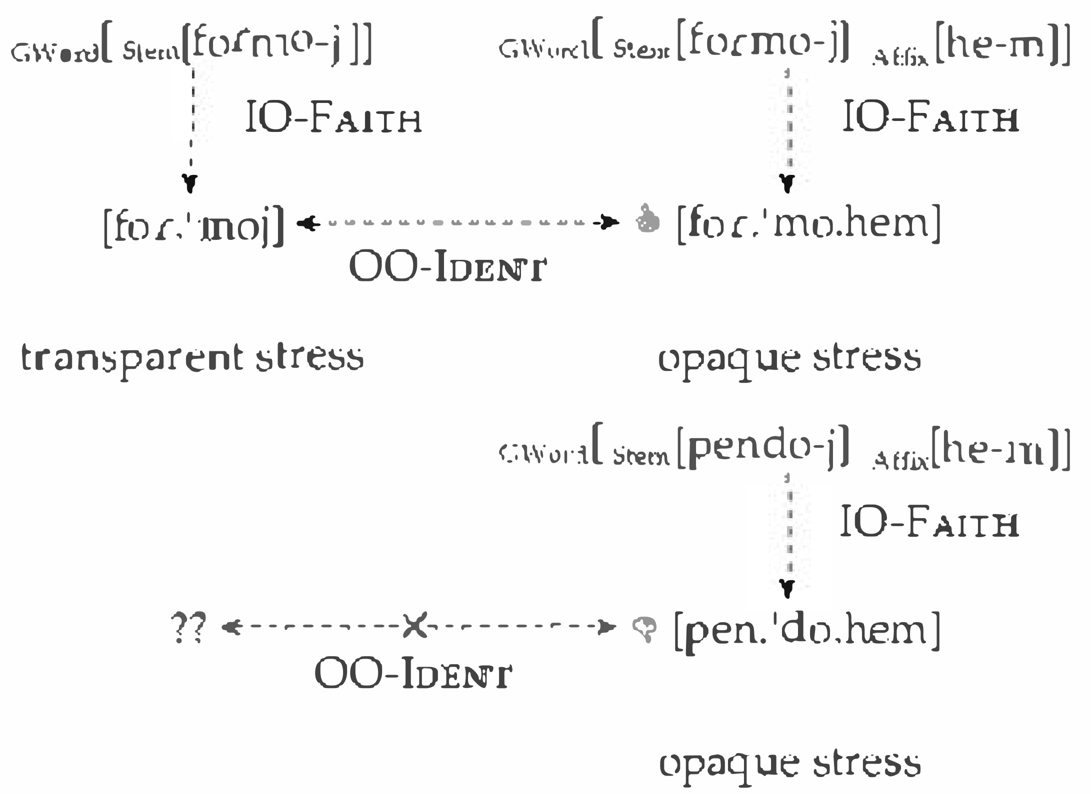
In contrast, OO-correspondence seems unable to account for the misapplication of stress assignment in the present indicative forms of Albanian deponent verbs: there are simply no suitable surface bases with transparent stress.

Table 85.3 The present indicative of the Albanian verb *pendohem* 'regret'

			UR	*SR*	*opaque stress?*
ACT			—	—	—
NACT	SG	1	$_{\text{GWord}}$[$_{\text{Stem}}$[pendo-j] $_{\text{Affix}}$[he-m]]	[pen.'do.hem]	yes: *[pen.do.'hem]
		2	$_{\text{GWord}}$[$_{\text{Stem}}$[pendo-j] $_{\text{Affix}}$[he-ʃ]]	[pen.'do.heʃ]	yes: *[pen.do.'heʃ]
		3	$_{\text{GWord}}$[$_{\text{Stem}}$[pendo-j] $_{\text{Affix}}$[he-t]]	[pen.'do.het]	yes: *[pen.do.'het]
	PL	1	$_{\text{GWord}}$[$_{\text{Stem}}$[pendo-j] $_{\text{Affix}}$[he-mi]]	[pen.'do.he.mi]	yes: *[pen.do.he.'mi]
		2	$_{\text{GWord}}$[$_{\text{Stem}}$[pendo-j] $_{\text{Affix}}$[he-ni]]	[pen.'do.he.ni]	yes: *[pen.do.he.'ni]
		3	$_{\text{GWord}}$[$_{\text{Stem}}$[pendo-j] $_{\text{Affix}}$[he-n]]	[pen.'do.hen]	yes: *[pen.do.'hen]

(42) a.

b.



Thus Trommer's analysis suggests that morphologically induced misapplication depends on syntagmatic structure, not on the contents of paradigms.

In the case of Quito Spanish /s/-voicing, the advocates of OO-correspondence shifted the burden of explanation to phonetics (§6); in the case of English linking and intrusive *r*, to the theory of representations (§7). A similar escape maneuver in the case of Albanian stress might conceivably appeal to morphology, e.g. by claiming that stress assignment in Albanian verbs has been partly or wholly morphologized. Whatever the merits of such an argument, OO-correspondence will remain in an anomalous position until enough languages are found in which systematic patterns of morphologically induced phonological misapplication fail to hold in defective, deponent, suppletive, and heteroclitic paradigms.

9 Further challenges to OO-correspondence

The case studies presented in §6–§8 provide the most direct challenge to the theory of OO-correspondence: in all three cases, the necessary surface bases appear to be unavailable. However, transderivational theories face other questions, briefly noted in §5: what expressions can qualify as surface bases, and how are they selected?; should OO-identity be symmetrical, base-prioritizing, or both?

(See CHAPTER 83: PARADIGMS.) Whilst these problems have attracted a great deal of attention in the literature, the fact that OO-correspondence fails to preserve much of the corroborated empirical content of cyclic theory has generally prompted less discussion. One key instance is the Russian Doll Theorem (§3).

For example, (43) reports the incidence of /l/-darkening in three English dialects where the process has not yet become foot-based (i.e. where /l/ remains light in *village*: see §7).[14] From this evidence one can reliably infer a pattern of diachronic evolution instantiating the Russian Doll Theorem: (43) is a perfect match for (12) and (13).

(43)

	RP	Am1	Am2
Healey	l	l	l
heal-ing	l	l	ɫ
heal it	l	ɫ	ɫ
heal	ɫ	ɫ	ɫ
darkening of rhymal /l/ applies at . . .	PL	WL	SL

conservative ⟶ innovative

In a transderivational analysis, however, the inevitability of the Russian Doll pattern disappears. For example, Hayes (2000: 102) proposes two separate OO-identity constraints to capture the facts in (43): Am1 shows an effect of high-ranking OO-IDENT(PHRASAL); Am2 shows effects of both OO-IDENT(PHRASAL) and OO-IDENT(MORPHOLOGICAL).

(44) a. hiːɫ ⇢ hiː.ɫɪt
OO-IDENT(PHRASAL)

b. hiːɫ ⇢ hiː.ɫɪŋ
OO-IDENT(MORPHOLOGICAL)

By factorial typology, however, these two constraints can generate an impossible dialect with [hiːɫ, hiː.ɫɪŋ, hiː.lɪt], in violation of the Russian Doll Theorem. All that is needed is a constraint hierarchy of the following type:

(45) *CODA[l] >> OO-IDENT(MORPHOLOGICAL) >> *[ɫ] >> OO-IDENT(PHRASAL)

To avoid this result, Hayes (2000: 102) resorts to stipulating an innate fixed ranking in Universal Grammar:

(46) OO-IDENT(PHRASAL) >> OO-IDENT(MORPHOLOGICAL)

The explanatory loss is plain to see: whereas (11) was a corollary, (46) is an axiom.

[14] For RP, see Cruttenden (2001: 201). The dialect I have here labeled "Am1" is the one described by Sproat and Fujimura (1993); see §7 above. For "Am2," see Olive *et al.* (1993: 212–215). The implicational relationships implicit in (43) are confirmed by the rates of variation reported by Hayes (2000: 98): see (16) above.

It thus looks as if a great deal of work remains to be done before OO-correspondence can claim to have superseded the cycle.

ACKNOWLEDGMENTS

In this chapter I have drawn on research previously presented at meetings in Groningen, Leipzig, Manchester, Rhodes, Toulouse, and Warsaw; I am grateful to the organizers and audiences on all these occasions for their comments and suggestions. I am also indebted to Sonia Colina, Andrew Nevins, Tobias Scheer, Patrycja Strycharczuk, and Jochen Trommer.

REFERENCES

Bachrach, Asaf & Andrew Nevins (eds.) 2008. *Inflectional identity*. Oxford: Oxford University Press.

Bailyn, John F. & Andrew Nevins. 2008. Russian genitive plurals are impostors. In Bachrach & Nevins (2008), 237–270.

Baker, Brett. 2005. The domain of phonological processes. In Ilana Mushin (ed.) *Proceedings of the 2004 Conference of the Australian Linguistics Society*. Sydney: Department of Linguistics, University of Sydney. Available (August 2010) at http://hdl.handle.net/2123/112.

Barry, Martin C. 1985. A palatographic study of connected speech processes. *Cambridge Papers in Phonetics and Experimental Linguistics* 4. 1–16.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Ph.D. dissertation, University of Massachusetts, Amherst (ROA-259).

Bermúdez-Otero, Ricardo. 1999. Constraint interaction in language change: Quantity in English and Germanic. Ph.D. dissertation, University of Manchester & University of Santiago de Compostela.

Bermúdez-Otero, Ricardo. 2003. The acquisition of phonological opacity. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.) *Variation within Optimality Theory: Proceedings of the Stockholm Workshop*, 25–36. Stockholm: Department of Linguistics, Stockholm University.

Bermúdez-Otero, Ricardo. 2006. Phonological change in Optimality Theory. In Keith Brown (ed.) *Encyclopedia of language and linguistics*, 2nd edn, vol. 9, 497–505. Oxford: Elsevier.

Bermúdez-Otero, Ricardo. 2007a. Word-final prevocalic consonants in English: Representation vs derivation. Paper presented at the Old World Conference in Phonology 4, Rhodes. Handout available (August 2010) at www.bermudez-otero.com/OCP4.pdf.

Bermúdez-Otero, Ricardo. 2007b. Diachronic phonology. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 497–517. Cambridge: Cambridge University Press.

Bermúdez-Otero, Ricardo. 2007c. Marked phonemes vs marked allophones: segment evaluation in Stratal OT. Paper presented at the workshop on segment inventories, 30th GLOW Colloquium, Tromsø. Available (August 2010) at www.bermudez-otero.com/GLOW2007.pdf.

Bermúdez-Otero, Ricardo. 2007d. Morphological structure and phonological domains in Spanish denominal derivation. In Fernando Martínez-Gil & Sonia Colina (eds.) *Optimality-theoretic studies in Spanish phonology*, 278–311. Amsterdam & Philadelphia: John Benjamins.

Bermúdez-Otero, Ricardo. 2008. [ðə swɪˈŋɒmɪtə ˈtɜːnd əˈgenst sə ˈmɪŋɪs ˈkæmbəl]: Evidence for Chung's generalization. Paper presented at the 16th Manchester Phonology Meeting. Handout available (August 2010) at http://www.bermudez-otero.com/16mfm.pdf.

Bermúdez-Otero, Ricardo & Richard M. Hogg. 2003. The actuation problem in Optimality Theory: Phonologization, rule inversion, and rule loss. In D. Eric Holt (ed.) *Optimality Theory and language change*, 91–119. Dordrecht: Kluwer.

Bermúdez-Otero, Ricardo & Ana R. Luís. 2009. Cyclic domains and prosodic spans in the phonology of European Portuguese functional morphs. Paper presented at the Workshop on the Division of Labour between Morphology and Phonology and 4th Meeting of the Network Core Mechanisms of Exponence, Meertens Institute, Amsterdam. Available (August 2010) at www.bermudez-otero.com/bermudez-otero&luis.pdf.

Bermúdez-Otero, Ricardo & April McMahon. 2006. English phonology and morphology. In Bas Aarts & April McMahon (eds.) *The handbook of English linguistics*, 382–410. Cambridge, MA & Oxford: Blackwell.

Bobaljik, Jonathan D. 2008. Paradigms (Optimal and otherwise): A case for skepticism. In Bachrach & Nevins (2008), 29–54.

Booij, Geert & Jerzy Rubach. 1984. Morphological and prosodic domains in Lexical Phonology. *Phonology Yearbook* 1. 1–27.

Borowsky, Toni. 1993. On the word level. In Hargus & Kaisse (1993), 199–234.

Bradley, Travis G. & Ann Marie Delforge. 2006. Systemic contrast and the diachrony of Spanish sibilant voicing. In Randall Gess & Deborah Arteaga (eds.) *Historical Romance linguistics: Retrospective and perspectives*, 19–52. Amsterdam & Philadelphia: John Benjamins.

Campbell, Fiona, Bryan Gick, Ian Wilson & Eric Vatikiotis-Bateson. 2010. Spatial and temporal properties of gestures in North American English /r/. *Language and Speech* 53. 49–69.

Carter, Paul & John Local. 2007. F2 variation in Newcastle and Leeds English liquid systems. *Journal of the International Phonetic Association* 37. 183–199.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.) *Ken Hale: A life in language*, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Chomsky, Noam, Morris Halle & Fred Lukoff. 1956. On accent and juncture in English. In Morris Halle, Horace G. Lunt, Hugh McLean & Cornelis H. van Schooneveld (eds.) *For Roman Jakobson: Essays on the occasion of his sixtieth birthday*, 65–80. The Hague: Mouton.

Colina, Sonia. 2006. Voicing assimilation in Ecuadoran Spanish: Evidence for Stratal OT. Paper presented at 36th Linguistic Symposium on Romance Languages, Rutgers University.

Colina, Sonia. 2009. Sibilant voicing in Ecuadorian Spanish. *Studies in Hispanic and Lusophone Linguistics* 2. 1–18.

Connell, Bruce & Amalia Arvaniti (eds.) 1995. *Papers in laboratory phonology IV: Phonology and phonetic evidence*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2007. Deponency, syncretism, and what lies between. In Matthew Baerman, Greville G. Corbett, Dunstan Brown & Andrew Hippisley (eds.) *Deponency and morphological mismatches*, 231–269. Oxford: Oxford University Press.

Cruttenden, Alan (ed.) 2001. *Gimson's pronunciation of English*, 6th edn. London: Edward Arnold.

Davis, Stuart. 2005. Capitalistic v. militaristic: The paradigm uniformity effect reconsidered. In Downing *et al.* (2005), 107–121.

Donegan, Patricia J. 1993. On the phonetic basis of phonological change. In Charles Jones (ed.) *Historical linguistics: Problems and perspectives*, 98-130. London: Longman.

Downing, Laura J., T. A. Hall & Renate Raffelsiefen (eds.) 2005. *Paradigms in phonological theory*. Oxford: Oxford University Press.

Dressler, Wolfgang U. 1985. *Morphonology: The dynamics of derivation*. Ann Arbor: Karoma.

Ellis, Lucy & W. J. Hardcastle. 2002. Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics* 30. 373–396.

Garrett, Andrew & Juliette Blevins. 2009. Analogical morphophonology. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Essays in honor of Paul Kiparsky*, 527–545. Cambridge, MA: MIT Press.

Gick, Bryan. 1999. A gesture-based account of intrusive consonants in English. *Phonology* 16. 29–54.

Giegerich, Heinz J. 1999. *Lexical strata in English: Morphological causes, phonological effects*. Cambridge: Cambridge University Press.

Hammond, Michael. 1999. *The phonology of English: A prosodic optimality-theoretic approach*. Oxford: Oxford University Press.

Hardcastle, W. J. 1995. Assimilations of alveolar stops and nasals in connected speech. In Jack Windsor Lewis (ed.) *Studies in general and English phonetics: Essays in honour of Professor J. D. O'Connor*, 49–67. London: Routledge.

Hargus, Sharon & Ellen M. Kaisse (eds.) 1993. *Studies in Lexical Phonology*. San Diego: Academic Press.

Harris, James W. 1983. *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: MIT Press.

Harris, John. 1985. *Phonological variation and change: Studies in Hiberno-English*. Cambridge: Cambridge University Press.

Harris, John. 2003. Release the captive coda: The foot as a domain of phonetic interpretation. In Local *et al.* (2003), 103–129.

Harris, John. 2006. Wide-domain *r*-effects in English. *UCL Working Papers in Linguistics* 18. 357–379.

Hay, Jennifer & Margaret Maclagan. 2010. Social and phonetic conditioners on the frequency and degree of intrusive /r/ in New Zealand English. In Dennis R. Preston & Nancy Niedzielski (eds.) *A reader in sociophonetics*,41–70. Berlin & New York: Mouton de Gruyter.

Hay, Jennifer & Andrea Sudbury. 2005. How rhoticity became /r/-sandhi. *Language* 81. 799–823.

Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13. 227–276.

Hayes, Bruce. 2000. Gradient well-formedness in Optimality Theory. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: Phonology, syntax, and acquisition*, 88–120. Oxford: Oxford University Press.

Heselwood, Barry. 2009. R vocalisation, linking R and intrusive R: Accounting for final schwa in RP English. *Transactions of the Philological Society* 107. 66–97.

Holst, Tara & Francis Nolan. 1995. The influence of syntactic structure on [s] to [ʃ] assimilation. In Connell & Arvaniti (1995), 315–333.

Inkelas, Sharon. 1989. Prosodic constituency in the lexicon. Ph.D. dissertation, Stanford University.

Jansen, Wouter. 2004. Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch. Ph.D. dissertation, University of Groningen.

Jensen, John T. 2000. Against ambisyllabicity. *Phonology* 17. 187–235.

Kager, René. 1999. Surface opacity of metrical structure in Optimality Theory. In Ben Hermans & Marc van Oostendorp (eds.) *The derivational residue in phonological Optimality Theory*, 207–245. Amsterdam & Philadelphia: John Benjamins.

Kahn, Daniel. 1976. Syllable-based generalizations in English phonology. Ph.D. dissertation, MIT.

Keating, Patricia. 1988. Underspecification in phonetics. *Phonology* 5. 275–292.

Kenstowicz, Michael. 1996. Base identity and uniform exponence: Alternatives to cyclicity. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, 365–394. Salford: ESRI.

Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10. 421–441.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–365.
Kochetov, Alexei & Marianne Pouplier. 2008. Phonetic variability and grammatical knowledge: An articulatory study of Korean place assimilation. *Phonology* 25. 399–431.
Ladd, D. Robert & James M. Scobbie. 2003. External sandhi as gestural overlap? Counter-evidence from Sardinian. In Local *et al.* (2003), 162–180.
Lipski, John M. 1989. /s/-voicing in Ecuadoran Spanish: Patterns and principles of consonantal modification. *Lingua* 79. 49–71.
Local, John, Richard Ogden & Rosalind Temple (eds.) 2003. *Phonetic interpretation: Papers in laboratory phonology VI*. Cambridge: Cambridge University Press.
Marvin, Tatjana. 2002. Topics in the stress and syntax of words. Ph.D. dissertation, MIT.
Mascaró, Joan. 1987. A reduction and spreading theory of voicing and other sound effects. Unpublished ms., Universitat Autònoma de Barcelona. Published 1995, *Catalan Working Papers in Linguistics* 4. 267–328.
McCarthy, John J. 1991. Synchronic rule inversion. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 17. 192–207.
McCarthy, John J. 1993. A case of surface constraint violation. *Canadian Journal of Linguistics* 38. 169–195.
McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.
McMahon, April. 2000. *Lexical Phonology and the history of English*. Cambridge: Cambridge University Press.
Minkova, Donka. 1991. *The history of final vowels in English: The sound of muting*. Berlin & New York: Mouton de Gruyter.
Mohanan, K. P. 1982. Lexical Phonology. Ph.D. dissertation, MIT. Distributed by Indiana University Linguistics Club.
Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
Nolan, Francis. 1992. The descriptive role of segments: Evidence from assimilation. In Gerard J. Docherty & D. Robert Ladd (eds.) *Papers in laboratory phonology II: Gesture, segment, prosody*, 261–280. Cambridge: Cambridge University Press.
Nolan, Francis, Tara Holst & Barbara Kühnert. 1996. Modelling [s] to [ʃ] accommodation in English. *Journal of Phonetics* 24. 113–137.
Olive, Joseph P., Alice Greenwood & John Coleman. 1993. *Acoustics of American English speech: A dynamic approach*. New York: Springer.
Orgun, Cemil Orhan & Sharon Inkelas. 2002. Reconsidering bracket erasure. *Yearbook of Morphology 2001*. 115–146.
Raffelsiefen, Renate. 2005. Paradigm uniformity effects versus boundary effects. In Downing *et al.* (2005), 211–262.
Robinson, Kimball L. 1979. On the voicing of intervocalic *s* in the Ecuadorian highlands. *Romance Philology* 33. 137–143.
Rubach, Jerzy. 1996. Nonsyllabic analysis of voice assimilation in Polish. *Linguistic Inquiry* 27. 69–110.
Scheer, Tobias. 2008. A lateral theory of phonology, vol. 2: How morpho-syntax talks to phonology: A survey of extra-phonological information in phonology since Trubetzkoy's Grenzsignale. Unpublished ms., University of Nice.
Selkirk, Elisabeth. 1996. The prosodic structure of function words. In James L. Morgan & Katherine Demuth (eds.) *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 187–213. Mahwah, NJ: Lawrence Erlbaum.
Sproat, Richard. 1993. Looking into words. In Hargus & Kaisse (1993), 173–195.
Sproat, Richard & Osamu Fujimura 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21. 291–311.
Steriade, Donca. 2000. Paradigm uniformity and the phonetics–phonology boundary. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, 313–334. Cambridge: Cambridge University Press.

Strycharczuk, Patrycja. 2010. What's in a word? Prosody in Polish voicing. Paper presented at the 18th Manchester Phonology Meeting. Available (August 2010) at http://personalpages.manchester.ac.uk/postgrad/patrycja.strycharczuk/18mfmslides.pdf.

Sugahara, Mariko & Alice Turk. 2009. Durational correlates of English sublexical constituent structure. *Phonology* 26. 477–524.

Szpyra, Jolanta. 1989. *The phonology–morphology interface: Cycles, levels and words*. London & New York: Routledge.

Trommer, Jochen. 2004. Albanian word stress. Unpublished ms., University of Osnabrück.

Trommer, Jochen. 2006. Stress uniformity in Albanian: Morphological arguments for cyclicity. Paper presented at the Workshop on Approaches to Phonological Opacity, 29th GLOW Colloquium, Barcelona.

Trommer, Jochen. 2009. Stress uniformity in Albanian: Morphological arguments for cyclicity. Unpublished ms., University of Leipzig.

Tuinman, Annelie, Holger Mitterer & Anne Cutler. 2007. Speakers differentiate English intrusive and onset /r/, but L2 listeners do not. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 1905–1908. Saarbrücken: Saarland University.

Vennemann, Theo. 1972. Rule inversion. *Lingua* 29. 209–242.

Wells, J. C. 1982. *Accents of English*. 3 vols. Cambridge: Cambridge University Press.

Wells, J. C. 1990. Syllabification and allophony. In Susan Ramsaran (ed.) *Studies in the pronunciation of English: A commemorative volume in honour of A. C. Gimson*, 76–86. London: Routledge.

Wheeler, Max W. 2005. *The phonology of Catalan*. Oxford: Oxford University Press.

Wright, Susan & Paul Kerswill. 1989. Electropalatography in the analysis of connected speech processes. *Clinical Linguistics and Phonetics* 3. 49–57.

Zsiga, Elizabeth C. 1995. An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. In Connell & Arvaniti (1995), 282–302.

90 Frequency Effects

STEFAN A. FRISCH

1 Introduction

This chapter provides a review of evidence that occurrence frequency has an influence on phonological patterns. The examination of quantitative or statistical patterns in phonology, and grammar more generally, pre-dates the development of modern linguistic theory. For example, Zipf (1965) examined statistical properties of texts and noted a variety of effects that still have relevance in current theoretical discussions, such as the tendency for reduction of high-frequency words. Bolinger (1961) and Herdan (1962) discussed gradience and statistical distribution as a natural middle ground between the grammatical extremes of the acceptable and the unacceptable. However, given the limitations both in the understanding of language structure and in the computational resources for conducting quantitative studies, this type of research did not gain much traction in the field of linguistics more generally. With the availability of lexical and usage corpora now available, and armed with the descriptive and theoretical advances of modern linguistic theory, a variety of authors are now arguing for the influence of frequency on phonological patterns in synchronic phonology and morphophonology, phonological acquisition, and diachronic phonology. In many ways, phonology is the ideal domain in which to study the potential role of frequency effects in grammar (Herdan 1962). The set of basic phonological units (phonemes or features) is relatively limited, these units are routinely combined in a reasonably static set of fixed forms (the lexicon), and productive combinations of these units in morphology are also limited in their variety of combination by that same fixed set of units. Morphophonological changes are triggered by phonological environments defined by features (CHAPTER 17: DISTINCTIVE FEATURES) or phonemes (CHAPTER 11: THE PHONEME) and the results of the changes are within the same set of featural or phonemic varieties.

1.1 *Definition of phonology*

In this chapter, I take phonology to be any aspect of language sound structure that can vary systematically between languages or dialects. This would therefore include most of what might traditionally be called phonetics. In the debate over the phonetics–phonology interface, or lack thereof, it has been shown by the laboratory phonology enterprise over the last 20 years that phonetic patterns often vary systematically between languages (Pierrehumbert *et al.* 2000). Articulatory and co-articulatory patterns are not fully determined by physical limitations, but are instead under some degree of (usually unconscious) control in the individual (e.g. Clumeck 1976; Manuel 1990; Beddor *et al.* 2002). Similarly, phonetic perception is tuned by language experience, and not wholly determined by auditory physiology (e.g. Werker and Tees 1984; Kuhl *et al.* 1992; Best and McRoberts 2003). While expanding the domain of phonology to include physically quantitative phonetic effects also expands the potential domain of variability to be dealt with by phonological analysis, this expansion is a necessary step in developing a complete science of language sound structure. While frequency effects can be demonstrated over purely symbolic phonological patterns, expanding the study of the frequency effects on phonology to include traditionally phonetic dimensions may help to integrate synchronic phonology with sociolinguistic and diachronic studies of language sound structure, where phonetic dimensions are often relevant.

Similarly, this chapter will touch on various aspects of morphophonology, where sound patterns vary systematically in morphologically complex words. Demonstrating frequency effects in this domain supports the argument for frequency effects as part of the theory of grammar regardless of the phonetics/phonology issue. In morphophonology, the data are indisputably grammatical in nature, and some aspect of cognitive computation is required. Frequency effects in morphophonology help to show that frequency effects are not merely a residue of performance, memory, general cognition, or diachrony. To the extent that there is a phonological system distinct from other cognitive structures, a variety of evidence has been gathered in support of the use of quantitative frequency information as part of the operating parameters of grammar.

1.2 Definition of frequency

Frequency is the rate of occurrence of a phonological unit, and is unrelated to acoustic frequency. But there are still many possible frequencies, depending on what is taken to be the domain over which occurrences are counted. In studies of language using corpora of language usage, frequency is usually the frequency of occurrence in the corpus. This type of frequency is referred to as token frequency or usage frequency. In English, for example, the token frequency of the phonemes /ð/ and /v/ is relatively high, due to their presence in frequently used words like *the* and *that*, and *of* and *very*. Token frequency for words affects phonetic reduction and the resistance of lexical or morphophonological forms to diachronic change (e.g. Bybee 2002), as detailed in later sections.

Abstracting away from repeated usages of a word, phonological patterns can also be examined on the basis of the number of times the pattern is used across different words. This frequency is referred to as type frequency or lexical frequency. The type frequency of the phonemes /ð/ and /v/ in English is relatively low, as they are used in relatively few words. An example of a consonant with a high type frequency is /b/, which is the most common word onset in English. Some consonants, such as /s/ and /t/, have both high token frequency and high type frequency, being used in many words, many of which are common. Other consonants, such as /θ/ and particularly /ʒ/, have both low token frequency and low type frequency, as they are used in few words, most of which are not common. A consonant cluster with a high type frequency is /st/, which is found in many different words. The token frequency of /st/ is also high, as many of these words are commonly used. Type frequency has been shown to influence metalinguistic judgments for novel word forms (also known as non-words, e.g. Frisch *et al.* 2000), repetition accuracy for non-words (e.g. Vitevitch 2002), and the propensity for phonological or morphophonological generalization for regular or irregular forms (e.g. Bybee 1995; Pierrehumbert 2001; Albright and Hayes 2003).

A few other variants of frequency have been examined in particular cases. Transitional frequency (or transitional probability) is the frequency of one form following another in sequence. Transitional probability for intersyllabic consonant phoneme sequences has been shown to influence the parsing of non-words as simple or morphophonologically complex by adults (e.g. Hay and Baayen 2002), and repetition accuracy for subsequences within non-words in children (e.g. Munson 2001). Neighborhood density is a commonly used measure of word pattern frequency that combines the concepts of frequency and similarity (see also CHAPTER 87: NEIGHBORHOOD EFFECTS). It is typically measured as the number of words that differ from a target word by a single phoneme substitution or a limited number of phoneme substitutions, insertions or deletions (e.g. Goldinger *et al.* 1989; Frisch *et al.* 2000). High neighborhood density has been shown to inhibit word recognition (presumably through competition for lexical access; see Luce and Pisoni 1998), and to facilitate non-word repetition (presumably through activation of frequently used phoneme sequences; Vitevitch 2002).

Neighborhood density can be seen as a more specific application of the general concept of analogy between phonological words or phonological forms. Like neighborhood density, analogy combines the ideas of frequency of occurrence and similarity. Presumably the influence that a phonological pattern could have via analogy is determined in some way by the frequency of the phonological pattern and the similarity of the pattern to the target context (e.g. Bybee 1995; Davidson 2006; see also CHAPTER 83: PARADIGMS). It is also conceivable that analogy could have an influence at a variety of phonological levels, from the phonetic to the phonemic, syllabic, lexical, or morphophonological. The frequency of a phonological pattern at any of these levels may be different, and there is no reason to assume a priori that only one frequency or only one measure of frequency could be relevant.

It is the goal of this chapter to demonstrate that more than one frequency is relevant to phonology, that different frequencies are relevant to phonology in different ways, and that different levels of phonological generalization are relevant, also potentially in different ways (Bybee 2007). Overall, the theoretical position is that of the "ladder of abstractions," where phonemic categories are generalizations over phonetic patterns, sub-syllabic, and syllabic categories are generalizations over phonemic patterns, lexical and morphophonological categories are generalizations over phonemic, sub-syllabic, and syllabic patterns, and so forth (Pierrehumbert 2003; Beckman and Edwards 2010). Under this view, higher-level phonological categories emerge as systematic generalizations over lower-level categories, where the lowest-level category is physical/articulatory/acoustic experience with language. This experience is parsed into more general, abstract categories as repeated similar experiences occur. Frequency may play a role in supporting these generalizations. Token frequency creates robust, entrenched, well-defined categories through frequent exposure. Type frequency leads to grammatical generalization, as categories occur in a variety of contexts, promoting generalization and parsing of units, presumably for cognitive efficiency (Bybee 2007). Note also that this view includes frequency information as part of the grammar, and not just as a tool that can be used to analyze a lexicon or corpus for phonologically relevant patterns (e.g. as in McCarthy 1994 or Hammond 1999; see the introduction in Bod *et al.* 2003 for discussion).

2 Frequency matching in metalinguistic judgments

Evidence for the relevance of frequency in phonology is found in behavioral experiments in which the frequency or probability of phonological units in stimuli is manipulated or examined, and native-speaker judgments are consulted. It has been found that native speakers are sensitive to phonological frequency in a variety of lexical and morphophonological patterns. In the absence of traditional phonological constraints, it has been shown that baseline phonotactic probability is an influence on well-formedness judgments for novel non-words. This suggests that native speakers have encoded frequency information that is present in language data. Whether this knowledge is properly a part of their grammatical competence, as opposed to extra-linguistic information that is accessed as a part of linguistic performance, is a topic of debate. In connection with data presented in later sections that frequency has an influence over the distribution of possible forms in the language, the influence of frequency on native-speaker performance in metalinguistic experiments has been used to argue that frequency information is part of the knowledge of language sound structure that speakers have, i.e. that frequency effects are psychologically real (Frisch *et al.* 2004; Zuraw 2007).

2.1 *Novel phonotactic combinations*

Frisch *et al.* (2000) presented recordings of novel non-words to naive undergraduate listeners. These non-words were constructed by combining attested onset and rime constituents into multisyllabic non-words with no obvious phonotactic violations. The onset and rime constituents (CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE) used were of relatively high or low frequency of occurrence, creating novel non-words that were relatively high or low in cumulative expected probability, but with the frequency of occurrence of any particular onset or rime constituent balanced across the experiment. Example non-words are given in (1).

(1) /sɪʃəp/
/zujɛθʊs/
/sɪrɔsɛnɔn/

Frisch *et al.* (2000) found a significant, but moderate, correlation between cumulative expected probability and native English speaker well-formedness judgments (r ~ 0.4). Well-formedness judgments were collected in two different experiments (CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY). In one experiment, speakers were asked to judge the wordlikeness of the novel non-words on a scale from 1 to 7 (1 = impossible, can't be a word of the English; 2, 3 = unlikely, doesn't sound much like a word of English; 4 = neutral, sounds somewhat like a word of English; 5, 6 = likely, sounds like it could be a word of English; 7 = definitely, sounds just like a word of English). In the other experiment, speakers were asked to judge whether a non-word was acceptable or unacceptable. In both cases, there was evidence for gradient acceptability of the novel non-words as a function of phonotactic probability (CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY). For acceptability judgments, gradient acceptability was apparent when judgments of acceptability were aggregated across participants. For very low probability non-words, very few speakers judged them to be acceptable. For very high probability non-words, most speakers judged them to be acceptable. For non-words of intermediate probability, some speakers judged them to be acceptable and some did not. For wordlikeness judgments, where individual participants were able to use a gradient scale to make their judgments, the analysis of individual speaker data found significant correlations with non-word expected probability for most participants. These findings are replicated in Frisch and Brea-Spahn (2010), where the non-words were constructed slightly differently, by random combination of onset and rime constituents (screening out any categorical phonotactic violations that were created by this random process, such as the creation of a geminate across the syllable boundary; CHAPTER 37: GEMINATES).

Bailey and Hahn (2001) presented monosyllabic non-words to British English speakers in two experiments, one orthographic and the other auditory. Their non-words were selected to differ from real English words by either one or two phonemes (e.g. *drump* and *drolf*) but otherwise violated no categorical phonotactic constraints. Participants rated the novel items on a 1 to 9 wordlikeness scale. Bailey and Hahn (2001) examined a variety of predictors for judgments, focusing on probabilistic phonotactics and lexical neighborhood density. They found that each factor provided an independent influence, and thus they argued that both direct lexical information and abstract probabilistic phonotactic information are used in the wordlikeness judgment task. Frisch *et al.* (2000), in a post hoc examination of their polysyllabic non-word data, found somewhat similar results. While polysyllabic words generally have fewer lexical neighbors, the highest probability non-words did show some evidence for lexical neighborhood effects. However, a recent study by Shademan (2006) failed to replicate the lexical neighborhood effects of Bailey and Hahn (2001). Given that phonotactic probability and lexical neighborhood density are confounded with one another, it is perhaps not surprising that it has been difficult to differentiate the two and demonstrate clear influences of both in metalinguistic experiments.

Frisch *et al.* (2000) was a replication of a study by Coleman and Pierrehumbert (1997). However, the Coleman and Pierrehumbert (1997) study was less systematic in its construction of non-words, and also included non-words with phonotactic violations in onset consonant clusters. They used an acceptability judgment task, and examined aggregate acceptability across the participants in the study. Coleman and Pierrehumbert (1997) found a similar correlation between non-word expected probability and acceptability. They also found that non-words with phonotactic violations were judged as more or less acceptable depending on the frequency of the other (non-violating) constituents in the word. In other words, high frequency elsewhere in a novel non-word could mediate the detriment to well-formedness caused by a phonotactic violation. The findings of Coleman and Pierrehumbert (1997) are compatible with models of phonological grammar that use a cumulative or aggregate well-formedness in evaluating the output of the grammar, but they are difficult to capture in a model of grammar where only grammatical violations or only the greatest grammatical violation is relevant to the output.

Albright (2009) also examined well-formedness judgments for monosyllabic English non-words containing a variety of phonotactically legal and illegal sequences, with variation in frequency of the phonotactically legal sequences. Albright used a model based on transitional probability between natural classes (feature groupings; CHAPTER 17: DISTINCTIVE FEATURES) in an attempt to create a more linguistically grounded model that might extend frequency patterns in attested sequences to linguistically related unattested sequences. The goal of the model was to differentiate unattested onset clusters like /bw bn bz/ that have been shown to vary in their wordlikeness (CHAPTER 55: ONSETS). While the Albright model was successful, he also found that the transitional frequencies between natural class generalizations and the transitional frequencies between segments each contribute to predicting participant judgments, and do not overlap completely. This suggests that multiple levels of generalization may be relevant to participant performance in well-formedness judgment tasks.

Frisch and Stearns (2006) examined onset consonant clusters, presenting a variety of CC initial monosyllabic non-words auditorily to participants (cf. similar non-words presented to child participants by Scholes 1966, reported in Albright 2009). The stimuli included attested English onset clusters with varying frequency, and a few unattested clusters that might be expected to occur in English but do not, based on relatively simple phonological categorization (/sr tl dl θl/). For the attested clusters, cluster frequency was a significant predictor of well-formedness judgments. Judgments of the unattested clusters were surprisingly high, however, and some follow-up investigation has suggested that these clusters were frequently misperceived (e.g. /tl/ heard as /pl/). When orthographic supports were provided (in an unpublished replication study), well-formedness judgments for the non-attested clusters were lower, though not necessarily very different from low-frequency attested clusters such as /gw sf dw/, which might be thought of as less consistent with the overall grammar of consonant clusters in English (Hammond 1999).

2.2 *Frequency matching in morphophonology*

It has also been shown that phonological frequency influences participant behavior in metalinguistic tasks that more directly reflect linguistic productivity. For example, Ernestus and Baayen (2003) presented novel verbs to Dutch speakers and asked them to produce past-tense forms. Dutch is one of many languages with a process of word-final devoicing. Dutch verbs contain a variety of examples where stem-final obstruents alternate, as in (2) (see also CHAPTER 80: MERGERS AND NEUTRALIZATION).

(2) /vɛrʋɛit/ 'widen (3SG PRES)'
/vɛrʋɛidən/ 'widen (INF)'

There are also examples where stem-final consonants do not alternate. The standard phonological analysis is that the stem-final consonant may be voiced or voiceless, with neutralization to the voiceless consonant when the stem-final consonant is word-final (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL

NEUTRALIZATION). Ernestus and Baayen (2003) analyzed the Dutch lexicon and found that cases of alternation are not equally distributed across different natural classes of consonants. For example, stops have a very low rate of alternation, while fricatives have a higher rate of alternation. The rate of alternation also depends on place of articulation (CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION). In an experiment, Ernestus and Baayen (2003) presented participants auditorily with a variety of novel verbs with potentially neutralized stem-final consonants (i.e. stimuli with voiceless final consonants), and asked participants to supply an alternative form where the underlying consonant is no longer in a neutralizing context (in this case, the past tense, which is spelled either *te* or *de* depending on the voicing of the underlying consonant). An analysis of responses found variation in the frequency with which participants guessed that the stem-final consonant had been underlyingly voiced or voiceless. The frequency distribution of responses matched the frequency distribution of alternation in the lexicon. In particular, even for consonants with a relatively low frequency of alternation, participants still occasionally guessed that there was alternation, rather than uniformly choosing the most likely result, which would be no alternation. Ernestus and Baayen (2003) attribute these frequency effects to the influence of groups of related lexical items, as in a lexical neighborhood effect.

Similar results are found in the case of Hungarian vowel harmony (with the most extensive analysis involving back vowel harmony patterns in the lexicon) and in experimental behavior (Ringen and Vago 1998; Siptár and Törkenczy 2000; Hayes and Londe 2006; Hayes *et al.* 2009; see also CHAPTER 123: HUNGARIAN VOWEL HARMONY). Hungarian back vowel harmony is sensitive to roundness for front vowels and proximity between the vowel trigger and the suffix. When a stem-final vowel is front rounded, the dative suffix takes a front vowel. When the stem-final vowel is back, the dative suffix takes a back vowel. When the stem-final vowel is front unrounded, some variation is found, depending on vowels earlier in the word. Cases where all vowels in the stem are front unrounded, or where there is a back vowel earlier in the word, can result in either back (3a) or front suffixes (3b), or in synchronic variation between these (3c) (examples from the web corpus study of Hayes and Londe 2006).

(3) a. /pɔlleːr-nɔk/ 'foreman-DAT'
 b. /mutɔgeːn-nɛk/ 'mutagen-DAT'
 c. /ɔrzeːn-nɔk/ /ɔrzeːn-nɛk/ 'arsenic-DAT'

The frequency of back vowel harmony is influenced by proximity of the triggering back vowel to the suffix. With one intervening front unrounded vowel, harmony is fairly common. With two intervening front unrounded vowels, harmony is fairly uncommon. The height of the intervening front unrounded vowel is also a factor, with /i/ more likely to be transparent (facilitating back harmony) and /e/ less likely to be transparent (blocking back harmony). Hayes and Londe (2006) found frequency matching by Hungarian participants in a pencil-and-paper experiment in which participants were asked to generate dative forms for novel Hungarian-like nouns by filling in blanks in a paragraph. This finding was replicated and extended by Hayes *et al.* (2009) in an experiment conducted via the Internet. In Hayes *et al.* (2009), frequency matching was demonstrated not only for the influences of trigger and potentially transparent vowels, but also less robustly for generalizations involving place and manner of intervening consonants and consonant clusters. Hayes *et al.* (2009) describe these as unnatural constraints, as there is no clear phonological connection between the consonant features involved and vowel backness, and suggest that there may be a bias against the learning of unnatural phonological constraints. Alternatively, the influence of these unnatural groupings may merely reflect the influence of lexical neighborhoods of related words.

Overall frequency matching was also found in a study by Zuraw (2007), in which the marginally productive process of nasal substitution in Tagalog was used in a two-part experiment. Tagalog speakers were asked to generate novel derived forms, as in Ernestus and Baayen (2003), and subsequently asked to judge the well-formedness of derived variants on a ten-point scale. A novel example from Zuraw (2007) is shown in (4).

(4) /bugnat/ 'to *bugnat*'
/mambubugnat/ '*bugnat*-er' (no nasal substitution)
/mamumugnat/ '*bugnat*-er' (nasal substitution)

Nasal substitution in Tagalog is found to varying degrees across the lexicon, and is influenced by the voicing and place of articulation of the consonant (CHAPTER 78: NASAL HARMONY). Zuraw (2007) found that speakers' use of nasal substitution and their judgments of substituted forms were influenced by voicing, but she did not find a significant effect of place of articulation. In addition, she found that the use of nasal substitution in novel forms was overall less frequent than its appearance in the lexicon, indicating that the participants in her study had a bias for the regular (non-substituted) form.

In summary, metalinguistic experiments have shown that speaker judgments of novel phonological forms can be influenced by the relative frequency of sub-syllabic constituents within the non-word, as well as by lexical patterns in the application of phonological processes. These data suggest that speakers have encoded frequency information within a variety of static phonotactic constituents as well as within the targets, results, and contexts of phonological and morphophonological processes. In other words, these data suggest that speakers have statistical knowledge of language sound structure at a variety of levels.

3 Frequency in phonotactics: OCP-Place

The previous section showed examples of the influence of phonological frequency on native speaker performance in phonological experiments. Quantitative patterns such as these have become a point of interest in the development of theories of grammar and grammar acquisition. In this section, elementary concepts are introduced, in conjunction with the application of these concepts to consonant co-occurrence data. These data have been a driving force in the development of quantitative theories of phonology.

3.1 *Base phonotactic frequency*

It has long been known that phonemes do not all occur with equal frequency (e.g. Zipf 1965; CHAPTER 11: THE PHONEME). Availability of electronic corpora in recent years has made analyzing the frequency distributions of phonemes easier, leading to comprehensive studies of a large number of languages and statistical distributions (e.g. Martindale *et al.* 1996; Tambovtsev and Martindale 2007), with substantiation of the claim that frequency differences between phonemes influence phonological processing.

Given that base phoneme frequencies may vary, the study of phonological combinations requires that these base frequency differences be taken into account (Pierrehumbert 1993, 1994). This approach was taken in a groundbreaking study by Greenberg (1950), in which the statistical combination of consonants in Arabic verbal roots was examined. Greenberg noted several series of consonants that tended not to co-occur with one another within a single root, primarily based on place of articulation. These generalizations were formalized by McCarthy (1994), using the more general Obligatory Contour Principle (OCP) applied to place of articulation features. In its simplest form, this constraint bans the co-occurrence of consonants within a root that share a place of articulation feature. It was also noted that the frequency of co-occurrence varied between places of articulation; that this variation was sensitive to manner features; and that distance between co-occurring consonants was relevant as well (Pierrehumbert 1993; McCarthy 1994; Padgett 1995). However, in a root lexicon modeling study, Frisch *et al.* (2004) found that a significant portion of consonant co-occurrence frequency across the lexicon could be predicted from base phoneme frequency alone (see also CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS).

3.2 *Phonotactic constraints and co-occurrence frequency*

In the statistical analysis of phonotactic co-occurrence, researchers typically attempt to determine whether some combination is significantly statistically underrepresented, indicating a constraint against co-occurrence (e.g. Greenberg 1950; McCarthy 1994). Presumably, statistical overrepresentation could also be found, indicating a preference for certain phonotactic combinations, such as in the co-occurrence of nasal consonants with homorganic stops in clusters. Alternatively, the rate of co-occurrence between related classes of consonants can be examined for a significant statistical variance, indicating a difference in the strength of application of a constraint between two co-occurrence classes (e.g. Frisch *et al.* 2004). These statistical approaches all use the notion that the observed frequency of a combination should be equal to the expected frequency of co-occurrence by chance (assuming independent combination of the individual units, such as phonemes, onsets, rimes, or syllables).

As a simple example, consider the following hypothetical table (see Table 90.1) of aggregated consonant co-occurrence data. The style of this example follows Wilson and Obdeyn (2009), and is similar to actual data in several studies (Padgett 1995; Frisch *et al.* 2004; Coetzee and Pater 2008). This could represent the data on consonant co-occurrence from a sample of 500 words that begin with CVC from a language, where the consonant co-occurrences have been aggregated within place of articulation classes to look for OCP-Place effects (cf. the examination of co-occurrence in 500 Italian verbs in Frisch *et al.* 2004). In this table, the number of CVC pairs with each place of articulation combination is given, and the row and column marginal totals show the base frequency of occurrence of consonants in C1 (right-hand column) or C2 (bottom row) position.

Table 90.1 Hypothetical co-occurrence data for place of articulation classes in 500 words (following Wilson and Obdeyn 2009)

	[labial]	[coronal]	[dorsal]	Mar C1
[labial]	12	83	33	128
[coronal]	104	44	70	218
[dorsal]	54	91	9	154
Mar C2	170	218	112	500

If consonants co-occurred randomly in these data, we would expect the place of articulation combinations to be evenly distributed across the different combination groups. In other words, the expected probability of combination would be equal to the marginal probability of C1 multiplied by the marginal probability of C2. The marginal probability of C1 or C2 is the ratio of the frequency of occurrence of the consonant divided by the total number of occurrences. These calculations are summarized in the following equation:

(5) $$Expected(C_1,C_2) = p(C_1) \times p(C_2) = \frac{obs(C_1)}{total(C_1)} \times \frac{obs(C_2)}{total(C_2)}$$

For the case of the [labial]–[labial] combination, for example:

(6) $$Expected([labial],[labial]) = \frac{170}{500} \times \frac{128}{500} = 0.08704$$

In a lexicon of 500 items, $0.08704 \times 500 = 43.5$ occurrences are expected. Performing this calculation for all combinations gives the following table (Table 90.2) of expected consonant co-occurrence, given random combination of C1 and C2. It's important to note that, in this view, the marginal probabilities for C1 and C2 are taken as a starting point for examining the combinations of C1 and C2.

Pierrehumbert (1993) attempted to simplify the inspection of statistical co-occurrence data such as these. She used the ratio of the observed (O) co-occurrence *vs.* the expected (E) co-occurrence to provide a single number that summarized the relative under- or overrepresentation of combinations compared to chance.

Table 90.2 Expected frequency of co-occurrence for hypothetical data in Table 90.1

	[labial]	[coronal]	[dorsal]	Mar C1
[labial]	43.5	55.8	28.7	128.0
[coronal]	74.1	95.0	48.8	218.0
[dorsal]	52.4	67.1	34.5	154.0
Mar C2	170.0	218.0	112.0	500.0

Table 90.3 Observed over expected (O/E) ratio for hypothetical data in Table 90.1

	[labial]	[coronal]	[dorsal]
[labial]	0.28	1.49	1.15
[coronal]	1.40	0.46	1.43
[dorsal]	1.03	1.36	0.26

Examining these ratios, the hypothetical data show a statistical pattern of low co-occurrence, O/E ratio less than 1, for combinations at the same place of articulation. This would be taken as evidence for the operation of a constraint against co-occurrence for homorganic consonants (i.e. OCP-Place). For co-occurrence for different places of articulation the O/E ratio is greater than 1, meaning co-occurrence is at or above the level expected by chance. This would be taken as evidence that there is no constraint against co-occurrence for consonants at different places of articulation.

Since McCarthy (1994), data such as these have been used not only to argue for the presence of a co-occurrence constraint for natural classes of phonemes, but also that there can be different degrees of constraint, or gradience, in phonotactic constraints. Quantitative differences in the frequency of occurrence of combinations (or in O/E ratios) have been used to argue that some combinations may occur, but are more or less restricted than others. In the case of OCP-Place, quantitative differences have been used to argue for stronger constraints for a variety of consonant combinations: a stronger constraint for identical consonant combinations in comparison to homorganic, but not identical, consonants (McCarthy 1994); a stronger constraint for consonants that are closer together within the root or word (Pierrehumbert 1993; Buckley 1997; Berkley 2000); a stronger constraint for consonants that share manner features or are more similar in general (McCarthy 1994; Padgett 1995; Frisch *et al.* 2004; Coetzee and Pater 2008).

As grammatical models of frequency effects have become more sophisticated, it has been suggested that quantitative O/E data for phonotactic combinations should be treated with caution. Wilson and Obdeyn (2009) demonstrate that differences in O/E such as those in Table 90.3 may arise due to a combination of constraints, without necessarily indicating a difference in the strength of a constraint. First, note that in the hypothetical data there are differences in the marginal probabilities for the different consonant places of articulation, as is common in natural data. The [coronal] class is overall more frequent both as C1 and as C2 in comparison to the [labial] and [dorsal] classes (CHAPTER 12: CORONALS). In the MaxEnt grammar model Wilson and Obdeyn (2009) use, discussed later in this section, these differences in marginal probability arise from constraints on individual consonant classes. These constraints would then interact with OCP-Place constraints for combinations of consonants.

Table 90.4 shows a set of hypothetical constraint weights for this general type of interaction model, based on the hypothetical data. The hypothetical constraint weights are given as fractions in order to more transparently reveal their impact on the frequency of occurrence of features and feature combinations, and so are

Table 90.4 Hypothetical constraint weights for a simplified interactive model of consonant phonotactics

	[labial]	[coronal]	[dorsal]	Mar C1
[labial]	0.25	1	1	0.256
[coronal]	1	0.25	1	0.436
[dorsal]	1	1	0.25	0.308
Mar C2	0.340	0.436	0.224	

not identical to the weights for a MaxEnt model of these data. These weights are most similar to the quantitative constraint values used in the Frisch *et al.* (2004) stochastic constraint model. In a MaxEnt model, the constraint weights would be inversely related to frequency and the connection from a particular weight value to frequency is less transparent. In this table, the marginal values are constraint weights for individual place of articulation features in C1 or C2 position (based directly on the marginal probabilities for simplicity of exposition), and the combination weights represent a uniform OCP-Place effect across different places of articulation. A 0.25 weight on combinations of the same feature is used in this example, providing a strong but not categorical restriction on co-occurrence. A weight of 1 is assigned to combinations for different features, suggesting these pairs are unrestricted.

From these constraint weights, it can be shown that differences in the cumulative weight for place combinations can result as a combination of the weights for marginal probabilities and the uniform OCP-Place weight for all combinations. For example, for [labial]–[labial] combinations:

(7) Weight(C1 [labial]) × Weight(C2 [labial]) × Weight(C1 [labial] & C2 [labial]) $= 0.256 \times 0.340 \times 0.25 = 0.02176$

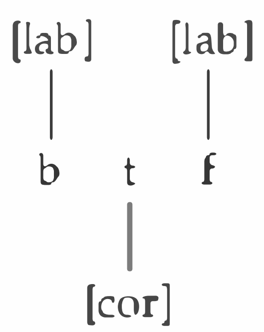
But for [coronal]–[coronal] combinations:

(8) Weight(C1 [coronal]) × Weight(C2 [coronal]) × Weight(C1 [coronal] & C2 [coronal]) $= 0.436 \times 0.436 \times 0.25 = 0.04752$

The ratio of the cumulative constraint weights is approximately 1 to 2, which roughly corresponds to the ratio of O/E for [labial]–[labial] *vs.* [coronal]–[coronal] combinations. Thus the same degree of direct constraint on consonant combination, interacting with positional constraints for differences in marginal consonant frequencies, produces differences in the cumulative constraint for combinations. In the Wilson and Obdeyn (2009) approach, the base phonotactic frequencies for consonant occurrence are modeled as part of the grammar, rather than taken as a starting point from which to model phonotactic combination.

3.3 *Quantitative OCP-Place effects*

A fruitful domain for the examination of frequency as a relevant dimension for phonological constraints has been the study of long-distance constraints on consonant combination, originally characterized for Arabic as an application of the Obligatory Contour Principle to place of articulation classes (McCarthy 1994). McCarthy (1988) first observed that the occurrence of Arabic verbal roots with identical consonants was restricted to forms of the type *sdd*, where C2 and C3 are identical. This pattern was analyzed using the formalism of autosegmental phonology as involving spreading from the C2 to C3 position in a root with only two consonants, e.g. *sd*, in order to fill three consonant positions in a template (CHAPTER 105: TIER SEGREGATION; CHAPTER 108: SEMITIC TEMPLATES). This analysis of the asymmetry in appearance of identical consonants in the Arabic verbal roots more or less still stands. McCarthy (1994) observed that the co-occurrence series originally presented by Greenberg (1950) could also be analyzed as an application of the OCP to place of articulation features where place of articulation features are located on distinct autosegmental tiers in feature geometry (CHAPTER 27: THE ORGANIZATION OF FEATURES). One advantage of this analysis was the ability to capture co-occurrence constraints between C1 and C3 in a three-consonant root, where the middle consonant would be invisible to the application of the constraint, due to the segregation of place features onto distinct tiers, as in (9). In this approach, two distinct OCP constraints are active in the Arabic roots, in order to prevent roots with identical C1–C2, and also to restrict roots with homorganic consonants in any position.

(9)


Frisch *et al.* (2004) argue that some type of gradient analysis of OCP-Place effects is necessary, as any set of categorical constraints on co-occurrence will either have an unacceptably large number of exceptions or fail to explain an unacceptably large number of non-occurrences. They constructed a variety of categorical OCP-Place models, based on excluding consonant pairs with different degrees of specificity of feature combination, and then counted the number of consonant pairs that occurred despite being restricted (exceptions) or did not occur relative to the expected frequency of occurrence for non-homorganic pairs (unexplained underrepresentation). Their data are presented in Table 90.5. The "enumerated pairs" model excludes only the consonant pairs that occur with zero frequency, which cannot be defined by a natural class. Previous models proposed a variety of feature combinations like the ones in this table, but cannot explain why

Table 90.5 Failure of categorical models of OCP-Place to capture exceptions or eliminate gradiently restricted pairs

Definition of classes	*Exceptions*	*Unexplained underrepresentation*
Place only	816	0
Place & [son]	123	160.8
Place & [son] & [cont]	36	312.7
Enumerated pairs	0	430.5

restricted pairs occur with such high frequency, or why unrestricted pairs occur with such low frequency, relative to the expected frequency of occurrence.

Quantitative OCP-Place type effects have been demonstrated for several languages: Arabic, Maltese, Italian, Thai (Frisch *et al.* 2004); Muna (Coetzee and Pater 2008); Shona, Wargamay (Wilson and Obdeyn 2009); Javanese (Mester 1986); Tigrinya (Buckley 1997); Russian (Padgett 1995); English, French, Latin (Berkley 2000). In all of these cases, featural similarity appears to play a role. Perhaps more importantly, there is yet to be an instance where a language is analyzed quantitatively and gradient OCP-Place effects are not found. Consonant place-based co-occurrence restrictions have been discussed for an even larger set of languages, though these have not been analyzed quantitatively from a similarity perspective (see Yip 1989, Frisch *et al.* 2004, Coetzee and Pater 2008 for references). Analogous constraints have also been investigated for laryngeal features (Steriade 1982; Itô and Mester 1986; MacEachern 1999), but these also have not yet been studied quantitatively.

3.4 *Subsidiary features and underspecification*

An interesting wrinkle to the OCP-Place constraint data for Arabic is that the constraint apparently applies more strongly within some places of articulation than others. In Arabic, for the labial, dorsal, and guttural places of articulation, co-occurrence within these places of articulation is virtually non-existent (with a few exceptions, depending exactly on how the classes are defined). Within the coronal place of articulation, however, co-occurrence can be quite frequent, particularly between obstruents and sonorants, and less so for obstruents differing on continuant, as suggested by Table 90.5. Table 90.6 shows the Arabic coronal

Table 90.6 Co-occurrence for coronal consonants in the lexicon of Arabic triconsonantal roots (totals include combinations with non-coronal consonants, which are not listed in the table)

	C2/C3														
C1/C2	t	d	tˤ	dˤ	θ	ð	s	z	sˤ	zˤ	ʃ	l	r	n	total
t	0	3	0	0	0	0	0	0	0	0	2	11	17	6	102
d	0	1	0	0	2	0	8	0	0	0	4	16	22	14	198
tˤ	0	1	0	0	0	0	5	0	0	0	5	15	19	8	140
dˤ	0	3	0	0	0	0	0	0	0	0	0	5	12	4	85
θ	1	0	0	0	0	0	0	0	0	0	0	6	13	1	55
ð	0	0	0	0	0	0	0	0	0	0	0	7	17	2	72
s	2	13	8	0	0	0	0	0	0	0	0	21	22	9	230
z	0	0	0	0	0	0	0	0	0	0	0	11	15	9	151
sˤ	1	14	0	0	0	0	0	0	0	0	0	17	13	5	138
zˤ	0	0	0	0	0	0	0	0	0	0	0	3	3	0	22
ʃ	4	9	9	0	0	2	0	1	0	2	0	6	23	9	200
l	5	9	13	0	3	2	15	4	5	2	2	0	0	1	270
r	9	21	11	12	7	1	21	13	10	1	16	0	0	9	400
n	10	12	10	7	4	2	15	10	9	3	14	0	2	0	324
total	117	243	151	83	61	55	200	123	100	29	141	365	497	226	

co-occurrence data for individual coronal consonants in C1–C2 or C2–C3 position, from Frisch *et al.* (2004). In Table 90.6, cells showing no co-occurrence are primarily clustered near the diagonal, where identical or nearly identical consonant pairs are found.

Theoretical accounts of the OCP-Place constraint have used the difference between coronal consonants and other consonants to argue for place of articulation-specific constraints on coronals. Davis (1991), examining co-occurrence data from English, claimed that coronal place of articulation involves underspecification, with [coronal] not present in the feature representation, and so not subject to the OCP. For Arabic, there are clearly restrictions on coronals evident in Table 90.6, and so an absolute restricted/unrestricted distinction cannot be maintained on the basis of place of articulation. Quantitative accounts of place differences in the strength of OCP-Place constraints vary in their method of accounting for these differences, as detailed in the next two sections.

3.5 Similarity and OCP-Place

Pierrehumbert (1993) attempted to unify the two distinct OCP constraints, and also to capture differences in the frequency of occurrence of distinct consonant combinations at different places of articulation. She proposed that consonant combinations are restricted on the basis of their perceived similarity, where perceived similarity is influenced by two factors. One factor is the featural similarity of the consonants involved. Identical consonants would be maximally similar. Homorganic consonants that also share manner features would be highly similar. Homorganic consonants that do not share manner features would be less similar. The second factor influencing perceived similarity is distance. By her hypothesis, the similarity of consonants that are closer together in the root (C1–C2 or C2–C3) is more easily perceived than the similarity of consonants that are further apart (C1–C3). Pierrehumbert (1993) argues that perceived similarity corresponds to the frequency of co-occurrence of consonant combination, providing a unified account of variation in the strength of OCP-Place effects for different consonant combinations (CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). Pierrehumbert (1993) also extended the perceived similarity account to include the influence of contrastiveness on similarity using contrastive underspecification. In this approach, places of articulation with a relatively large number of contrasts (e.g. Arabic coronals) will have relatively more features influencing similarity in comparison to places of articulation with a relatively small number of contrasts (e.g. Arabic labials). In other words, the size of the inventory at a particular place of articulation would predict how weak or strong the co-occurrence constraint is at that place.

The similarity account of OCP-Place effects in Arabic was more concretely formalized in Frisch *et al.* (2004), using a categorization function model of a gradient phonotactic constraint. In this model, similarity was related to a co-occurrence weight similar to the fractional weights used in Table 90.4. High-similarity homorganic consonants had a co-occurrence weight near zero, reflecting their near complete absence from Arabic. Non-homorganic consonants were analyzed as having no similarity and a co-occurrence weight near one. Homorganic consonants with intermediate degrees of similarity had weight values in between these extremes along a smooth curve. An example of this type of constraint is shown in Figure 90.1. The application of this type of constraint to single consonant co-occurrence pairs

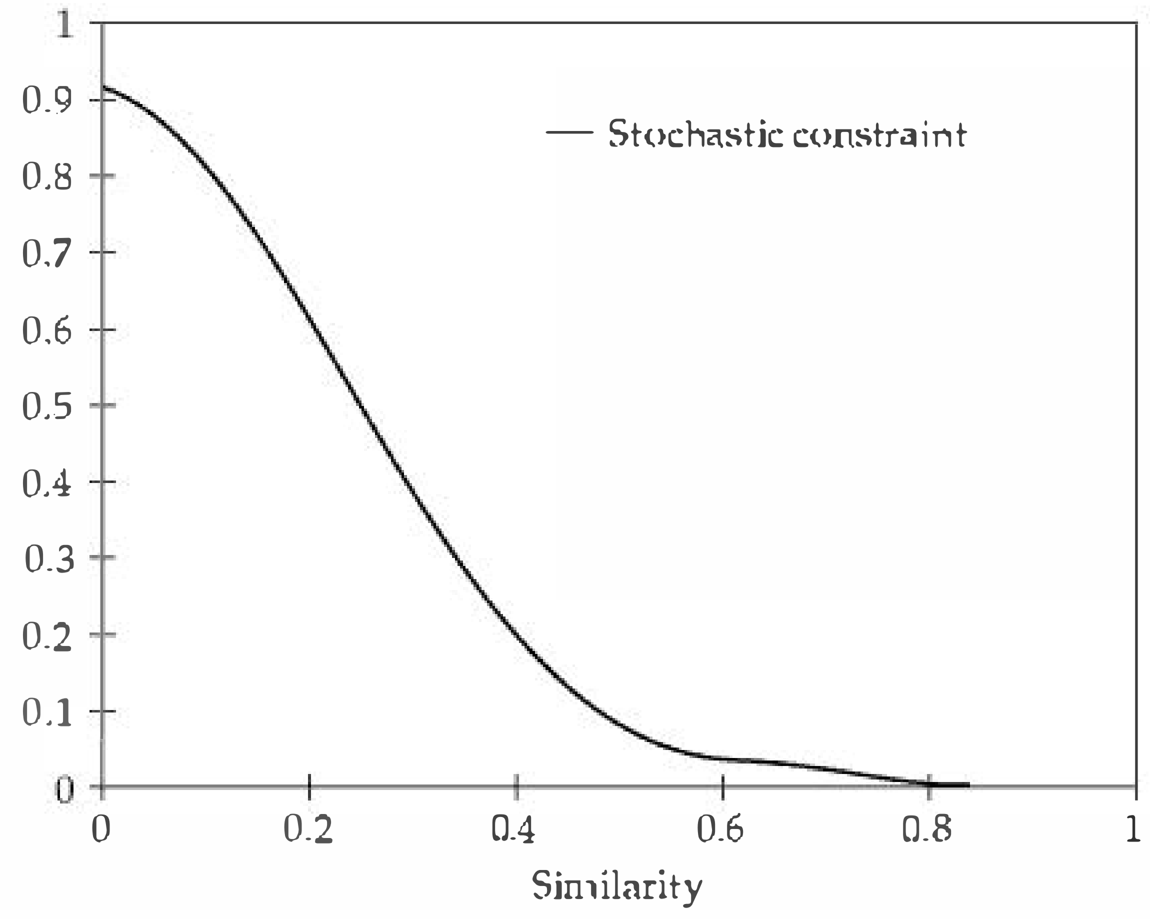


Figure 90.1 Stochastic constraint model of gradient similarity avoidance for consonant pairs in Arabic (after Frisch *et al.* 2004)

was extended in Frisch (2000) to model entire three-consonant roots by combining together constraints on C1–C2, C2–C3, and C1–C3 combinations.

3.6 *Modern interactive constraint models of OCP-Place effects*

Parallel to the development of the Pierrehumbert (1993) and Frisch *et al.* (2004) stochastic constraint model, formal phonological models of categorical constraint interaction were developed using Harmonic Grammar (Legendre *et al.* 1990; Smolensky and Legendre 2006) and Optimality Theory (Prince and Smolensky 2004). Quantitative implementations of these grammatical approaches have been applied to the OCP-Place data of Arabic and other languages. These models differ in a variety of theoretical points that are, to some extent, orthogonal to the issue of frequency effects. They will be discussed primarily from the perspective of their differences in modeling frequency effects in corpus data.

One way in which these models differ is in the degree of restriction on independent free parameters used to fit the frequency data. In Frisch *et al.* (2004), the degree of restrictiveness of consonant combination was proportional to perceived similarity, resulting in a relatively low number of free parameters. Anttila (2008a) makes a proposal most analogous to Frisch *et al.* (2004), stating that OCP-Place consonant co-occurrences can be predicted on the basis of grammatical complexity, where grammatical complexity comes from the required number of pairwise grammatical constraint rankings that are required to achieve a faithful output in an optimality-theoretic grammar. The restrictions on more specific natural classes of co-occurrence require more constraint rankings in comparison to more general natural classes of co-occurrence. For the case of Arabic coronals, for example, Anttila (2008a) presents the following constraints and rankings for representative consonant pairs:

(10) a. t-d
FAITH >> OCP-COR ∧
FAITH >> OCP-COR[−son] ∧
FAITH >> OCP-COR[−son, αcont]

b. t-s
FAITH >> OCP-COR ∧
FAITH >> OCP-COR[−son]

c. l-n
FAITH >> OCP-COR ∧
FAITH >> OCP-COR[+son]

d. t-n
FAITH >> OCP-COR

A relatively larger number of rankings must hold to allow co-occurrence to the *t-d* case (combination of coronal stops or coronal fricatives) for the input to map faithfully to the output, in comparison to the *t-n* case (combination of a coronal obstruent with a coronal sonorant). The complexity hypothesis is analogous to the account of laryngeal OCP constraints in MacEachern (1999), where the parallel between OCP constraints with multiple features and the notion of similarity due to multiple feature matches is explicitly discussed. There is an implicit hierarchy among these OCP constraints, where less specific constraint rankings can make more specific constraint rankings irrelevant. If OCP-COR >> FAITH, any coronal pair will be unable to surface faithfully, regardless of the ranking of OCP-COR[−son], for example. Since more specific feature matches for the OCP constraints means higher similarity between phonemes, the complexity hypothesis and the similarity account of the OCP make analogous predictions and are both fairly restrictive in the number of free parameters that can be used to fit the data.

Coetzee and Pater (2008) take a different approach from constraint ranking, using an accumulation of constraint violations for the same sort of constraint set as above (Anttila 2008a credits Coetzee and Pater 2008 for the constraint set that he uses). In the Coetzee and Pater (2008) approach, however, constraints are weighted relative to one another, rather than ranked relative to one another, as an application of Harmonic Grammar to the consonant co-occurrence problem. Constraints with higher weight have a greater impact on the likelihood of appearance of a form that violates that constraint. Forms that violate multiple highly weighted constraints are unlikely to appear. The most frequent forms will be those that violate the fewest and lowest-weighted constraints. In the Coetzee and Pater (2008) model, constraint weights were determined by a gradual learning acquisition algorithm, where data were presented to the learning model at rates that reflected the frequency of occurrence of forms in the lexicon. These constraint weights were then used to compute a cumulative harmony value for each type of input consonant combination that was compared to O/E values in the data for those consonant combinations. Coetzee and Pater (2008) conducted this analysis for consonant co-occurrence data from Arabic and the Austronesian language Muna. On the basis of differences in the co-occurrence patterns between these two languages, they concluded that language-specific constraint weightings are required for a relatively large number of OCP constraints, resulting in a relatively large number of free parameters. In particular, in order to account for differences between languages in the effects of subsidiary features such as [sonorant] and [continuant] for different places of articulation, Coetzee and Pater (2008) have different constraint weights for OCP-Cor[+son] and OCP-Lab[+son], for example. These differences were motivated by differences between the languages in rate of co-occurrence of consonant pairs that could not be reduced to contrastiveness, demonstrating that the proposals of Pierrehumbert (1993) and Frisch *et al.* (2004) do not generalize to Muna.

As introduced earlier, Wilson and Obdeyn (2009) attempt to account for the language-specific differences found by Coetzee and Pater (2008) with a restrictive model of constraint weighting with fewer free parameters. Wilson and Obdeyn (2009) use the MaxEnt model of Hayes and Wilson (2008). Like the Harmonic Grammar model of Coetzee and Pater (2008), the MaxEnt model uses constraint weights to predict frequency of occurrence, and gathers together cumulative violations to determine relatively high or low frequency of occurrence. Wilson and Obdeyn (2009) conclude from their modeling of Arabic and Muna, as well as data from Shona and Wargamay, that the influence of subsidiary features on consonant co-occurrence within a language can be held constant across different places of articulation, thus reducing the number of free parameters needed by the model (and consequently the number of constraints needed by the grammar). In other words, the impact of the feature [sonorant] is argued to be the same across different places of articulation within a language, though it may differ from one language to the next. Wilson and Obdeyn (2009) argue that apparent differences in the effect of subsidiary features for different places of articulation result from differences in the weights of the place features themselves. For places of articulation where the place feature weight is relatively low (e.g. [coronal] in Arabic), the effect of subsidiary features can be seen in gradient differences in frequency of co-occurrence. For places of articulation where the place feature weight is relatively high (e.g. [labial] in Arabic), co-occurrence is already at very low frequency due to the place feature, and so the effects of the subsidiary features are not visible. Wilson and Obdeyn (2009) argue that differences between their modeling results and those of Coetzee and Pater (2008) come primarily from three differences in their approach to modeling. First, the MaxEnt grammar directly determines constraint weights from frequency data, rather than learning constraint weights indirectly via a learning algorithm. Second, the success of the different models was determined by comparison directly to frequency data, rather than by comparison to O/E values: recall the discussion above on potential issues with O/E data when consonants differ in their base frequency of co-occurrence (and base frequency of occurrence is also modeled by the MaxEnt grammar). Third, the Wilson and Obdeyn (2009) modeling includes a penalty in the evaluation of model success based on the number of free parameters allowed in the model. As a result, they provide an incentive for parsimony in their model fitting. While parsimony is often implicitly part of the discussion in model comparison (e.g. Frisch *et al.* 2004; Coetzee and Pater 2008), their proposal explicitly incorporates a quantitative penalty in the numerical evaluation of a model fit for free parameters.

3.7 *Metalinguistic judgments for OCP-Place*

To test the psychological reality of the OCP-Place constraint, Frisch and Zawaydeh (2001) conducted a novel non-word study with native Arabic speakers in Jordan. Their non-words were constructed with a variety of consonant pairs along the spectrum of violations and non-violations of the OCP-Place constraint. They found that the Arabic speakers were sensitive to OCP-Place violations, including violations of the constraint by more than one consonant pair (i.e. cumulative constraint violations). Frisch and Zawaydeh (2001) also found that Arabic speakers were sensitive to different degrees of OCP-Place violation for a single consonant pair, parallel to the frequency distribution of these consonant pairs in the lexicon. Like the Coleman and Pierrehumbert (1997) data, these findings are difficult to model in a grammar that considers only a single or maximal constraint violation, and are more naturally captured by a grammatical model that considers cumulative violations in determining well-formedness. The cumulative nature of well-formedness is compatible with a grammar that includes frequency information, as the frequency of occurrence of phonological structures reflects the well-formedness of phonological forms within a language in much the same way as typological frequency reflects markedness between languages. In this way, the synchronic frequency effects seen in corpus studies of language lexicons appear to be transparently related to cross-linguistic typological patterns. Thus there is converging evidence that frequency effects at a variety of levels can contribute to our understanding of the nature of language sound structure.

4 Synchronic variation

Synchronic variation refers to cases where there are multiple alternatives for a phonological form. In general, synchronic variation is not expressed by variation between equally frequent alternatives. The cases that have been studied have shown that various phonological and morphological factors bias the output so that the frequency distribution of alternatives is unequal. In addition, there may be any number of non-phonological factors (from syntax or discourse) that influence alternatives. These types of interactions have not been studied to any great extent. Synchronic variation provides evidence that frequency information is part of the phonological knowledge possessed by speakers of a language, as the factors that condition synchronic variation are the same factors that are seen elsewhere in traditional phonological analyses. Providing a completely independent account of synchronic variation from the rest of the (non-variable) grammar seems redundant, as the same phonological factors are used in both cases. The simpler approach, then, is to begin with the same phonological building blocks for both variation and non-variation and provide a mechanism within the grammar that allows for variation to occur. Several mechanisms are discussed, in general developing from less complex to more complex alternatives. The majority of the examples use the framework of Optimality Theory, as the most recent analyses adopt constraints on possible outputs as the fundamental description of phonological patterns.

4.1 Variation as multiple optimality-theoretic grammars

One extensively studied case of synchronic variation is found in so-called *t/d*-deletion in English (e.g. Guy 1991; see also CHAPTER 92: VARIABILITY). Deletion of coda *t/d* is relatively common in English, particularly in dialects such as African American English. The deletion process is more common in consonant clusters, and depends on the following phonological context (among other factors). For example, in African American English deletion is relatively likely when the next word is consonant-initial (11a), or before a pause (11b), and least likely when the next word is vowel-initial (11c).

(11) a. *That's a fast* /fæs/ *car.*
b. *That's fast* /fæs/.
c. *That's a fast* /fæst/ *answer.*

Anttila (2007) proposes a relatively simple account of this systematic variation, using the ranking schema of Optimality Theory. In many ways, the approach is non-quantitative at its core, which makes the analysis of variation compatible with classic generative phonology, in which variation is not considered an important aspect of the theory. Using a set of basic parsing constraints, Anttila (2007) shows that any constraint ranking that allows deletion before a pause or vowel will also allow deletion before a consonant. Thus, the synchronic variation within a language is compatible with the phonological differences between languages (e.g. Davidson 2006). In order to generate variation, we can assume that speakers have multiple grammars with different constraint rankings (also called co-phonologies). If these grammars are used variably, perhaps with different rates, they can be used to model different frequencies of occurrence of *t/d*-deletion. However, the use of multiple constraint rankings does not result in a state where any pattern of variation is possible, as the set of proposed constraints does restrict the possible grammars. In the case of *t/d*-deletion, these constraints ensure that deletion before a consonant is always possible if deletion elsewhere is possible. So deletion before a consonant will always be more frequent than deletion elsewhere.

There are other factors that condition *t/d*-deletion, and their inclusion in grammar has led to more complex models of variation. For example, *t/d*-deletion is sensitive to the preceding phonological context: deletion is more likely in /st/ clusters than /ft/ clusters, more likely following obstruents than sonorants (e.g. *fast vs. pant*) and least likely after a vowel. Deletion is also more likely if *t/d* is part of the root than if *t/d* comes from suffixation, as in the past-tense morpheme (e.g. *fast vs. raced*). There are, however, confounds among the possible conditioning factors. Coetzee (2004) notes that voicing is sometimes mentioned as a conditioning factor, but that voicing is not discussed as an influence on *t/d*-deletion in many studies. However, coda clusters with /d/ are rarely found outside of cases where the /d/ is a suffix or preceded by a sonorant, so it is not clear whether voicing itself is a conditioning factor, or whether these cases of deletion are subsumed by the influence of preceding environment and morphological status. Another alternative analysis is that the effect of the manner of the preceding phoneme is actually an effect of similarity, such that deletion is more likely when *t/d* are next to a highly similar segment (i.e. obstruents are more similar to *t/d* than sonorants and vowels, and similarity also makes some predictions about place of articulation effects that appear to be correct).

4.2 Variation as unranked constraints

Anttila (2008b) takes a slightly different approach to modeling synchronic variation in Optimality Theory in a study of Finnish, in which vowel hiatus is resolved by coalescence of two distinct vowel qualities into a long vowel with a single quality. Vowel coalescence in Finnish is variable, but more common when the first vowel is a mid vowel (12a) than a high vowel (12b).

(12) a. /suomea/ ~ /suomee/ 'Finnish-PARTITIVE'
     b. /ruotsia/ ~ /ruotsii/ 'Swedish-PARTITIVE'

Assuming underlying forms with two distinct vowel qualities, Anttila (2008b) uses three distinct constraints to capture the basic patterns of possible variation: FAITH, *EA, *IA. FAITH preserves the underlying contrast, while *EA and *IA are constraints against vowel hiatus. If *EA is always higher-ranked than *IA, the height asymmetry in vowel coalescence will be preserved. In order to allow for variation, Anttila (2008b) takes some constraints to be unranked in the grammar, to be ranked only upon evaluation and determination of an output (also called floating constraints). For example, if only the universal ranking *EA >> *IA is present, then FAITH can either outrank both of these constraints, leading to no coalescence; only the *IA constraint, leading to coalescence only for /ea/; or neither constraint, leading to coalescence for both /ia/ and /ea/. Due to the asymmetry between *EA and *IA in the typological grammars, there is a predicted asymmetry in the frequency that these patterns are output at evaluation. The full analysis of Anttila (2008b) also includes morphological conditioning involving coalescence in root forms *vs.* coalescence when hiatus is created by a suffix, and also some influences of part of speech on coalescence rate, similar to the *t/d* case. This involves expanding the number of constraints used in the analysis, but the basic approach using some unranked constraints to account for variation is retained.

4.3 Variation through violation of non-crucial constraints

A third Optimality Theory-based approach to variation is presented in Coetzee (2006), where, in contrast to the two discussed above, the ranking of all phonological constraints in the grammar is maintained, but the differences in violations accrued by non-optimal candidates are relevant to frequency. Coetzee (2006) discusses Failense Portuguese, where unstressed vowels are variably deleted. This process is similar in some contexts to unstressed vowel deletion in English (e.g. deletion preceding *r/l* discussed below). It is sensitive to the quality of the unstressed vowel, and also to position within the word. Unlike English, unstressed vowel deletion can occur word-finally (e.g. *paga* /pagɐ̆/ ~ /pag/ 'she pays') and is generally more common in this position. In Coetzee's approach to modeling variability, markedness constraints that are motivated elsewhere are used. These constraints are also used to account for vowel reduction phenomena, thus adding generality to the analysis. These constraints penalize parsing vowels into unstressed syllables, roughly based on their sonority. Coetzee also includes a constraint against unstressed vowels in final position in a prosodic word as a type of markedness constraint against weak prosodic word boundaries. In classical Optimality Theory, faithfulness constraints would outrank these constraints in order for word-final unstressed vowels to appear at all.

Coetzee proposes two extensions to classical Optimality Theory to account for variability. First, non-optimal candidates are ranked relative to one another by the grammar on the basis of their non-crucial violations, so that more marked possibilities are disfavored relative to less marked possibilities overall. If these degrees of non-optimality correspond to frequencies of occurrence, then non-optimal candidates that violate the fewest number of additional constraints will be more frequent, and those which violate many additional constraints will be less frequent. Second, in order to limit the number of possible outputs and allow for categorical effects, there is a boundary in the constraint hierarchy that is a relative cut-off for frequency influences on non-optimal candidates. Constraints above this cut-off have categorical influences on the appearance of alternatives, so that relatively highly ranked constraints will limit the possible outputs qualitatively. Constraints below this cut-off will limit the possible outputs quantitatively, so that variation can be observed when only these relatively low-ranked constraints distinguish alternative outputs. This is the primary distinction between the Coetzee (2006) approach, where there is a cut-off for potential variability in the hierarchy, and the previous alternatives that generate variability with re-ranking. In a re-ranking analysis, constraints anywhere in the hierarchy can potentially be unranked and allow variation.

4.4 Variation by quantitatively ranked constraints

Boersma and Hayes (2001) present a more quantitative approach to optimality-theoretic constraint ranking, which is similar to the unranked constraints approach. In their theory, Stochastic Optimality Theory, each constraint has an associated average ranking value that represents its typical location within the constraint hierarchy. When constraint evaluation takes place to generate an output, each constraint is perturbed from its average ranking value probabilistically, so that a constraint at evaluation may be at a higher or lower point than its average location. If two constraints are relatively close together in average ranking value, it is possible that their relative position will change with respect to one another (and, if they had the exact same average ranking value, they are essentially unranked with respect to one another, and could appear in either order at evaluation). Constraints that are relatively far apart in average ranking value are very unlikely to occur in a different ranking at evaluation. In this way, Stochastic Optimality Theory can have relatively categorical constraint rankings and relatively variable constraint rankings, with the potential to fine-tune the average ranking values so that the relative ranking of two constraints at evaluation occurs with any desired frequency. With an appropriate choice of constraints, then, frequency effects can be modeled rather precisely. Criticism of Stochastic Optimality Theory has noted that it can produce no truly categorical effects, as it is always possible in any particular evaluation for the ranking value for a constraint to be very distant from its average ranking value; however, this possibility is extremely unlikely (e.g. Anttila 2008b). It has also been suggested that tuning the frequencies predicted by a grammar too closely to a corpus of data is unrealistic as a model of the knowledge of any individual speaker, as there is variation between individuals and there are extra-grammatical influences on the frequency of occurrence of variant forms. Coetzee (2006) argues that a better model of variability in grammar only provides relative frequency differences between alternatives and does not model exact frequencies of occurrence.

4.5 *Token frequency as an influence in variation*

Word token frequency has been shown to be an influence on phonetic reduction and also on the diachronic spread of sound change. Though these processes are arguably diachronic, the synchronic language learner acquires and uses a system in which there is synchronic variation. There are numerous studies documenting these types of changes (see Bybee 2000 for a review). One of the earliest studies of this type examined deletion of unstressed /ə/ preceding a sonorant /r/ or /l/. In low-frequency words, such as *artillery* and *cursory*, unstressed /ə/ is preserved (i.e. /artıləri/ not /artılri/). In high-frequency words, such as *memory*, *salary*, and *nursery*, unstressed /ə/ may be deleted or replaced with a syllabic sonorant making syllable count in these words somewhat ambiguous. Bybee (2000) argues that this variation is the result of a sound change in progress that is gradual both in its diffusion through the lexicon and in its phonetics for any particular word.

A second example of this type of reduction is *t/d*-deletion, mentioned above. The deletion of *t/d* is conditioned by a variety of social and grammatical factors (see Guy 1991). One potential grammatical factor is the distinction between function words and content words. Function words have a higher rate of *t/d*-deletion. However, function words are also highly frequent, so this potential grammatical factor is hopelessly confounded with frequency. Another potential grammatical factor is a decrease in the frequency of *t/d*-deletion when *t/d* is used as a regular past-tense marker *vs.* ordinary stem-final *t/d* in content words. Taking this factor into account, Jurafsky *et al.* (2001) show that frequency has an influence on deletion rate in content words. In a study of the phonetically transcribed Switchboard database, they found that content words with the highest frequency were twice as likely to have final *t/d*-deletion as content words with the lowest frequency. They found word frequency to be a significant factor above and beyond other potential phonological (e.g. word length) and morphophonological influences on reduction.

Bybee (2000, 2002) proposes that the other morphophonological influences on *t/d*-deletion may also reduce to frequency effects. In Bybee's model, words are represented in the lexicon in full phonetic detail, with gradient degrees of reduction being part of the idiosyncratic knowledge that individual speakers have about a word. Moreover, word usage leads to additional phonetic reduction, so that high-frequency words are more phonetically reduced than low-frequency words. As this gradual reduction proceeds to the eventual endpoint of deletion, higher-frequency words are more likely to reach this state in comparison to low-frequency words. The gradual erosion accounts for frequency effects on the reduction and deletion of unstressed /ə/ as well as *t/d*-deletion. Bybee's model also includes listing of morphologically complex forms in the lexicon, so these forms are also subject to the gradual erosion through usage that is found in simple words. Bybee claims that the supposed difference in deletion rate between morphologically complex words and simple words is actually coincidental, as morphologically complex words are more likely to contain a vowel preceding the *t/d* in comparison to simple words, and the post-vocalic environment is less favorable to *t/d*-deletion (e.g. *missed vs. said*). Finally, she notes that speech production is flexible, allowing a non-reduced form to be produced if needed. Such a form might be more likely to be used for a morphologically complex word, where the presence of a suffix conveys potentially distinctive information, in comparison to a morphologically simple word, where other parts of the word may be sufficient to indicate the lexical identity.

5 Summary

A review of the evidence for occurrence frequency having an influence on phonological and morphophonological patterns finds support for frequency as an influence on language sound structure at a variety of levels in a variety of ways. Corpus studies based on the dictionary as a model of the mental lexicon have found systematic variation in phonotactic type frequency that influence native-speaker judgments in novel non-word processing tasks. Corpus studies of spoken language corpora have found that speech production is influenced by type and token frequency that can also be observed in novel word production experiments. Theoretical models of frequency effects in grammar are being developed, and this is an area of growing research interest in phonology and other areas of grammar (e.g. Bod *et al.* 2003). Combined with laboratory research techniques, these models provide new opportunities to understand language sound structure in the grammar, as well as in language use.

In the view presented in this chapter, language sound structure is built up from language use, starting with the phonetic level. Phonemic categories are generalizations over phonetic patterns; sub-syllabic and syllabic categories are generalizations over phonemic patterns; and lexical and morphophonological categories are generalizations over phonemic, sub-syllabic, and syllabic patterns. Basic abstractions at each level (e.g. segments at the phonemic level) provide material for abstraction and generalization at higher levels. As models and theories of language sound structure bridge the gap between abstract symbolic knowledge representations and actual occurrence in language use, they also provide new opportunities to integrate the study of phonology with the study of phonetics, sociolinguistics, and psycholinguistics. In these other domains, frequency effects are well-known and well-documented. Advances in theoretical phonology to model frequency effects may also be useful to researchers in these fields, allowing them to model their data with more sophisticated phonological representations (Dell 2000). This potential for interdisciplinary exploitation of probabilistic models of phonology is likely to make for the most robust models of phonological phenomena that provide a real understanding of how language sound structure is represented in the mind/brain.

REFERENCES

Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26. 9–41.

Albright, Adam & Bruce Hayes. 2003. Rules *vs.* analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119–161.

Anttila, Arto. 2007. Variation and optionality. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 519–536. Cambridge: Cambridge University Press.

Anttila, Arto. 2008a. Gradient phonotactics and the Complexity Hypothesis. *Natural Language and Linguistic Theory* 26. 695–729.

Anttila, Arto. 2008b. Derived environment effects in colloquial Helsinki Finnish. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Essays in honor of Paul Kiparsky*, 433–460. Cambridge, MA: MIT Press.

Bailey, Todd M. & Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44. 568–591.

Beckman, Mary E. & Jan Edwards. 2010. Generalizing over lexicons to predict consonant mastery. *Laboratory Phonology* 1. 319–343.

Beddor, Patrice Speeter, James D. Harnsberger & Stephanie Lindemann. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics* 30. 591–627.

Berkley, Deborah M. 2000. Gradient Obligatory Contour Principle effects. Ph.D. dissertation, Northwestern University.

Best, Catherine T. & Gerald W. McRoberts. 2003. Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech* 46. 183–216.

Bod, Rens, Jennifer Hay & Stefanie Jannedy (eds.) 2003. *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45–86.

Bolinger, Dwight L. 1961. *Generality, gradience, and the all-or-none*. The Hague: Mouton.

Broe, Michael B. & Janet B. Pierrehumbert (eds.) 2000. *Papers in laboratory phonology V: Acquisition and the lexicon*. Cambridge: Cambridge University Press.

Buckley, Eugene. 1997. Tigrinya root consonants and the OCP. *Penn Working Papers in Linguistics* 4. 19–51.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10. 425–455.

Bybee, Joan. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In Michael Barlow & Suzanne Kemmer (eds.) *Usage-based models of language*, 65–85. Stanford: CSLI.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14. 261–290.

Bybee, Joan. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Bybee, Joan & Paul Hopper (eds.) 2001. *Frequency and the emergence of linguistic structure*. Amsterdam & Philadelphia: John Benjamins.

Clumeck, Harold. 1976. Patterns of soft palate movements in six languages. *Journal of Phonetics* 4. 337–351.

Coetzee, Andries W. 2004. What it means to be a loser: Non-optimal candidates in Optimality Theory. Ph.D. dissertation, University of Massachusetts, Amherst.

Coetzee, Andries W. 2006. Variation as accessing "non-optimal" candidates. *Phonology* 23. 337–385.

Coetzee, Andries W. & Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26. 289–337.

Coleman, John & Janet B. Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In John Coleman (ed.) *Proceedings of the 3rd Meeting of the ACL Special Interest Group in Computational Phonology (SIGPHON 3)*, 49–56. Somerset, NJ: Association for Computational Linguistics.

Davidson, Lisa. 2006. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34. 104–137.

Davis, Stuart. 1991. Coronals and the phonotactics of nonadjacent consonants in English. In Carole Paradis & Jean-François Prunet (eds.) *The special status of coronals: Internal and external evidence*, 49–60. San Diego: Academic Press.

Dell, Gary S. 2000. Commentary: Counting, connectionism, and lexical representation. In Broe & Pierrehumbert (2000), 335–348.

Ernestus, Mirjam & R. Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79. 5–38.

Frisch, Stefan A. 2000. Temporally organized representations as phonological units. In Broe & Pierrehumbert (2000), 283–298.

Frisch, Stefan A. & María R. Brea-Spahn. Forthcoming. Metalinguistic judgements of phonotactics by monolinguals and bilinguals. *Laboratory Phonology* 1.

Frisch, Stefan A. & Adrienne M. Stearns. 2006. Linguistic and metalinguistic tasks in phonology: Methods and findings. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.) *Gradience in grammar: Generative perspectives*, 70–84. Oxford: Oxford University Press.

Frisch, Stefan A. & Bushra A. Zawaydeh. 2001. The psychological reality of OCP Place in Arabic. *Language* 77. 91–106.

Frisch, Stefan A., Nathan R. Large & David B. Pisoni. 2000. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42. 481–496.

Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22. 179–228.

Goldinger, Stephen D., Paul A. Luce & David B. Pisoni. 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language* 28. 501–518.

Greenberg, Joseph H. 1950. The patterning of root morphemes in Semitic. *Word* 6. 162–181.

Guy, Gregory R. 1991. Contextual conditioning in variable lexical phonology. *Language Variation and Change* 3. 223–239.

Hammond, Michael. 1999. *The phonology of English: A prosodic optimality-theoretic approach.* Oxford: Oxford University Press.

Hay, Jennifer & R. Harald Baayen. 2002. Parsing and productivity. *Yearbook of Morphology* 2001. 203–236.

Hayes, Bruce & Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440.

Hayes, Bruce, Kie Zuraw, Péter Siptár & Zsuzsa Cziráky Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 82. 822–863.

Herdan, Gustav. 1962. *The calculus of linguistic observations.* The Hague: Mouton.

Itô, Junko & Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17. 49–73.

Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee & Hopper (2001), 229–254.

Keating, Patricia (ed.) 1994. *Phonological structure and phonetic form: Papers in laboratory phonology III.* Cambridge: Cambridge University Press.

Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens & Björn Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255. 606–608.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky. 1990. Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness – theoretical foundations. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, 388–395. Mahwah, NJ: Lawrence Erlbaum.

Luce, Paul A. & David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19. 1–36.

MacEachern, Margaret R. 1999. *Laryngeal cooccurrence restrictions.* New York: Garland.

Manuel, Sharon Y. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America* 88. 1286–1298.

Martindale, Colin, S. M. Gusein-Zade, Dean Mckenzie & Mark Y. Borodovsky. 1996. Comparison of equations describing the ranked frequency distributions of phonemes and graphemes. *Journal of Quantitative Linguistics* 3. 106–112.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Keating (1994), 191–233.

Mester, Armin. 1986. Studies in tier structure. Ph.D. dissertation, University of Massachusetts, Amherst.

Munson, Benjamin. 2001. Phonological pattern frequency and speech production in children and adults. *Journal of Speech, Language, and Hearing Research* 44. 778–792.

Padgett, Jaye. 1995. *Stricture in feature geometry*. Stanford: CLSI.

Pierrehumbert, Janet B. 1993. Dissimilarity in the Arabic verbal roots. *Papers from the Annual Meeting of the North East Linguistic Society* 23. 367–381.

Pierrehumbert, Janet B. 1994. Syllable structure and word structure: A study of triconsonantal clusters in English. In Keating (1994), 168–188.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee & Hopper (2001), 137–157.

Pierrehumbert, Janet B. 2003. Probabilistic phonology: Discrimination and robustness. In Bod *et al.* (2003), 177–228.

Pierrehumbert, Janet B., Mary E. Beckman & D. Robert Ladd. 2000. Conceptual foundations of phonology as a laboratory science. In Noel Burton-Roberts, Philip Carr & Gerard J. Docherty (eds.) *Phonological knowledge: Conceptual and empirical issues*, 273–303. Oxford: Oxford University Press.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Ringen, Catherine & Robert M. Vago. 1998. Hungarian vowel harmony in Optimality Theory. *Phonology* 15. 393–416.

Shademan, Shabnam. 2006. Is phonotactic knowledge grammatical knowledge? *Proceedings of the West Coast Conference on Formal Linguistics* 25. 371–379.

Siptár, Péter & Miklós Törkenczy. 2000. *The phonology of Hungarian*. Oxford: Oxford University Press.

Smolensky, Paul & Géraldine Legendre (eds.) 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, MA: MIT Press.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT.

Tambovtsev, Yuri & Colin Martindale. 2007. Phoneme frequencies follow a Yule distribution. *SKASE Journal of Theoretical Linguistics* 4(2). 1–11.

Vitevitch, Michael S. 2002. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 28. 735–747.

Werker, Janet F. & Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7. 49–63.

Wilson, Colin & Marieke Obdeyn. 2009. Simplifying subsidiary theory: Statistical evidence from Arabic, Muna, Shona, and Wargamay. Unpublished ms., Johns Hopkins University.

Yip, Moira. 1989. Feature geometry and cooccurrence restrictions. *Phonology* 6. 349–374.

Zipf, George K. 1965. *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: MIT Press. First published 1935, Boston: Houghton Mifflin.

Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog infixation. *Language* 83. 277–316.

91 Vowel Harmony: Opaque and Transparent Vowels

Adamantios I. Gafos
Amanda Dye

1 Introduction

In linguistic theory, there is a tension between two core aims of the field: theoretical parsimony and empirical coverage of the remarkable diversity and specificity seen in linguistic data. Vowel harmony, and more specifically the two phenomena of transparency and opacity in vowel harmony that concern us in this chapter, provide prime examples of phonological research concerned with resolving this tension.[1] Recent findings and promising new theoretical ideas on transparency and opacity indicate that our understanding of these phenomena is at a critical stage. The crucial future step will be the integration of the new data patterns with insights from theoretical analyses. This will require an improved understanding of the phonetic basis of vowel harmony, of the nature and origin of the grammatical forces governing transparency and opacity, and finally of the organizational principles governing the resolution of these forces within and across languages.

2 Definitions and exemplification of the basic patterns

Vowel harmony is a regularity found in many languages requiring vowels in certain grammatical domains to agree in terms of specific phonological features. For example, the vowels of Hungarian can be divided into two subsets, the "front" vowels /i iː e eː ø øː y yː/ and the "back" vowels /u uː o oː ɔ ɑː/ (Vago 1980a; van der Hulst 1985; Siptár and Törkenczy 2000; see also CHAPTER 123: HUNGARIAN VOWEL HARMONY). In terms of phonological features, the vowels in the "front" set share the feature [−back] and those in the "back" set share the feature [+back]. The existence of vowel harmony is most readily observed in

[1] For excellent reviews of vowel harmony, see van der Hulst and van der Weijer (1995) and Archangeli and Pulleyblank (2007). See Gordon (2006) for a short article on the phonetics of transparency in vowel harmony. At an introductory level, a cogent discussion of different proposals on vowel harmony can be found in Kenstowicz (1994). For recent dissertation-length treatments of vowel harmony, see Baković (2000), Krämer (2003), and Przezdziecki (2005).

suffix vowel alternations where the [±back] quality of the suffix vowel is determined by the [±back] quality of the stem vowel. For example, as shown in (1), the dative suffix alternates between two forms, one with the front vowel /e/ and another with the back vowel /a/, as a function of the stem vowel: *ház-nak* 'house-DAT', *város-nak* 'city-DAT', but *kéz-nek* 'hand-DAT', *öröm-nek* 'joy-DAT' (in Hungarian orthography, the acute accent denotes a long vowel, the umlaut denotes a front round vowel, and the double acute indicates a long front round vowel). In such combinations of stems and suffixes, then, the suffixes adjust or harmonize to the [±back] "harmonic feature" of the stem vowels. Because it determines the suffix form, the stem vowel is called the trigger and the suffix vowel is called the target of the harmony pattern.

(1) *Regular vowel harmony*

a. *stem vowels are front* ([−back])

vidék-től	[videːktøːl]	'country-ABL'
öröm-nek	[ørømnɛk]	'joy-DAT'
hegedű-nél	[hɛgɛdyːneːl]	'violin-ADESS'
víz-ben	[viːzbɛn]	'water-INESS'

b. *stem vowels are back* ([+back])

város-tól	[vaːroʃtoːl]	'town-ABL'
mókus-nak	[moːkuʃnɔk]	'squirrel-DAT'
harang-nál	[hɔrɔŋgnaːl]	'bell-ADESS'
ház-ban	[haːzbɔn]	'house-INESS'

A central aim in phonological theory has been to characterize the range and possible forms of such systematic sound patterns in different languages (Chomsky and Halle 1968). A particularly fruitful research strategy in this regard has been to explore the *phonetic basis* of sound patterns, i.e. the extent to which such patterns can be seen as adaptations to biological constraints on speech production and perception (Lindblom 1983; cf. Anderson 1981). For vowel harmony, it has been proposed that a natural basis for it can be traced to the low-level phonetic influences among vowels in consecutive syllables (Fowler 1983; Ohala 1994a, 1994b; Beddor *et al.* 2002; Przezdziecki 2005). The crucial fact is that vowels exert influences on neighboring vowels across intervening consonants, the so-called *V-to-V co-articulation* (Öhman 1966). Acoustically, this means that in a V_1CV_2 sequence, V_2's formant values are influenced by the identity of V_1 and vice versa. Articulatorily, the basis of this phenomenon is well understood. Vowels are produced by movements of the tongue body and the jaw, with relatively slow changes in the global shape of the vocal tract, because of the larger mass involved. In the case of a V_1CV_2 sequence, while the consonantal constriction is localized at some area in the vocal tract, the free part of the tongue is allowed to assume different shapes and thus move in a "smooth transition" (Öhman 1966) from the shape of the first to the shape of the second vowel. Vowels, in other words, provide a continuous articulatory substrate on which consonants are superimposed. V-to-V co-articulation is a direct consequence of this articulatory continuity. This phonetic effect can develop, via processes of phonologization, into vowel harmony (Przezdziecki 2005).

However, within the attested range of phonological vowel harmony patterns there exist phenomena that are systematic, yet unexpected given the phonetic grounding sketched above. Specifically in Hungarian, but also in other languages with vowel harmony, we find as well a set of "neutral" vowels, which do not undergo harmony. Neutral vowels come in two species: transparent and opaque. The best known property of transparent vowels is that they may intervene between the trigger and the target vowel even when they bear the opposite value for the harmonizing feature. The examples in (2) show that the first stem vowel dictates the backness value for the suffix vowel across the intervening transparent vowels *í* [iː], *i* [i], and *é* [eː]. In (2a), stems with initial front vowels followed by a transparent vowel select front suffixes. In (2b), however, the initial stem vowel and the suffix vowel are back, despite the front quality of the intervening transparent vowel. Hence the initial and suffix vowels seem to establish a harmony relationship across both consonants and transparent vowels.

(2) *Transparent vowels: Agreement across the medial vowel*

a.	*first stem vowel is front*		
	emír-nek	[ɛmiːrnɛk]	'emir-DAT'
	zefír-ből	[zɛfiːrbøːl]	'zephyr-ELAT'
	bili-vel	[bilivɛl]	'pot-INSTR'
	művész-nek	[myːveːsnɛk]	'artist-DAT'
b.	*first stem vowel is back*		
	papír-nak	[pɔpiːrnɔk]	'paper-DAT'
	zafír-ból	[zɔfiːrboːl]	'sapphire-ELAT'
	buli-val	[bulivɔl]	'party-INSTR'
	kávé-nak	[kaːveːnɔk]	'coffee-DAT'

The other species of neutral vowels, opaque vowels, block harmony by overriding the expected consequences of another potential trigger in the phonological form. For example, whereas the front unround Hungarian vowels are transparent, front round vowels in Hungarian and other palatal vowel harmony systems are opaque, e.g. *papír-nak* 'paper-DAT' *vs. parfüm-nek* 'perfume-DAT'. In the latter form, the front round *ü* [y] initiates its own harmony domain, imposing a [−back] specification on its following suffix. Thus, [y] is opaque but [iː] is transparent.

Whereas the existence of opaque vowels is consistent with the generally accepted idea that the phonetic basis of vowel harmony inheres in V-to-V co-articulation effects, transparent vowels pose a challenge to that idea. Consider, for example, the difference in the suffix vowels between words like *papír-nak* and *zefír-nek*. The backness of the suffix in *papír-nak* must be linked to the backness of the stem-initial vowel. However, it is not clear how this can be achieved through the acoustic consequences of V-to-V co-articulation, given that the stem-initial and suffix vowels are not adjacent. Although long-distance co-articulation across schwa has been found in English (Magen 1997), studies also show that [iː] is resistant to co-articulation from the preceding vowel(s) in terms of perception (e.g. Magen 1984 for Japanese; Farnetani *et al.* 1985 for Italian; Recasens 1987 for Spanish and Catalan). As we will see in §3, because of this long-distance character, transparency has been a recalcitrant problem in theoretical treatments of the phenomenon.

A notable generalization in "palatal" vowel harmony systems is the existence of a relation between vowel height and transparency. In Hungarian, stems in which a back vowel is followed by [ɛ] are commonly described as "vacillating," because they allow both front and back suffixes (Vago 1980a).[2] Hence, stems such as *hotel* vacillate: *hotel-nak/nek* 'hotel-DAT'; but stems such as *papír* do not, *papir-nak*, **papír-nek*. Thus, the generalization is that the lower and more retracted [ɛ] is phonologically less transparent than the higher and more front [i] (see Beňuš 2005 for the first characterizations of these vowels based on ultrasound and electromagnetic articulometry data). A similar generalization is true for other palatal vowel harmony systems (L. B. Anderson 1980; Beňuš 2005). As far as we know, the opposite generalization, that lower vowels are more transparent than higher vowels, is not attested.

Another notable generalization concerns the relationship between suffix choice and the number of transparent vowels in stems: stems where a back vowel is followed by two transparent vowels are more likely to vacillate or take front suffixes than stems where a back vowel is followed by only one transparent vowel. For example, *mam-i* and *mam-csi* (both 'mother-DIM') select back suffixes: *mami-nak*, *mamcsi-nak* 'mother-DIM-DAT'. However, when the two diminutive suffixes are combined in *mamicsi*, both front and back suffixes are acceptable: *mamicsi-nak/nek* 'mother-DAT' (Farkas and Beddor 1987; Ringen and Kontra 1989; Beňuš 2005; Hayes and Londe 2006). Beňuš (2005), highlighting this aspect of the Hungarian data, concludes that transparency is not a categorical property of vowels but it is determined contextually. The same vowel can be transparent in one context (e.g. *mamcsi-nak*) but opaque in another (e.g. *mamicsi-nek*).

Our examples of transparency and opacity so far come from a so-called *"palatal" vowel harmony* system where, following standard description, the phonological feature exhibiting harmonic behavior is [±back]. Transparency and opacity can be instantiated regardless of the vowel feature exhibiting harmony. In what follows, we exemplify the other major classes of harmony: "tongue root" harmony, "rounding" harmony, and "height" harmony.

2.1 "Tongue root" harmony

In languages described as having "tongue root" harmony, vowels harmonize for features that correspond to the position of the tongue root or pharyngeal expansion/compression. In these languages, following common phonological description, vowels may be "advanced," i.e. articulated with the tongue root in an advanced position, or "retracted," i.e. articulated with a non-advanced or retracted tongue root. The relevant phonological dimension of distinction, it has been proposed, is [±ATR] (originally from Halle and Stevens 1969) with advanced vowels sharing [+ATR] and retracted vowels [−ATR]. This has come to be the standard way of describing the phonetic basis and the phonological feature in question. But work by Lindau (1978), using radiographic and acoustic data, and by Tiede (1996), using MRI data, indicates that the relevant difference is in terms of pharyngeal expansion *vs.* compression, which can be achieved in different ways, one of which is by positioning the tongue root in the way implicated by [±ATR]. At a first approximation, however, we can talk about patterns of transparency and opacity in "tongue root" harmony without being precise about the phonetic dimensions involved.

[2] This statement is a first approximation to a set of complex data patterns. Farkas and Beddor (1987), Kaun (1995), Beňuš (2005), Beňuš and Gafos (2007), and Hayes and Londe (2006) provide details, data, and relevant discussion on this height–transparency link in Hungarian.

Our first example of "tongue root" harmony comes from Akan (Ladefoged 1964; Schachter and Fromkin 1968; Lindau 1978; Dolphyne 1988). The Akan verb /di/ 'to eat', which contains a [+ATR] vowel, imposes that value on a personal pronoun prefix, so that we get forms such as [mi-di] 'I eat' and [wu-di] 'you eat'. By contrast, the verb /dɪ/ 'to be called' imposes its [−ATR] value to its prefix, so that the same prefixes surface as [mɪ-dɪ] 'I am called' and [wʊ-dɪ] 'you are called'.

In Akan, both high and mid vowels are subject to harmony, but in other "tongue root" harmony languages the vowels that alternate comprise a more restricted set. Thus, in Wolof, only the mid vowels /e ɛ o ɔ/ alternate in harmony; for example, the past tense suffix appears as [-ɔɔn] with a [−ATR] stem, as in [rɛɛr-ɔɔn] 'had dinner', and as [-oon] with a [+ATR] stem, as in [reer-oon] 'was lost'.

In Wolof, the high vowels are neutral (Archangeli and Pulleyblank 1994). In particular, high vowels in Wolof are transparent to vowel harmony.[3] In (3), the initial stem vowels determine the [ATR] values of the suffix vowels across the medial transparent vowels. In (3a), stems with initial [+ATR] vowels take [+ATR] suffixes. In these cases the intervening high vowel has the same [ATR] value as the initial vowel, but in (3b) the opposite is true. In the forms on the right, stems with [−ATR] vowels take [−ATR] suffixes, despite the fact that the intervening high vowel has the opposite feature value for ATR.

(3) *Transparent vowels in Wolof* (Archangeli and Pulleyblank 1994)

a.	toxi-leen	'go and smoke'	b.	tɛkki-lɛɛn	'untie!'
	gəstu-leen	'do research!'		mɔytu-lɛɛn	'avoid!'
	təriji-leen	'go sleep!'		sɔppiwu-lɛɛn	'you have not changed'
	seenu-woon	'tried to spot'		tɛɛru-wɔɔn	'welcomed'

Like high vowels, low vowels may also be neutral. In Pulaar, for instance, low vowels are opaque to vowel harmony. Pulaar shows regressive harmony of mid vowels, such that the [ATR] value of suffixes migrates to stems, as in [peetʃ-i] / [pɛɛtʃ-ɔn] 'crack-PL/DIM PL', where the plural suffix /-i/ causes the stem to surface with a [+ATR] vowel, while the diminutive plural suffix /-ɔn/ imposes its [−ATR] value on the stem. The behavior of the opaque low vowel /a/ is exemplified in (4).

(4) *Opaque low vowels in Pulaar* (Archangeli and Pulleyblank 1994)

bɔɔt-aa-ri	*boot-aa-ri	'dinner'
pɔɔf-aa-li	*poof-aa-li	'breaths'
nɔdd-aa-l	*nodd-aa-li	'call'
ᵑgɔr-aa-gu	*ᵑgor-aa-gu	'courage'

In (4), the [+ATR] value of the rightmost suffix cannot spread to the low vowel /a/, which surfaces as [−ATR] everywhere in Pulaar. The low vowel not only does not accept the feature value of the suffix to its right, it also blocks vowel harmony from propagating to the stem to its left. The vowel starts its own harmonic

[3] High-vowel-initial roots are always harmonic (for [+ATR]) in Wolof, but this is the only circumstance in which high vowels trigger harmony.

domain, as shown by the starred examples; when a stem appears to the left of /a/, it cannot contain a [+ATR] vowel.

Archangeli and Pulleyblank (1994) discuss plausible phonetic bases for the behavior of high vowels in languages like Wolof and low vowels in languages like Pulaar. Tongue body raising, which is required for [+high] vowels, and tongue root advancement, which is required for [+ATR] vowels, are compatible gestures (CHAPTER 21: VOWEL HEIGHT). Both serve to lower F1, producing the percept of a higher vowel. In the same way, lowering of the tongue body (for [+low] vowels) and retracting the tongue root are compatible, and in concert they help create a perceptually lower vowel by producing a higher F1. Tongue body raising/tongue root retraction and tongue body lowering/tongue root advancement, however, can be said to constitute pairs of incompatible gestures, because one member of each pair serves to counteract the acoustic effect of the other. This phonetic basis provides plausible reasons for the lack of [+high, −ATR] vowels contrasting with transparent vowels in Wolof, and [+low, +ATR] vowels contrasting with the opaque vowels in Pulaar. Further, Archangeli and Pulleyblank (1994) propose that such lack of contrast (CHAPTER 2: CONTRAST) may underlie the neutrality of these vowels. As is often the case with phonetic motivations for phonological patterns (Anderson 1981), identifying plausible phonetic grounds for neutrality does not fully explain the behavior of neutral vowels in these harmony systems. The fundamental question is: what principles determine when a vowel is transparent or opaque? Consider the case of Wolof, in which high vowels are transparent to vowel harmony, but where, in some circumstances, the low vowel may be opaque.[4] There exist plausible phonetic grounds for saying that both [+high, −ATR] and [+low, +ATR] vowels are marked. But it remains unclear how exactly these similar markedness considerations are translated into the different phonological behavior that these vowels seem to exhibit. The reader is referred to Archangeli and Pulleyblank (1994), who discuss cases of cross-linguistically variable behavior of [+high, −ATR] and [+low, +ATR] vowels.

We can highlight the issue of how deterministic phonetic forces may be of phonological behavior when we also consider that what appears to be the same vowel may be transparent in some languages but opaque in others. For example, in Pulaar, as described above, and in Tangale (Jungraithmayr 1971; Kidda 1985; van der Hulst and van de Weijer 1995) the low vowel /a/ is opaque, whereas in Kinande (Schlindwein 1987; Steriade 1987; Cole and Kisseberth 1994; Mutaka 1994) /a/ has traditionally been analyzed as transparent to vowel harmony. Predicting the transparency *vs.* opacity of /a/ in these languages cannot be reduced to a question of phonological contrast, since /a/ has no phonological [+ATR] counterpart in any of the relevant languages. On the surface, then, the cross-linguistically variable behavior of neutral /a/ seems to present a puzzle. However, recent articulatory work (Beňuš 2005; Gick *et al.* 2006) suggests that the real question may not be one of opacity *vs.* transparency, but rather of opacity *vs.* participation in harmony: experimental data show that /a/ in Kinande is affected by vowel harmony. It should be noted that, like low vowels, neutral high vowels show cross-linguistically variable behavior in ATR harmony. For instance, high vowels are transparent to

[4] The long low vowel /aː/ is always opaque in Wolof, but short /a/ usually participates in vowel harmony. There are exceptions to the behavior of short /a/, however, in which it is also opaque. See Archangeli and Pulleyblank (1994) for a thorough discussion.

harmony in Wolof, as discussed above, but are opaque in Yoruba (Archangeli and Pulleyblank 1994; see Pulleyblank 1996 for discussion).

2.2 *"Rounding" harmony*

In various languages, including those of the Turkic, Mongolian, and Tungusic families, as well as Yawelmani (now called Yokumne), an American Indian language of California, vowels harmonize for lip posture (CHAPTER 19: VOWEL PLACE). As we remarked in the case of "tongue root" harmony, the phonetic basis of the harmonizing "rounding" feature is a separate, non-trivial issue (see Goldstein 1991). With this in mind, we will follow standard descriptions in saying that the feature showing harmony in these languages is [±round]. In Turkish "rounding" harmony, both high and non-high vowels can trigger rounding harmony, but only high vowels are eligible as targets. The 1st person singular possessive suffix, for example, alternates depending on the value of the stem for the feature [round], as *ip-im* 'my rope' *vs.* *süt-üm* 'my milk'. See CHAPTER 118: TURKISH VOWEL HARMONY for extensive discussion.

A well-known case of transparency in rounding harmony is observed in Khalkha Mongolian. In (5a) we see the [+round] feature of the stem propagating across /i/ to the (rightmost) suffix vowel (data from Svantesson *et al.* 2005). The examples in (5b) show that these suffix vowels surface as [−round] when the initial stem vowel is [−round].

(5) *Transparent* [i] *in the rounding harmony of Khalkha Mongolian*

a.	poor-ig-o	'kidney-ACC-RFL'	b.	piir-ig-e	'brush-ACC-RFL'
	xɔɔʤ-ig-ɔ	'food-ACC-RFL'		teeʤ-ig-e	'gown-ACC-RFL'
	ɔʤʲi-ʤɔ	'to squint-DPST'		tʰɛ.xʲi-ʤa	'to be bent-DPST'

Previous theoretical discussions of Mongolian rounding harmony (S. R. Anderson 1980; Steriade 1995) have considered the transparency of /i/ to be problematic. As we discuss in the following section, a popular account of transparency in vowel harmony maintains that the harmonizing feature, in the present case [+round], is first transmitted to all vowels in the word. For example, in [poor-ig-o], according to this account, the transparent vowel /i/ would be assigned the feature [+round]. This is a standard way to express formally the intuition that the rounding of the suffix vowel in [poor-ig-o] is due to the rounding of the first stem vowel, while respecting one of the fundamental phonological principles that such dependencies among vowels in words are local and hence rounding must propagate also to the intervening /i/. As a result of this step, however, the intermediate representation obtained is [poor-yg-o], where /i/ has changed to /y/. In order to derive the correct surface form, this intermediate form would need to be repaired via a rule like /y/ → [i]. However, such a rule would not be tenable in the case of Mongolian, because /y/ is thought to be a phoneme of the language, and there is no principled way to distinguish derived [y] from underlying /y/ in the application of such rules. But acoustic data from Rialland and Redouanne (1984), as well as from Svantesson *et al.* (2005), show that the phoneme notated with /y/ in past analyses is not a front vowel. Therefore, /i/ and /y/ are not phonemically contrastive in this language. Rather, the data show that the purported [y] is a back vowel. In light of the phonetic evidence, then, this case of transparency is not unique as it was thought to be. As we have seen in the cases of transparency in "palatal" and "tongue root" harmony systems reviewed so far, the transparent vowels do not contrast with respect to the harmonizing feature, e.g. /i/ in Hungarian does not have a back counterpart, /i/ in Wolof does not have a [–ATR] counterpart, and /a/ in Kinande does not have a [+ATR] counterpart.

2.3 *"Height" harmony*

In phonological terms, "height" harmony systems require uniformity across vowels in terms of the phonological feature [±high] (CHAPTER 21: VOWEL HEIGHT). Some instances of what have been called "height" harmony systems have been an issue of discussion in connection with "tongue root" harmony. This issue arises because tongue body vertical position, an undisputed correlate of phonological height, and pharyngeal size have very similar effects in the frequency of the first formant – see Lindau (1978) for a particularly clear illustration of this from Akan. As an expected consequence, then, in some harmony systems, such as that of Kinande, the harmonic classes seem to differ according to vowel height as well as tongue root position. Clements (1990, 1991) proposes an account for such systems in terms of an "aperture" theory of vowel height. In his theory, the harmonic classes are distinguished on the basis of their value for the feature [open], used to implement height distinctions. More recent discussion of the same issue, "tongue root" *vs.* "height" harmony, can be found in the instrumental studies of Gick *et al.* (2006) and Kenstowicz (2009), who provide converging evidence in favor of the "tongue root" interpretation for the Kinande vowel harmony system.

In part because of this issue, the "height" harmony systems in dialects of Italian (CHAPTER 110: METAPHONY IN ROMANCE) have not received an interpretation in terms of "tongue root" harmony and thus can be considered as good representatives of "height" harmony (see van der Hulst and van de Weijer 1995 for further relevant discussion of the "tongue root" *vs.* "height" harmony issue). This position must be qualified as tentative. The phonetic bases of the distinctions implicated in the Italian vowel harmony systems have not been the subject of systematic studies.

3 Previous accounts

Transparency and opacity have provided persistent challenges to phonological theory. A fundamental understanding of these phenomena implicates long-standing phonological issues: the range of possible rules or constraints (CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS), the nature of representations (CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATIONS; CHAPTER 19: VOWEL PLACE; CHAPTER 21: VOWEL HEIGHT; CHAPTER 27: THE ORGANIZATION OF FEATURES), the central issue of locality (CHAPTER 81: LOCAL ASSIMILATION), markedness (CHAPTER 4: MARKEDNESS), and the phonetic bases of phonological patterning. In this section, we review various analyses of transparency and opacity. Our aim is to highlight the central issues and the key insights, and to point out the areas where improvements can be expected.

The first generative account of vowel harmony is considered to be Lightner (1965). Lightner promoted the intuition that stems in vowel harmony languages are marked for a value of the feature exhibiting harmony. In this "root-marker" approach, the Akan stem for 'pierce' [fiti] would be marked for the feature [+ATR] and 'show' [cɪrɛ] would be marked for [−ATR] (data from the Asante dialect of Akan; Clements 1985). This feature is then realized on all vowels of the stem and on harmonizing affixes that may appear in combination with the stem. Representative derivations of [o-fiti-i] 'he pierced it' and [ɔ-cɪrɛ-ɪ] 'he showed it' are given in (6). The intuition that the harmonizing feature is a property of the stem or, more precisely, that it is not a property of any particular segment in the stem invited attempts to treat vowel harmony in ways parallel to contemporary proposals on the treatment of tonal phenomena (Goldsmith 1976). In the domain of tone and around the same time in the development of phonological theory, substantial attention was devoted to data where tonal melodies on a tier independent from that of the segmental sequence are associated to the segmental sequence via general principles of association. Spreading, specifically, is the operation that adds association lines between, say, a single low or "L" tonal specification and each vowel in the stem. The same formal mechanism of spreading could then be applied to vowel harmony (Clements 1976). The main intuition of the "root-marker" approach, feature sharing among multiple vowels in stems or in stem plus affix combinations, could thus be expressed via this independently motivated formal mechanism of spreading. Employing the same formal mechanisms for such disparate phenomena as tone, vowel harmony, and non-concatenative morphology provided the main momentum for the appeal of the autosegmental approach.

(6) *Analysis of ATR harmony in Akan*

a. o-fiti-i → o-fiti-i

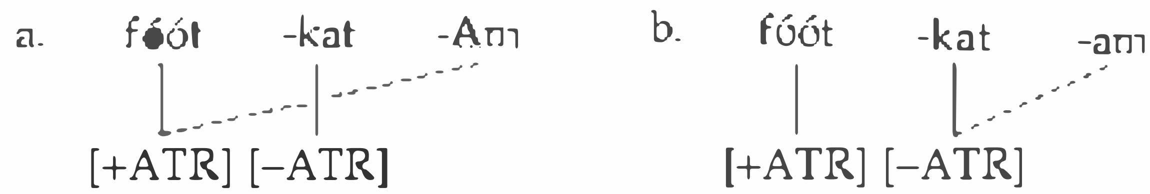
[+ATR] [+ATR]

b. ɔ-cɪrɛ-ɪ → ɔ-cɪrɛ-ɪ

[−ATR] [−ATR]

An immediate consequence of pursuing such an approach to vowel harmony is that the behavior of opaque vowels seems to follow from independently motivated constraints on autosegmental representations. For instance, in Clements (1976, 1980, 1985), opaque vowels are specified with their values for the harmonic feature, so that such a pre-existing association on an opaque vowel blocks spreading from the stem beyond the opaque vowel. The blocking is due to the well-known constraint against crossing association lines. To make this concrete, recall that in Wolof /a/ is in some cases opaque to [ATR] harmony and, in those cases, the vowel surfaces as [−ATR]. For instance, when the suffix [kat] intervenes between a stem and another suffix, it blocks harmony from a [+ATR] stem to the following suffix. In (7), we show how this fact follows from autosegmental principles. The feature specification of the stem vowel cannot spread to the suffix, as shown by the dashed line in (7a), because it would result in line crossing; 'A' denotes a low vowel unspecified for the value of [ATR]. The only allowable result is (7b), where the [−ATR] specification of [kat] spreads to the following suffix. In different words, opacity is seen here as a reflex of the fundamental phonological principle of locality. The final suffix receives its featural specification from its closest eligible donor, its preceding vowel, not from the first stem vowel, which is further away.

(7)



Unlike opacity, transparency has proved to be a recalcitrant problem for phonologists throughout the past approximately fifty-year period of theoretical developments. In autosegmental approaches, the behavior of transparent vowels was not as directly derivable from general principles of representations as was the case with opaque vowels.[5] Consider, for example, the Hungarian diminutives of the proper names *Erzsébet* and *Klára*, *Erzsi* [ers-i] and *Klári* [klaːr-i]. When the diminutives are further combined with the endearment suffix [-kɔ/-kɛ], the result is [ers-i-kɛ] *vs.* [klaːr-i-kɔ]. How is the backness of the suffix transmitted to the suffix vowel? Clements (1976) proposes that, in a first step, the feature [+back] spreads to all vowels of the stem–suffix–suffix combination, including the medial [i]. This step expresses the intuition, by means of spreading, that the suffix is back because the stem vowel is back. But it also results in a representation where [+back] is linked to the medial high, unround vowel. Hungarian lacks such non-low, unround, back vowels. In Clements's (1976) analysis, then, this intermediate representation is corrected by linking [−back] to that vowel. This latter step is conceptually equivalent to the absolute neutralization rules of rule-based analyses, as in Vago (1976, 1980a). Its technical implementation in autosegmental accounts, however, effectively deconstructs the result of the first spreading step to a representation where each vowel has its own separate specification of backness (S. R. Anderson 1980, in a critique of such approaches to vowel harmony, calls this the "mitosis" step; Kenstowicz 1994 brings out independent issues with such manipulations on autosegmental representations). Kiparsky (1981), Clements and Sezer (1982), and Ringen (1988) propose variants of this basic analysis that proceed to either parameterize the harmonic feature-bearing units or keep the features at a different plane from that in which vowel harmony applies (or keep the features at a different tier, as in those accounts that hold the relevant featural specifications of transparent vowels "locked in" the so-called segmental core). The effect of such elaborations is that, in languages with transparent vowels the spreading envisioned in the first step of Clements's (1976) analysis above would not link to these vowels at all. The "correction" step could thus be avoided. Representative works pursuing this perspective are Kiparsky (1981), Goldsmith (1985), Steriade (1987), and Ringen (1988); for critical reviews, see Kenstowicz (1994: 357–359), Farkas and Beddor (1987), and Steriade (1995: 135ff.).

The issue of transparent vowels first participating in harmony and then being readjusted in some way persists across different phonological theories. Thus,

[5] Rule-based accounts of transparency either write in the rule the optional transparent material that can intervene between trigger and target (Ringen 1975: 24) or apply the rule of vowel harmony to all vowels in the stem. In the latter case, the representations are then adjusted so that transparent vowels end up with the expected values for the harmonizing feature (via neutralization rules, as in Vago 1980a).

in Ringen and Vago's (1998) Optimality Theory (OT) analysis of Hungarian, both the representations and the constraints encode a distinction between transparent and all other vowels. On the side of representations, the transparent vowel in e.g. [klaːr-i-ka] is assumed to be unspecified for backness. In Ringen and Vago's (1998) analysis, this assumption crucially enters into the evaluation of the constraint, ensuring that the harmonizing feature links throughout every vowel in the word. That constraint is defined in such a way that it is not violated by the presence of a transparent vowel placed between two back vowels (see e.g. Ringen and Vago 1998: tableau (12)). On the side of the constraints, the distinction between transparent and non-transparent vowel is encoded in the constraint set used in Ringen and Vago (1998) (see e.g. the definition of their constraint in (5b)). Baković and Wilson (2000) and Ní Chiosáin and Padgett (2001) are two later approaches illustrating the same issue. The technical details among the different accounts are of course rather different, but the problem of transparency has remained a persistent challenge.

In early OT (Prince and Smolensky 1993) accounts of vowel harmony, the key notion of spreading is inherited from autosegmental approaches. The formal implementation of spreading pursued an extension of the constraint schema known as ALIGN (McCarthy and Prince 1993). ALIGN constraints were originally used for aligning morpheme edges with prosodic domain edges. For feature spreading, as in vowel harmony, alignment constraints require the alignment of harmonizing features with selected edges of domains. Apart from this, however, within the scope of early OT (Prince and Smolensky 1993) applications, attempts to address the transparency challenge were less reliant on representational assumptions. Thus Smolensky (1993), in his treatment of transparency in Finnish harmony, proposes that the reason for the transparency of [i] is that the feature combination that would result from imposing [+back] on [i] is marked – his constraint *[+B/I]. This constraint is ranked higher than *EMBEDDED, which penalizes embedded feature domains. Then, if ALIGN[+back] is ranked higher than the constraint against *EMBEDDED domains, this results in transparency (the other ranking results in opacity; see Kirchner 1993, Akinlabi 1994, Pulleyblank 1996, Cole and Kisseberth 1995, and Ringen and Vago 1998 for related analyses and discussion). This solution, then, excludes transparent vowels from the domain harmony. But because these vowels intervene between two vowels that agree in terms of the harmonizing feature, this analysis forces the conclusion that harmony can be non-local. Arguably, on both formal theoretical and typological grounds (McCarthy 1989; Clements and Hume 1995; Gafos 1999; Ní Chiosáin and Padgett 2001), this conclusion is considered unwarranted. Subsequent accounts of vowel harmony in OT, discussed below, make efforts to maintain locality in vowel harmony while accounting for opacity and transparency.

In one such account, Kiparsky and Pajusalu (2003) construct a typology of vowel harmony, with a special focus on neutral vowels. In Kiparsky and Pajusalu's system, the presence of vowel harmony is a consequence of a high-ranked constraint AGREE[F], where F is the harmonizing feature; thus, AGREE[ATR] for "tongue root" vowel harmony languages or AGREE[back] for "palatal" vowel harmony languages, and so on. The constraint is formulated in (8) in a way that maintains strict locality – evaluating a candidate with respect to this constraint never requires comparing segments that are not adjacent in the articulatory domain (Gafos 1999; Baković 2000: 4).

(8) AGREE[back]

Articulatorily adjacent vowels must have the same specification for the feature [back].

In addition to AGREE constraints, the Kiparsky and Pajusalu system works with two more types of constraints: markedness constraints such as "front vowels must be non-low and unrounded (*ä, *ö, *ü)," which oppose harmony, and positional faithfulness constraints such as IDENT-FOOT$_1$, which forces faithfulness to an initial foot, and IDENT-STEM. Kiparsky and Pajusalu (2003) use the latter constraints to account for "derived environment" asymmetries in harmonic patterning. Specifically, such domain-specific constraints can capture the pattern that harmony may be stricter in derived environments (CHAPTER 88: DERIVED ENVIRONMENT EFFECTS) than within a morpheme, and stricter in non-initial feet than in the more salient and thus phonologically privileged initial foot. This is formally implemented using the concept of constraint conjunction. In OT, Smolensky (1993, 1995, 1997) has argued that two constraint violations are worse when they occur locally than when they occur non-locally (CHAPTER 62: CONSTRAINT CONJUNCTION). This idea appears to be useful in capturing patterns of optimization in a wide range of cases (see Smolensky 1993, 1995, 1997, on segmental markedness, sonority profiles, and vowel harmony, respectively; Alderete 1997 and Itô and Mester 1998 on dissimilation; Gafos and Lombardi 1999 on consonant transparency; among others). To express this property of constraint violation, Smolensky proposed that the constraint component of OT includes "an operation in UG [Universal Grammar] by which two constraints governing substructures of a given local domain are conjoined into a higher ranked constraint" (Smolensky 1993). This operation, called Local Conjunction, is defined in (9).

(9) *The Local Conjunction of* $_1$ *and* $_2$ *in domain* D

$_1$ &$_l$ $_2$ is violated when there is some domain of type D in which both $_1$ and $_2$ are violated. Universally, $_1$ &$_l$ $_2$ >> $_1$, $_2$.

One use of constraint conjunction in the Kiparsky and Pajusalu (2003) system concerns patterns of stem-internal harmony. Thus, in Finnish, front and back vowels contrast in the second syllable of a stem, *sinä* 'you' *vs.* *kina* 'squabble', *riittä* 'suffice' *vs.* *viitta* 'cloak'. But beyond the first foot this potential for contrast disappears: *kipinä* 'spark', **kipina*. Furthermore, in morphologically complex contexts or "derived environments" such as stem–suffix combinations, the harmonic forces are stronger than within stems: thus, *viit-tä* 'five-ABESS', **viit-ta*. To account for this patterning, a constraint conjoining the individual IDENT-FOOT$_1$[back] and IDENT-STEM[back] constraints is employed.

(10) *Locally conjoined constraint*

IDENT-FOOT$_1$[back] & IDENT-STEM[back]: an [αback] input segment in the first foot of a stem cannot have a [$-\alpha$back] output correspondent.

As a consequence of this constraint, AGREE violations are allowed within morphological or prosodically privileged domains such as the stem or the first foot. In (11) this is formally expressed by the ranking IDENT-FOOT$_1$[back] & IDENT-STEM[back] >> AGREE[back].

(11) *Agreement violations admitted within the privileged domain of the stem/first foot*

	IDENT-FT$_1$[bk] & IDENT-STEM[bk]	AGREE[bk]
☞ a. ([viita])$_{Ft}$		*
b. ([viitä])$_{Ft}$	*!	

Since stems impose their backness value as in *viit-tä*, we must also posit the ranking AGREE[back] >> *ä, *ö, *ü. For *viit-tä*, the conjoined constraint is silent, since the suffix vowel is foot-initial but not stem-internal. But in Uyghur, when the neutral vowel is the only vowel in the stem, the suffix is back: [[i] a], *[[i] ä], where brackets indicate morphological constituency. This difference between Finnish and Uyghur in the way sole transparent vowels combine with suffixes is strikingly simple, and Kiparsky and Pajusalu (2003) propose a formal analysis that seems correct. In Finnish, the ranking is AGREE[back] >> *ä, resulting in transparent vowels imposing their frontness on suffixes despite the markedness of low front vowels: [[i] ä], *[[i] a]. In Uyghur, *ä >> AGREE[back], resulting in transparent vowels failing to impose their frontness on suffixes: *[[i] ä], [[i] a].

When in the context following a back vowel, however, Finnish and Uyghur exhibit the same pattern, namely [[a i] a], *[[a i] ä]. This is the case of transparency, to which we turn next. To account for transparency, Kiparsky and Pajusalu (2003) build parsimoniously on the ideas used in capturing the stem-internal *vs.* derived environment effects described above. As discussed, stem-internally in Finnish, some disharmony is allowed, but [ä a] sequences are not permitted. The key insight here from Kiparsky and Pajusalu (2003) is that harmony violations with marked vowels are avoided, or in other words "disharmony with marked vowels is both worse than disharmony alone, and worse than markedness alone" (Kiparsky and Pajusalu 2003: 8). Formally, this insight enters the grammar as another locally conjoined constraint, shown in (12). As a direct effect of this constraint, *[a ä] is excluded because [ä] is marked and the sequence is disharmonic, but [i a] is admitted because the sequence is disharmonic but none of the vowels is marked.

(12) *Generalized Marked Harmony*

AGREE[back] & *ä, *ö, *ü

This key intuition, capturing such facts about stem-internal harmonic patterning, can be extended to account for transparency. Specifically, as shown in (13), the ranking Generalized Marked Harmony >> AGREE[back] results in transparency. Given /a i – ä/ as input, and comparing candidates [[a i] ä] and [[a i] a], the former candidate incurs a violation of Generalized Marked Harmony: the first two vowels in this string disagree in terms of the harmonic feature, and the last vowel is marked with respect to that feature. The other candidate, [[a i] a], does not violate Generalized Marked Harmony, because it does not contain a marked vowel, though it does incur two violations of AGREE[back].[6] This is the ranking accounting for the case of transparency in Finnish and Uyghur.

[6] This use of constraint conjunction does away with the domain specification in the original definition of conjoined constraints (see (10)), in that evaluation of the conjoined constraint pools violations of the individual constraints irrespective of their locus in the form being evaluated; hence the "generalized" in the constraint reference.

(13) *Transparency*

/a i - ä /	Agree[bk] & *ä,*ö,*ü	Agree[bk]
a. [[a i] ä]	*!	*
☞ b. [[a i] a]		**

In this approach, then, the backness of the suffix vowel is related to the backness of the first stem vowel in a novel way. The reason why markedness in the suffix is a decisive factor is that the first two vowels, [a i], disagree. Since the second vowel is a neutral vowel (hence, front), disagreement implies that the first vowel must be back. Constraint conjunction imposes in this indirect way the same value of backness between the first and the third vowel.

Flipping the ranking of the constraints needed to get transparency results in opacity. That is, with the opposite ranking, Agree[back] >> Agree[back] & *ä, *ö, *ü, and for input /a i – ä /, Agree[back] decides in favor of the opaque candidate [[a i] ä]. This is the pattern observed, for example, in Eastern Khanty. However, note that, in the ranking necessary to express opacity, Agree[back] >> Agree[back] & *ä, *ö, *ü, the conjoined constraint is outranked by one of its constituent constraints. This is generally ruled out in OT (see Kager 1999: 393).

(14) *Opacity*

/a i - ä /	Agree[bk]	Agree[bk] & *ä,*ö,*ü
☞ a. [[a i] ä]	*	*
b. [[a i] a]	**!	

It is a notable corollary of the Kiparsky and Pajusalu (2003) system that back vowels in "palatal" harmony languages are opaque rather than transparent. That is, in the system entertained here, [[ä o] ä] is harmonically bound by [[ä o] a]. The sequence [[ä o] a] violates Generalized Marked Harmony (12) once and Agree[back] once. The sequence [[ä o] ä] violates Generalized Marked Harmony twice and Agree[back] twice. Thus, regardless of the ranking between the two constraints, Generalized Marked Harmony and Agree[back], [ä o ä] is less harmonic than [ä o a]. This is a remarkable result, following from two "first principles" in the Kiparsky and Pajusalu system: markedness and agreement.

We now return to the Hungarian facts, to work out the implications of transparency and opacity being present in the same system. To account for transparency, we must posit the ranking Agree[back] & *e >> Agree[back] (henceforth, when referring to Hungarian forms, the symbol for the low front vowel will be "e" in place of Kiparsky and Pajusalu's "ä" symbol). The candidate illustrating opacity, [[a i] e], violates the conjoined constraint Agree[back] & *e, *ö, *ü, because "e" is marked for the harmonizing feature [back] and [i] is disharmonic for [back]. This entails the desired harmonic ordering, that is, [[a i] e] (opacity) < [[a i] a] (transparency).

(15)

	AGREE[bk] & *e,*ö,*ü	AGREE[bk]
☞ a. [[a i] a]		**
b. [[a i] e]	*!	*

In this formal system, then, markedness as embodied by the *e constraint plays a crucial role in deriving transparency. However, Hungarian presents a problem here. The vowel *é* [eː] is a mid vowel, whereas *e* is a low or at least lower vowel, typically transcribed as [ɛ] (Siptár and Törkenczy 2000: 385, 426). This is the standard phonological classification. Furthermore, ultrasound imaging data show a rather clear difference in tongue height between the group [iː i eː] and [ɛ]. The latter vowel shows a much lower and more retracted tongue body posture than the other vowels. In Hungarian, there are suffixes with front non-low surface alternants such as the adessive [-neːl]/[-naːl]. The markedness constraint "front vowels must be non-low and unrounded (*e, *ö, *ü)" would not be violated by the vowel in [-neːl]. The conjoined constraint Generalized Marked Harmony would therefore also not be violated in [[a i] eː], and the decision is made by the other constraint AGREE[back], which declares [[a i] eː] > [[a i] aː]. This is the wrong outcome. In Hungarian, the choice between [-neːl], [-naːl] is made in the same way as for other suffixes such as the dative [-nɛk], [-nɔk].

Furthermore, unlike Finnish and Turkish, Hungarian admits disharmonic stems with marked vowels such as [fotel] 'armchair' and [farmer] 'jeans'. Hungarian also has back-initial stems followed by front round vowels as in *sofőr* 'driver', *parfüm* 'perfume', and *kosztüm* 'costume'. These words are relatively recent loans (Kertész 2003: 67), perhaps belonging to a lexical "periphery" in the sense of Itô and Mester (1995). Nevertheless, in their harmonic behavior they are completely regular. To account for the opacity pattern exhibited in these stems, the ranking schema must be AGREE[back] >> Generalized Marked Harmony. But recall that the transparency pattern required Generalized Marked Harmony >> AGREE[back]. The relevant datum is sequences of the [[a y] e] profile, the opaque candidate, *vs.* [[a y] a], the transparent candidate (see (16)). The crucial decision rests on how AGREE[back] & *e, *ö, *ü is evaluated. In [[a y] e], [a y] disagree and e is marked. This is the same profile of violations for the transparent winner [[a i] a] discussed above, and implies at least one violation of AGREE[back] & *e, *ö, *ü. Furthermore, the subsequence [a y] may also be considered to violate AGREE[back] & *e, *ö, *ü exactly as assumed for the stem-internal [ä a] sequences in Finnish above: [a y] disagree and [y] is marked. If two violations of AGREE[back] & *e, *ö, *ü are assigned, then [[a y] a] is the predicted winner. If only one violation of AGREE[back] & *e, *ö, *ü is assigned, then [[a y] e] is the correctly predicted winner.

(16)

	AGREE[bk] & *e,*ö,*ü	AGREE[bk]
a. [[a y] a]	*	**
b. [[a y] e]	*(*?)	*

In this system, it is conceivable to split the conjoined constraint into two separate ones: AGREE[back] & *e and AGREE[back] & *y. This permits more freedom in differentially weighing violations of the corresponding constraints, as shown in (17). But in this case, this change does not help, since to account for transparency we must posit AGREE[back] & *e as the highest-ranked constraint. This constraint is violated by the attested output [[a y] e], which is thus predicted to be sub-optimal.

(17) *Splitting the conjoined constraint*

	AGREE[bk] & *e	AGREE[bk]	AGREE[bk] & *y
☞ a. [[a y] ə]		**	*
b. [[a y] e]	*!	*	*

The challenges reviewed so far derive from a single formal property of the analysis: the rankings of constraints needed to account for transparency and opacity are different. There is no hard evidence a priori that this must be true of any system that aims at capturing both transparency and opacity. After all, focusing on palatal vowel harmony systems and without loss of generalization, within any given language, the transparent vowel is different from the opaque one, e.g. [i] is transparent and [y] is opaque in Hungarian/Finnish. An alternative approach to transparency and opacity could capitalize on inherent properties of the vowels in question while leaving the general schema of grammatical forces the same.

The need for different grammar rankings for transparency and opacity can be traced to a crucial assumption. That assumption is that transparent vowels categorically do not partake in agreement, as required by harmony. In other words, (all) transparent vowels categorically disagree with their adjacent vowels. A diametrically opposed perspective is to abandon this key assumption of disagreement and devise a system admitting differential agreement relations between vowels – that is, not simply agree *vs.* disagree, but a spectrum of agreement qualities tailored to the specific combination of vowels involved in the relation for which agreement is assessed. Beňuš (2005) develops this approach on the basis of a rigorous experimental study of Hungarian transparent vowels. The general schema governing both transparency and opacity in Beňuš's account is expressed by a single ranking: PERCEPTUAL FAITHFULNESS >> AGREEMENT >> ARTICULATORY FAITHFULNESS. The broad correspondence between the constraints in this ranking and those discussed so far is as follows. AGREEMENT corresponds to AGREE[back], requiring overlapping vowels to agree as much as possible in constriction location. Its opposing but lower-ranked constraint is ARTICULATORY FAITHFULNESS. This corresponds to IDENT[back] of previous accounts. The ranking AGREEMENT >> ARTICULATORY FAITHFULNESS expresses the simple fact that the language exhibits vowel harmony. The dominating PERCEPTUAL FAITHFULNESS restricts the harmonic agreement forces to degrees of agreement that leave the perceptual identity of the vowels in question unchanged. The closest parallel to PERCEPTUAL FAITHFULNESS in contemporary OT accounts is Baković and Wilson's (2000) similarity scale – in that analysis, transparent vowels are by-products of an optimization starting with the full assimilation candidate and changing that to a less marked candidate, while maintaining similarity with the full assimilation candidate. Overall, then, the main idea in Beňuš's proposal is that vowel harmony is driven by articulatory agreement between overlapping vowel gestures and that this agreement is constrained perceptually. In an A–I sequence, where "A" represents any back vowel and "I" any front vowel, vowels differ with respect to their potential for agreement: following any given back vowel, a high front vowel like /i/ is most retractable, a mid front vowel like /e/ is somewhat retractable, and the high front round /y/ is minimally retractable. Articulatory–perceptual quantal relations are crucial in determining the degree to which a vowel can be retracted without losing its perceptual identity. For transparent vowels, such as the [i] in the sequence [[a i] V_{suffix}], articulatory agreement with the preceding back vowel can be substantial without changing the perceptual identity of the [i]. In turn, the degree of articulatory backing induced on the transparent vowel by agreement with its preceding back vowel is sufficient to trigger a back suffix, hence [[a i] a]. For opaque vowels, the ranking of constraints remains the same. The difference in behavior between opaque and transparent vowels derives from the quantal characteristics of the different vowels. In [[a y] V_{suffix}], the front round vowel cannot be retracted to the same degree as the vowel [i] without losing its perceptual identity as a front vowel. The limited potential of agreement with a back vowel for [y] results in less articulatory backing on the [y], which in turn results in the selection of a front suffix, hence [[a y] e]. In other words, the high ranking of PERCEPTUAL FAITHFULNESS prevents establishing articulatory agreement to a degree that would induce a back suffix alternant just when the front vowel is [y].

The typological implications of the phonetically informed approach pursued in Beňuš (2005) have yet to be pursued systematically. That approach crucially relies on rigorous phonetic records. There is a great disparity between the range of languages whose vowel harmony patterns have been described or even analyzed in the theoretical literature and those for which we have systematic phonetic investigations for the same or even a subset of the same data. Apart from Beňuš's studies on Hungarian and some more recent data on transparency from Kinande, to be discussed in the next section, there are no other studies we are aware of that have systematically examined transparency and opacity patterns in vowel harmony using rigorous experimental methods. It would not be an exaggeration to state that we have only recently begun to register the phonetic record for transparency and opacity using rigorous (i.e. non-impressionistic) data registration methods.

Thus two key insights have emerged. In the typological study of Kiparsky and Pajusalu (2003), markedness and its prioritization with other generally accepted grammar principles such as AGREE go a long way in capturing essential distinctions between the different languages reviewed in that study. The other idea, from Beňuš's (2005) experimentally based work, is that considerations of the relation between articulation and acoustics play a crucial role in transparency and opacity. It seems that a combination of the two insights, supported by a parallel pursuit of theoretical and experimental work, would provide a solid basis for gaining a better understanding of the phenomena at hand.

4 Phonetic bases

Understanding the phonetic bases of opacity and transparency must begin with an understanding of the phonetic bases of vowel harmony in general. Since Gay (1977, 1978), it has been known that a non-contiguous sequence of identical vowels such as [u–u] in *kutup* is produced by speakers of English with a discontinuity in both the articulatory and the electromyographic measures of lip rounding. For example, in the electromyographic signal there is a trough coincident with the production of the intervening consonant. The cessation of muscle activity during the consonant is consistent with the hypothesis that the linguistic representation underlying the production of lip rounding schedules the rounding of the two identical vowels as two independent events, $[u]_{Round}C[u]_{Round}$ (C is a variable for any permissible intervocalic consonant or consonant cluster). A number of other studies have documented the same trough pattern in the production of non-contiguous identical vowels in Spanish, French (Perkell 1986), and Swedish (McAllister 1978; Engstrand 1981). In contrast to these cases, Boyce (1988, 1990) found in Turkish a plateau of continuous activity for /uCu/ utterances both in muscle excitation patterns (of the orbicularis oris) and in lower lip protrusion kinematics. This pattern of results, the English trough *vs.* the Turkish plateau, seems to reflect the fact that Turkish but not English has vowel (rounding) harmony. Furthermore, the linguistic representation underlying the production of lip rounding in Turkish is consistent with a central idea of autosegmental theory, namely that assimilation and harmony involve representations in which a single instance of the assimilating or harmonizing property extends over a domain encompassing all segments required to agree on that property. Thus, in Turkish, the rounding property of the two vowels would extend over a domain encompassing both vowels in /uCu/ (Clements 1976; Kiparsky 1981; Goldsmith 1985), and this is what gives rise to the plateau seen in Boyce's study. Open issues do remain, however, with establishing a link between linguistic representation and the experimental record here with some degree of confidence (Gafos and Goldstein, forthcoming).

The picture becomes more complex when we consider the facts of transparency. As we have seen, many languages with harmony include vowels that can intervene between the trigger and the target of vowel harmony even when they bear the opposite value for the harmonizing feature. In Hungarian, *papír* selects [+back] suffixes, such as *nak* 'DAT', *húz* 'ALL', *tól* 'ABL', *ban* 'INESS', in agreement with the [+back] value of the initial stem vowel and despite the intervening [−back] value of /iː/. At the heart of the problem that transparent vowels pose for the phonetic basis of vowel harmony is an assumption about their representation. The assumption is that the phonological category of a transparent vowel is invariant across different contexts and irrelevant to the quality of the suffix following the transparent vowel. In an impressionistic sense, the transparent vowels in words like *buli-nak* 'party-DAT', *híd-nak* 'bridge-DAT', and *mamicsi-nak* 'mother-DIM-DAT' are not perceptually different from those in *bili-nek* 'pot-DAT' and *víz-nek* 'water-DAT'. Hence, they are assumed to be invariant across these different contexts. However, it is well known that, for vowels, a relatively stable acoustic output can be produced using multiple articulatory strategies and constriction locations. Stevens (1972, 1989) promotes the idea that the relation between acoustic and articulatory dimensions of phonetic form displays discontinuous characteristics. In "stable" regions of an abstract articulatory–acoustic space, change along an articulatory dimension does not result in significant change in acoustics. In "unstable" regions, however, comparable articulatory change can cause significant difference in acoustics. Stevens argued that Universal Grammar utilizes the presence of such discontinuities in the dual articulatory–acoustic phonetic substance to encode

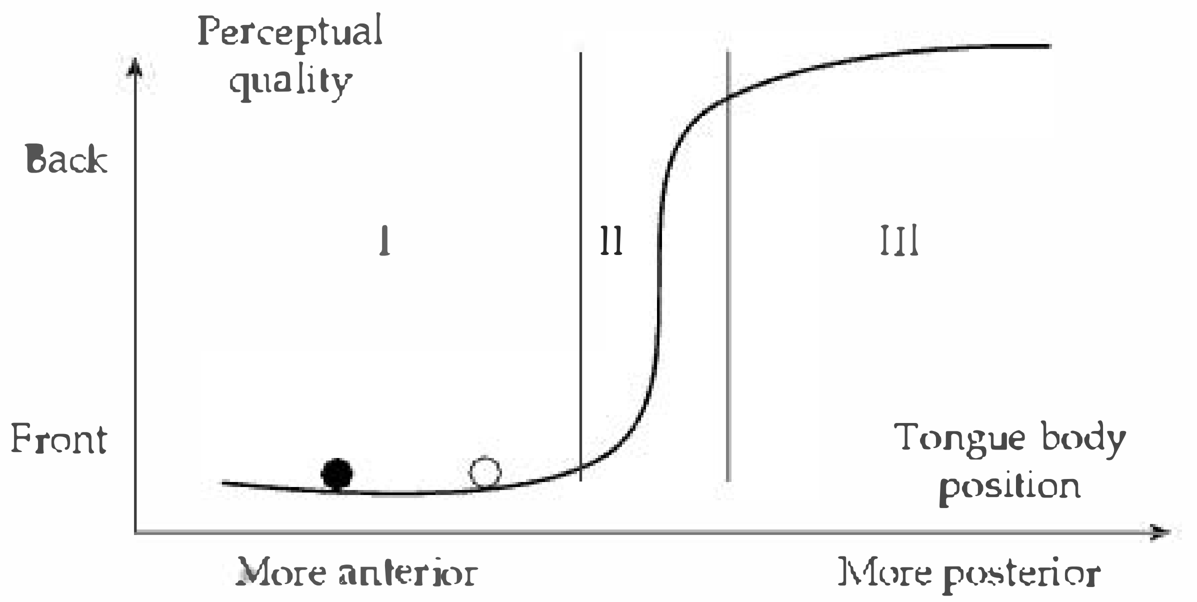


Figure 91.1 Non-linearity in front non-low unround vowels. Tongue body retraction is shown as the difference between the x-coordinates of the two circles, while the minimal perceptual effect of this retraction is shown on the y-axis

contrasts in phonological systems. Moreover, the presence of such regions, according to Stevens, explains why the abundance of co-articulation in natural speech does not hinder perception.

A group of sounds with well-documented discontinuities in the relation between articulation and acoustics is that of the non-low front unround vowels. Calculations using both simple tubes (Stevens 1989) and natural human vocal tract profiles (Wood 1979) show that the acoustic outputs for non-low front vowels are insensitive to a limited amount of variation in the horizontal position of the tongue body. For example, the vowel [i] may be articulatorily retracted to some degree without losing its perceptual identity. The central result is illustrated in Figure 91.1. The S-like curve divides the abstract phonetic space into the stable Regions I and III and the unstable Region II. The horizontal coordinate of the ball sitting on the curve represents the locus of a palatal constriction formed by the tongue body articulator. The black circle corresponds to a tongue body position with the palatal constriction of a prototypically front vowel. The slightly retracted tongue body position illustrated with the white circle falls in the stable region of perceptual stability; a vowel with this constriction location is still considered a front vowel.

We stress that the foundational results of Stevens and Wood above are not about a specific language. Rather, they characterize properties of articulatory–acoustic relations for specific sets of vowels, the non-low front unround vowels, language-independently. These are precisely the transparent vowels of palatal vowel harmony systems like Hungarian and Finnish.

The articulatory–acoustic relations reviewed above provide a plausible phonetic basis for transparency: transparent vowels in palatal vowel harmony are those vowels that can be articulatorily retracted to a certain degree, but maintain their perceptual quality of being front. This hypothesis has been pursued in studies of Hungarian transparent vowels [iː], [i], and [eː] using a combination of electromagnetic articulometry and ultrasound (Beňuš 2005: ch. 4). In these studies, one set of stimuli consisted of word pairs where transparent vowels occur in stems triggering front or back suffixes. In the first set, all words were trisyllabic, e.g. *zefír-ben* 'zephyr'-INESS *vs.* *zafír-ban* 'sapphire'-INESS. Such pairs permitted comparisons of the tongue posture for /iː/ in the two vowel harmony contexts, front and back. The second set of stimuli consisted of monosyllabic words. For example, /eː/ in *szél* 'wind' was compared to /eː/ in *cél* 'aim'. The forms *szél*, *cél* correspond to the nominative case of the respective nouns, where there is no overt suffix. When these stems appear with overt suffixes, *szél* triggers a front suffix while *cél* triggers a back suffix: *szél-nek* 'wind-DAT' *vs.* *cél-nak* 'aim-DAT'. Once again, such pairs permit comparisons of the tongue posture for /eː/ in the two vowel harmony contexts, front and back. However, they differ from pairs like *zefír vs. zafír* in one crucial respect. For *zefír vs. zafír*, the difference in suffix choice, front for *zefír vs.* back for *zafír*, is typically ascribed to the presence of a front *vs.* back stem-initial vowel. In the *szél vs. cél* pairs of stimuli, if systematic subcategorical differences are found in the transparent vowels, then the source of these differences cannot be ascribed to another vowel within the same stem. The results were, first, that transparent vowels in back harmony contexts show a less advanced (more retracted) tongue body posture than phonemically identical vowels in front harmony contexts: e.g. [i] in *buli-val* is less advanced than [i] in *bili-vel*. Second, transparent vowels in monosyllabic stems selecting back suffixes are also less advanced than phonemically identical vowels in stems selecting front suffixes: e.g. [iː] in *ír*, taking back suffixes, compared to [iː] of *hír*, taking front suffixes, is less advanced when these stems are produced in bare form (no suffixes). Because these monosyllabic stimuli were presented in isolation, the observed sub-phonemic differences cannot be attributed to contextual co-articulation. These differences must be part of the speakers' knowledge of these stems.

To review, then, the main result of the above experiments is that the harmonic type of a stem is realized as a sub-phonemic difference in the tongue body position of transparent vowels. Transparent vowels in the front-selecting stems are produced with the tongue body more advanced than the phonemically identical vowels that occur in back-selecting stems. The non-linearity in the articulatory–acoustic relations in these vowels provides a way to understand why the articulatory differences revealed in these experiments cause minimal differences in the acoustic output of these vowels. In turn, we can then understand why in impressionistic transcriptions these vowels are denoted with a phonemically invariant category, even though there are systematic articulatory differences.

The hypothesis that transparency has a basis in non-linearities between articulation and acoustics can be extended to at least two other cases of transparency in harmony. In ATR harmony, we discussed articulatory and perceptual factors contributing to limited contrast potential in the class of vowels commonly exhibiting transparency in these harmony systems. As noted, in Kinande "tongue root" harmony, articulatory, and acoustic data also indicate that the low vowel /a/ is an integral part of the harmony domain (Gick *et al.* 2006; Kenstowicz 2009). In the case of Mongolian, where [i ɪ] are transparent to the spreading of rounding, we see no principled reason why the same hypothesis would not be applicable. If the phonetics of "rounding" is pursued with some care (Disner 1983; Goldstein 1991), lip posture can be hypothesized to spread through the intervening [i] without a substantial effect on its acoustics. Overall, then, the plausible hypothesis is that transparency is not failure to participate in harmony but failure to produce salient acoustic consequences of harmony on a specific class on segments. Kaun (1995: 142) expresses a similar intuition, that for transparent vowels their "occurrence does not constitute a substantial interruption of the signal associated with the extended feature."

The hypothesis grounding transparency in articulatory–acoustic relations may also allow us to understand why certain vowels exhibit transparency, but other similar vowels exhibit opacity. To illustrate this, consider that one important generalization in the phonological patterning of transparency in "palatal" harmony systems concerns front round vowels. Phonologically, front round vowels do not behave transparently. In Hungarian, for example, front round vowels in stem-final position are followed by front suffixes, irrespective of the quality of the preceding vowels (*parfüm-nek*, **parfüm-nak* 'perfume-DAT', *tök-nek*, **tök-nak* 'pumpkin-DAT'). In contrast, front unround vowels can be followed by front or back suffixes. Whence the difference in harmonic patterning between [y] and [i]? The two vowels are very similar in terms of their lingual articulation (Beňuš *et al.* 2003; Beňuš 2005: ch. 4). The only remaining source of difference between [y] and [i] is their rounding. Both Wood (1986) and Stevens (1989) show that rounding in front vowels significantly affects their quantal properties. In front round vowels the degree of perceptually tolerated tongue body retraction is more limited than for unround vowels. The reason for this is the difference in the effective position of the constriction relative to the length of the vocal tract. Lip rounding, as for [y], increases the length of the vocal tract. This effectively advances the stable region in which horizontal articulatory perturbations have minimal acoustic effects (Beňuš 2005: §4.6.2). As a result, the potential degree of tongue body retraction for [y] is minimal as compared to [i]. There is therefore a plausible phonetic basis for the horizontal tongue body position for [y] to be more constrained than that of [i] in the context of adjacent back vowels. Hence, the binary (phonological) choice in suffix form correlates with differences in the quantal characteristics of stem-final vowels: front vowels for which some articulatory retraction is perceptually tolerated are followed by either front or back suffixes, whereas front vowels for which comparable retraction is not tolerated are followed by front suffixes only.

To summarize, the evidence reviewed indicates that a plausible phonetic basis for transparency and opacity can be formulated by reference to the link between articulation and acoustics.

5 Prospects

We have reviewed theoretical and experimental work on transparent and opaque vowels in vowel harmony. Here we highlight some directions for future work.

Just as in other areas of phonological patterning, our understanding of transparency and opacity has been increasingly informed by the availability of new experimental methods. Yet much remains to be done on the experimental side. The main aim should be gaining a more rigorous empirical foundation. Phonological theorizing has relied heavily on impressionistic descriptions of data. In the few cases where harmony data have been examined with experimental methods (Rialland and Redouanne 1984; Beňuš 2005; Svantesson *et al.* 2005; Gick *et al.* 2006; Kenstowicz 2009), the results provide new information that in turn requires revision of past theoretical analyses or inspires new ones.

Closely related to the aim of gaining a better grip on the data is research on perception. The experimental findings from Hungarian and Kinande reviewed above show that transparent vowels have two variants, depending on the harmonic type in which they appear. For example, the Hungarian vowels [iː i eː] show a more advanced variant in the front harmony type words and a less advanced one in the back harmony type words. Yet the phonological literature and impressionistic intuitions of Hungarian native speakers suggest that the two variants are non-contrastive. Ultimately, then, connected to this line of research must be perception experiments, testing the ability of listeners to differentiate transparent vowels embedded in different harmonic contexts.

On the side of theory, in turn, insights from theoretical analyses should be integrated with the new data patterns. Another way to express this is to say that high-level phonological theory should make predictions about results that could be obtained in future experiments (CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY). It is this integration between theory and rigorous data that seems to us most likely to lead to a deeper understanding of the phenomena at hand.

REFERENCES

Akinlabi, Akinbiyi. 1994. Featural alignment. Unpublished ms., Rutgers University.

Alderete, John. 1997. Dissimilation as local conjunction. *Papers from the Annual Meeting of the North East Linguistic Society* 27. 17–32.

Anderson, Lloyd B. 1980. Using asymmetrical and gradient data in the study of vowel harmony. In Vago (1980b), 271–340.

Anderson, Stephen R. 1980. Problems and perspectives in the description of vowel harmony. In Vago (1980b), 1–48.

Anderson, Stephen R. 1981. Why phonology isn't "natural." *Linguistic Inquiry* 12. 493–539.

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Archangeli, Diana & Douglas Pulleyblank. 2007. Harmony. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 353–378. Cambridge: Cambridge University Press.

Baković, Eric. 2000. Harmony, dominance, and control. Ph.D. dissertation, Rutgers University.

Baković, Eric & Colin Wilson. 2000. Transparency, strict locality, and targeted constraints. *Proceedings of the West Coast Conference on Formal Linguistics* 19. 43–56.

Beddor, Patrice Speeter, James D. Harnsberger & Stephanie Lindemann. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics* 30. 591–627.

Beňuš, Štefan. 2005. Dynamics and transparency in vowel harmony. Ph.D. dissertation, New York University.

Beňuš, Štefan & Adamantios I. Gafos. 2007. Articulatory characteristics of Hungarian "transparent" vowels. *Journal of Phonetics* 35. 271–300.

Beňuš, Štefan, Adamantios I. Gafos & Louis Goldstein. 2003. Phonetics and phonology of transparent vowels in Hungarian. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 29. 485–497.

Boyce, Suzanne. 1988. The influence of phonological structure on articulatory organization in Turkish and in English: Vowel harmony and coarticulation. Ph.D. dissertation, Yale University.

Boyce, Suzanne. 1990. Coarticulatory organization for lip rounding in Turkish and English. *Journal of the Acoustical Society of America* 88. 2584–2595.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1976. Neutral vowels in Hungarian vowel harmony: An autosegmental interpretation. *Papers from the Annual Meeting of the North East Linguistic Society* 7. 49–64.

Clements, G. N. 1980. Vowel harmony in nonlinear generative phonology: An autosegmental model. Bloomington: Indiana University Linguistics Club.

Clements, G. N. 1985. Akan vowel harmony: A nonlinear analysis. In Didier Goyvaerts (ed.) *African linguistics: Essays in memory of M. W. K. Semikenke*, 55–98. Amsterdam: John Benjamins.

Clements, G. N. 1990. The hierarchical representation of vowel height. Unpublished ms., Cornell University.

Clements, G. N. 1991. Vowel height assimilation in Bantu languages. *Working Papers of the Cornell Phonetics Laboratory* 5. 37–76.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

Clements, G. N. & Engin Sezer. 1982. Vowel and consonant disharmony in Turkish. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part II, 213–255. Dordrecht: Foris.

Cole, Jennifer & Charles W. Kisseberth. 1994. An Optimal Domains Theory of vowel harmony. *Studies in the Linguistic Sciences* 24. 1–13.

Cole, Jennifer & Charles W. Kisseberth. 1995. Paradoxical strength conditions in harmony systems. *Papers from the Annual Meeting of the North East Linguistic Society* 25. 17–29.

Disner, Sandra F. 1983. Vowel quality: The relation between universal and language-specific factors. Ph.D. dissertation, University of California, Los Angeles.

Dolphyne, Florence A. 1988. *The Akan (Twi-Fante) language: Its sound systems and tonal structure*. Accra: Ghana Universities Press.

Engstrand, Olle. 1981. Acoustic constraints of invariant input representation? An experimental study of selected articulatory movements and targets. *Reports of the Uppsala University Department of Linguistics* 7. 67–94.

Farkas, Donca & Patrice Speeter Beddor. 1987. Privative and equipollent backness in Hungarian. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 91–105.

Farnetani, Edda, Kyriaki Vagges & Emanuela Magno-Caldognetto. 1985. Coarticulation in Italian VtV sequences: A palatographic study. *Phonetica* 42. 78–99.

Fowler, Carol A. 1983. Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General* 112. 386–412.

Gafos, Adamantios I. 1999. *The articulatory basis of locality in phonology*. New York: Garland.

Gafos, Adamantios I. & Louis Goldstein. Forthcoming. Articulatory representation and organization. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.) *The handbook of Laboratory Phonology*. Oxford: Oxford University Press.

Gafos, Adamantios I. & Linda Lombardi. 1999. Consonant transparency and vowel echo. *Papers from the Annual Meeting of the North East Linguistic Society* 29. 81–96.

Gay, Thomas. 1977. Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America* 62. 182–193.

Gay, Thomas. 1978. Articulatory units: Segments or syllables? In Alan Bell & Joan B. Hooper (eds.) *Syllables and segments*, 121–132. Amsterdam: North Holland.

Gick, Bryan, Douglas Pulleyblank, Fiona Campbell & Ngessimo Mutaka. 2006. Low vowels and transparency in Kinande vowel harmony. *Phonology* 23. 1–20.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.

Goldsmith, John A. 1985. Vowel harmony in Khalka Mongolian, Yaka, Finnish and Hungarian. *Phonology Yearbook* 2. 253–275.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.

Goldstein, Louis. 1991. Lip rounding as side contact. *Proceedings of the 12th International Congress of Phonetic Sciences*, vol. 1, 97–101. Aix-en-Provence: Université de Provence.

Gordon, Matthew. 2006. Phonetics of harmony systems. In Keith Brown (ed.) *Encyclopedia of language and linguistics*, 2nd edn., vol. 9, 418–425. Amsterdam: Elsevier.

Halle, Morris & Kenneth N. Stevens. 1969. On the feature "advanced tongue root." *MIT Research Laboratory of Electronics Quarterly Progress Report* 94. 209–215.

Hayes, Bruce & Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104.

Hulst, Harry van der. 1985. Vowel harmony in Hungarian: A comparison of segmental and autosegmental analyses. In Harry van der Hulst & Norval Smith (eds.) *Advances in nonlinear phonology*, 267–303. Dordrecht: Foris.

Hulst, Harry van der & Jeroen van de Weijer. 1995. Vowel harmony. In Goldsmith (1995), 495–534.

Itô, Junko & Armin Mester. 1995. Japanese phonology. In Goldsmith (1995), 817–838.

Itô, Junko & Armin Mester. 1998. Markedness and word structure: OCP effects in Japanese. Unpublished ms., University of California, Santa Cruz (ROA-255).

Jungraithmayr, Hermann. 1971. The Tangale vowel system reconsidered. *Journal of African Languages* 10. 28–33.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kaun, Abigail. 1995. The typology of rounding harmony: An optimality theoretic approach. Ph.D. dissertation, University of California, Los Angeles.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kenstowicz, Michael. 2009. Two notes on Kinande vowel harmony. *Language Sciences* 31. 248–270.

Kertész, Zsuzsa. 2003. Vowel harmony and the stratified lexicon of Hungarian. *Odd Yearbook* 7. 62–77.

Kidda, Mairo. 1985. Tangale phonology: A descriptive analysis. Ph.D. dissertation, University of Illinois, Urbana.

Kiparsky, Paul. 1981. Vowel harmony. Unpublished ms., MIT.

Kiparsky, Paul & Karl Pajusalu. 2003. Towards a typology of disharmony. *The Linguistic Review* 20. 217–241.

Kirchner, Robert. 1993. Turkish vowel harmony and disharmony: An optimality theoretic analysis. Unpublished ms., University of California, Los Angeles (ROA-4).

Krämer, Martin. 2003. *Vowel harmony and correspondence theory*. Berlin & New York: Mouton de Gruyter.

Ladefoged, Peter. 1964. *A phonetic study of West African languages*. Cambridge: Cambridge University Press.

Lightner, Theodore M. 1965. On the description of vowel and consonant harmony. *Word* 19. 376–387.

Lindau, Mona. 1978. Vowel features. *Language* 54. 541–563.

Lindblom, Björn. 1983. Economy of speech gestures. In Peter F. MacNeilage (ed.) *The production of speech*, 217–245. New York: Springer.

Magen, Harriet S. 1984. Vowel-to-vowel coarticulation in English and Japanese. *Journal of the Acoustical Society of America* 75. Supplement S41.

Magen, Harriet S. 1997. The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics* 25. 187–205.

McAllister, Robert. 1978. Temporal asymmetry in labial coarticulation. *Working Papers, Institute of Linguistics, Stockholm University* 35. 1–29.

McCarthy, John J. 1989. Linear order in phonological representation. *Linguistic Inquiry* 20. 71–99.

McCarthy, John J. & Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology 1993*. 79–153.

Mutaka, Ngessimo. 1994. *The lexical tonology of Kinande.* Munich: Lincom Europa.

Ní Chiosáin, Máire & Jaye Padgett. 2001. Markedness, segment realization, and locality in spreading. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 118–156. Cambridge: Cambridge University Press.

Ohala, John J. 1994a. Towards a universal, phonetically-based theory of vowel harmony. *Proceedings of the 3rd International Conference on Spoken Language Processing*, vol. 2, 491–494. Yokohama: Acoustical Society of Japan.

Ohala, John J. 1994b. Hierarchies of environments for sound variation; Plus implications for "neutral" vowels in vowel harmony. *Acta Linguistica Hafniensia* 27. 371–382.

Öhman, Sven E. G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39. 151–168.

Perkell, Joseph S. 1986. Coarticulation strategies: Preliminary implications of a detailed analysis of lower lip protrusion movements. *Speech Communication* 4. 47–68.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Przezdziecki, Marek. 2005. Vowel harmony and coarticulation in three dialects of Yoruba: Phonetics determining phonology. Ph.D. dissertation, Cornell University.

Pulleyblank, Douglas. 1996. Neutral vowels in Optimality Theory: A comparison of Yoruba and Wolof. *Canadian Journal of Linguistics* 41. 295–347.

Recasens, Daniel. 1987. An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics* 15. 299–312.

Rialland, Annie & Djamouri Redouanne. 1984. Harmonie vocalique, consonantique et structures de dépendance dans le mot en mongol khalkha. *Bulletin de la Société de Linguistique de Paris* 79. 333–383.

Ringen, Catherine. 1975. Vowel harmony: Theoretical implications. Ph.D. dissertation, Indiana University. Published 1988, New York: Garland.

Ringen, Catherine. 1988. Transparency in Hungarian vowel harmony. *Phonology* 5. 327–342.

Ringen, Catherine & Miklós Kontra. 1989. Hungarian neutral vowels. *Lingua* 78. 181–191.

Ringen, Catherine & Robert M. Vago. 1998. Hungarian vowel harmony in Optimality Theory. *Phonology* 15. 393–416.

Schachter, Paul & Victoria A. Fromkin. 1968. A phonology of Akan: Akuapem, Asante, Fante. *UCLA Working Papers in Phonetics* 9.

Schlindwein, Deborah. 1987. P-bearing units: A study of Kinande vowel harmony. *Papers from the Annual Meeting of the North East Linguistic Society* 17. 551–567.

Siptár, Péter & Miklós Törkenczy. 2000. *The phonology of Hungarian.* Oxford: Oxford University Press.

Smolensky, Paul. 1993. Harmony, markedness, and phonological activity. Paper presented at the Rutgers Optimality Workshop 1, Rutgers University (ROA-87).

Smolensky, Paul. 1995. On the structure of the constraint component *Con* of UG (ROA-86).

Smolensky, Paul. 1997. Constraint interaction in generative grammar II: Local conjunction or random rules in Universal Grammar. Paper presented at the Hopkins Optimality Theory Workshop/University of Maryland Mayfest 1997.

Steriade, Donca. 1987. Redundant values. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 339–362.

Steriade, Donca. 1995. Underspecification and markedness. In Goldsmith (1995), 114–174.

Stevens, Kenneth N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In Edward E. David & Peter B. Denes (eds.) *Human communication: A unified view*, 51–66. New York: McGraw-Hill.

Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17. 3–45.

Svantesson, Jan-Olof, Anna Tsendina, Anastasia Karlsson & Vivan Franzén. 2005. *The phonology of Mongolian.* Oxford: Oxford University Press.

Tiede, Mark K. 1996. An MRI-based study of pharyngeal volume contrasts in Akan and English. *Journal of Phonetics* 24. 399–421.
Vago, Robert M. 1976. Theoretical implications of Hungarian vowel harmony. *Linguistic Inquiry* 7. 243–263.
Vago, Robert M. 1980a. *The sound pattern of Hungarian*. Washington, DC: Georgetown University Press.
Vago, Robert M. (ed.) 1980b. *Issues in vowel harmony*. Amsterdam: John Benjamins.
Wood, Sidney. 1979. A radiographic analysis of constriction locations for vowels. *Journal of Phonetics* 7. 25–43.
Wood, Sidney. 1986. The acoustic significance of tongue, lip, and larynx maneuvers in rounded palatal vowels. *Journal of the Acoustical Society of America* 80. 391–401.

92 Variability

Gregory R. Guy

1 Introduction

Variability is a term used in phonology with several meanings. One common meaning is sociolinguistic diversity: speakers of different social backgrounds speak differently, and all speakers vary in speech style and register. Thus in New York City, as Labov (1966) famously demonstrates, a consonantal articulation of coda /r/ is preferred by higher status speakers, and by all speakers in their more careful styles, while a vocalized or deleted realization is preferred by lower status speakers and in casual styles.

Another sense refers to acoustic and articulatory diversity: English voiced stops can be articulated with voice onset times ranging from negative values to +10 msecs, with a mode around +5 msecs, while their voiceless counterparts typically show VOTs from +20 to +90 msecs. Successive articulations by the same speaker under the same conditions are not identical, but wander around these ranges. Vowels are similarly variable in realization. As Peterson and Barney (1952) showed, a vowel cannot be defined as an articulatory point, or as a particular acoustic realization, but rather as a region in articulatory and acoustic space; a series of measurements of a speaker saying a given vowel will show considerable scatter in this range, and sometimes items that fall outside it.

A third sense of variability addresses simple optionality: phonological characteristics or processes that may or may not occur in certain circumstances. Thus English voiceless stops in final position may be aspirated or not, although those in initial position show aspiration systematically. Hence we can describe aspiration as obligatory in initial position, but optional or variable in final position.

But, despite this range of meanings, variability is simple to define: in its broadest sense it is the inverse of generality. A phonological generalization is a statement about a sound system that is true everywhere, in every relevant occurrence; when some statement is not true everywhere, we encounter a case of variability. A phonological phenomenon that occurs in some, but not all, of its possible instances is not fully general; it is in some respect variable in occurrence (sometimes it happens, and sometimes it doesn't) or in realization (sometimes it happens one way, and sometimes another).

Variability as lack-of-generality is therefore a chronic problem for linguistic analysis. Generalizations are, of course, privileged in linguistics. Linguists are trained to seek generalizations; indeed, we see regularly recurring structures as a defining property of language, and the absence of all regularity as the defining property of noise (i.e. non-language). Consequently, phonology has repeatedly addressed issues associated with defining the frontiers of generality and the treatment of partial generality.

Limits on generality. The search for ever broader and deeper generalizations has been a prominent theme in the history of linguistics. The broadest generalizations about sound systems are phonological universals: statements that are true of all utterances in all languages. A truly universal property cannot be absent or contradicted in some language or some speakers, cannot have lexical exceptions, and must be apparent in 100 percent of cases. Hence a context-free universal can be characterized as in (1):

(1) *Maximum generality: Phonological universals*

For all human speakers (of all languages),
in all linguistic contexts,
in all lexical items,
x is always true.

Any phonological generalization that cannot satisfy all the quantifiers in (1) is less than universal. But this is true of most of phonology; although universals have important status, most work in phonology deals with generalizations that are limited in some respect, for example to a particular language or particular context. Hence, in a sense, most of phonology deals with variability, with partial generalizations that leave a region of variation where non-conforming realizations occur. The phonologist who pursues generality and regularity is therefore always confronted with the task of identifying the limits on the generalization, and the alternatives that occur beyond those limits. This is the problem of variability.

The problem can be approached in terms of the various quantifiers in (1). These are of two types: the "all" quantifiers, which deal with issues of scope, and the "always" quantifier, which addresses prevalence. When these are less than universal, they delineate the several problems of variability. First is social scope: generalizations that are true of only some human speakers (some language, speech community, or ethnic or other social group) constitute cases of social variation. Second is contextual scope: many generalizations are context-sensitive, i.e. valid only for items in a particular structural position. But the definitions of contexts, and indeed of what may count as a relevant context, are substantive theoretical and empirical questions. Various forms for the interaction between context and item have been proposed: a context may be seen as having a categorical effect, a gradient effect, or some other non-categorical outcome. And third is the problem of lexical scope: do phonological processes apply to all relevant words (all that have a given phonological structure in the right context)? Or do some words of the appropriate phonological shape nevertheless fail to conform to an applicable generalization by virtue of their lexical identity? If phonological statements are limited in their applicability to subsets of the lexicon, leaving words or sets of words in which different conditions prevail, we confront problems of lexical variability.

Orthogonal to the problems of scope is the problem of variable prevalence – does a given state of affairs always prevail, or is it encountered only some of the time? Although some theoretical approaches treat this as a scope problem, for example by seeking to define a narrow social or contextual domain in which prevalence is categorical, the logical problem of prevalence exists nonetheless: if all relevant dimensions of scope are held constant, is a given phonological generalization valid for all successive occurrences of relevant forms? If we listen to the same speakers producing the same words in the same contexts, do we always hear the same productions, or do they vary? Does a given process apply 100 percent of the time in the relevant domains, or less than 100 percent? And how does phonology cope with the two scenarios – how does it model categorical prevalence, and how does it account for variable prevalence?

In what follows we consider in turn each of these potential limitations on generality – each of these types of variability. We begin with the question of prevalence.

2 The quantitative limits of generality: Variable prevalence

A linguistic universal has universal prevalence: it always occurs wherever possible. We can describe a phonological generalization that is always true as categorical or obligatory. But how does phonology treat a state of affairs with less than universal prevalence – a generalization that is not categorically true, or a process that is not obligatory? This question lies at the core of what many linguists consider variability.

To illustrate this issue, consider common alternations, found in a number of languages, between presence and absence of coda consonants (CHAPTER 68: DELETION). In natural speech in English, words containing final coronal stops, such as *best*, *old*, are often articulated without those stops (cf. *bes' friend*, *ol' man*). In colloquial Caribbean Spanish and Brazilian Portuguese, syllable- and word-final /s/ is similarly variable in realization: *estamos*, *menos* are often articulated as *etamo*, *meno*. Such facts can be accounted for by phonological deletion processes, which are plausibly motivated by markedness considerations that are likely universal: simpler codas are universally less marked than more complex ones. But in the cases cited, the prevalence of these deletion processes is less than categorical. Not all utterances of eligible words undergo deletion.

These are not questions of scope. The alternations are not restricted to particular words. Although social groups vary in deletion rates, these societies are not composed of some groups or individuals who always delete and others who never do (Guy 1980, 1981). And although these processes show contextual conditioning – for example, all three languages delete more before consonants than before vowels – the contexts do not define domains in which deletion always applies or never applies. No matter how we slice up the data in terms of scope, we always encounter both deleted and non-deleted forms. Hence, the prevalence of these phenomena is less than "always." How does phonology deal with such cases? Or indeed, is phonology responsible for such facts at all?

A phonological state of affairs that does not always prevail needs some statement of limitations on its prevalence. This can be done with quantitative vagueness, by replacing the "always" quantifier in (1) with non-universal quantifiers like "sometimes" or "optionally," or with existential quantifiers like "may occur." Or, a more precise quantification can be used, specifying some frequency or probability of occurrence (CHAPTER 90: FREQUENCY EFFECTS). Thus the prevalence clause on generality can be restated as (2a) or (2b):

(2) *Quantifying prevalence*

a. *Optional generalization*
. . . x is optionally true.

b. *Probabilistic generalization*
. . . x is true with probability p.

As we shall see, these choices are the subject of considerable theoretical debate. Some schools of thought strenuously argue that grammar has nothing to say about frequency or quantification, and consequently deny that statements like (2b) are permissible in formal phonology. Other frameworks embrace to a greater or lesser extent the quantification implied by (2b), seeking to account for variable prevalence by grammatical means. Let us consider the range of arguments bearing on variable prevalence.

The preference for categorical prevalence. Since so much linguistic analysis is inductive, it is unsurprising to observe a long, articulate tradition in linguistic theory of preferring generalizations that are categorical, i.e. true of 100 percent of relevant cases. Generalizations that are not fully general are often treated as valueless. The typical heuristic in linguistic analysis is to hypothesize a generalization, and, if counterexamples are discovered, to seek a reformulation of the generalization that either properly excludes the counterexamples, or accounts for them in another way, perhaps as a consequence of some other generalization (see also CHAPTER 106: EXCEPTIONALITY).

The historical prototype for this heuristic is Verner's (1877) refinement of Grimm's Law, which describes the sound changes characteristic of the Germanic languages. In the early nineteenth century Grimm and others discovered a set of common correspondences between the obstruents of the Germanic branch and those of other Indo-European languages (Rask 1818; Grimm 1819). Among these was the correspondence between Proto-Indo-European voiceless stops and Germanic voiceless fricatives (thus Latin *ped, tres, cornu vs.* English *foot, three, horn*). But there were also many recognized exceptions to these correspondences, so they appeared to fall well short of being categorical "laws." Thus in many words PIE voiceless stops end up as voiced stops or fricatives in Germanic (cf. Old English *fæder, hundred* vs. Latin *pater, centum*). Based on the data known to early nineteenth-century linguists, the Germanic sound-shift might more accurately have been described as a variable process with several outcomes, among which one particular set (those known as Grimm's Law) were quantitatively prominent.

In 1877, this picture was dramatically altered when Karl Verner published a paper showing that one large class of exceptions was regularly conditioned by the position of the word-stress in PIE: the regular fricative outcome for voiceless stops occurs only in initial and post-tonic positions, while the exceptional voiced outcome occurs elsewhere. This discovery removed many counterexamples to Grimm's Law, reducing the domain of apparent variability, and greatly increasing the prevalence of the Grimm's Law generalizations. This allowed the conjecture that a correct definition of contexts would yield two categorical generalizations – two separate conditioned sound changes, each of which *always* occurred in its appropriate context. This conjecture received explicit formulation as the Neogrammarian hypothesis: sound change is "exceptionless," admitting no variation (Osthoff and Brugmann 1878). By this hypothesis, any variability in the data was spurious, deriving from alternate sources such as borrowing, neologism, or dialect mixture, or from an inaccurate statement of the context of the change, calling for a Verner-like amendment of the generalization.

With the emergence of modern linguistics since Saussure (1916), the idea of exceptionlessness became dominant in synchronic phonology: generalizations should be categorical. If a generalization is nearly categorical, the analyst should seek to make it categorical by redefining the context or by explaining the exceptions. Variability, in the sense of a phonological state of affairs for which no redefinition of context yields categorical prevalence, is considered an unhappy outcome by many schools of phonological theory, including structuralism, generative phonology and its various developments, and mainstream Optimality Theory, and it receives little theoretical attention in these frameworks, being tolerated only where empirically necessary.

Optionality. Despite this widespread theoretical preference for categorical statements, the task of accounting for a body of data has always led linguists to the necessity of non-categorical descriptive statements. In American structuralism, this necessity was formally treated in phonemic theory in terms of "free variation" (Swadesh 1934; Hockett 1942). Thus, to account for variable realizations of final voiceless stops in English, a structuralist analysis would present a list of possible allophones in this position which included the aspirated and unaspirated variants. Declaring that these allophones occurred in free variation was equivalent to a statement that their occurrence was random, not subject to further principled (i.e. structural or contextual) analysis. Such an account therefore adopted the strategy of (2a), declaring only that these alternatives all occurred with limited prevalence, but making no attempt to quantify their respective frequencies.

In rule-based generative frameworks, the equivalent mechanism is the "optional" rule: a rule which may apply, generating its output, or fail to apply, which leaves its input to surface unaltered (Chomsky and Halle 1968). Thus the English voiceless stop alternations could be modeled in generative phonology with an optional rule that rewrites the feature matrix of a voiceless stop to include aspiration in final position. When it applies, aspirated forms are generated, but when it optionally fails to apply, an unaspirated realization is generated. Again, this reflects the (2a) approach to variable prevalence, stating only that it exists, without further quantification.

Variable prevalence in OT. The treatment of variability-as-optionality has survived the transition from rule-based to constraint-based models. In constraint-based phonology, optionality can be captured with variable or incompletely specified constraint rankings (Anttila 1997). In Optimality Theory, a fixed constraint ranking is expected to yield a unique outcome for every evaluated form. This is the ranking equivalent of "obligatory" rules: if constraint A outranks constraint B, and the requirements of A and B conflict in some situation, candidate forms must obligatorily satisfy A over B in order to be selected. In most circumstances, the conjunction of the ranked demands of a set of constraints will rule out all but a unique candidate, excluding the possibility of variation.

But Optimality Theory also allows for the existence of a great variety of rank orders, and makes extensive explanatory use of ranking differences. Although OT postulates a universal constraint set, it captures phonological differences between languages by differing constraint rankings. Similarly, dialect differences, and, by extension, all sorts of sociolinguistic differences between individuals, class and ethnic groups, even speech styles and registers, can be modeled as differences in constraint rankings. The theory also models change across time as changes in the rank order of the constraint set. Since different orders generate different outcomes, variable orders within a given grammar will generate variable outcomes, and variable prevalence of any given outcome.

To illustrate, consider the English case mentioned above, final coronal stop deletion (CSD), where alternations like *best* ~ *bes'*, *old* ~ *ol'* occur. To simplify for illustrative purposes, we might model this alternation in OT with two constraints, one that disfavors complex codas (*ComplexCoda), and a generalized faithfulness constraint stating that underlying segments should be realized on the surface (Faith). An OT grammar with these two constraints will generate full forms (*best*, *old*) if Faith is ranked higher, but will generate deleted forms (*bes'*, *ol'*) if *ComplexCoda has the higher ranking. Consequently, if these constraints are variably ranked, the grammar generates both forms in variation. Variable (or underspecified) constraint ranking is thus the OT equivalent of the optional rule.

Competence and performance. All the above accounts – structuralist, generative, and optimality-theoretic – treat variable prevalence in terms of optionality, as in (2a), while eschewing the quantification in (2b). In these schools of thought, the "optionality" approach is considered empirically adequate. For the structuralists, it was typically adequate to account for all the structural patterns in a corpus; hence, a list of options was sufficient. An adequate generative grammar, according to Chomsky (1965), must be able to generate all and only the possible grammatical utterances of a language (or, more properly, of an idealized homogeneous idiolect), and this criterion is largely maintained in OT. In all these frameworks the theory and grammar are responsible only for accounting for the *existence* of a possible form, but not for a more precise account of its likelihood or frequency of occurrence.

The formal models we have considered – the sequence of frameworks that runs from Saussure through the structuralists and generativists to OT – thus say nothing about whether a form is common or rare, preferred or exceptional. In fact, they mostly define such facts as lying outside the purview of grammar and formal linguistics. The limited empirical responsibility of the grammar is seen as a theoretical necessity in these models, as a consequence of a set of assumptions about the organization of language. They claim a fundamental distinction between the system of language (which defines grammaticality, possible structures, etc.), and the usage of that system and the utterances it generates. Linguistic theory, and the grammars that encapsulate linguistic knowledge, are concerned with the former, termed *langue*, competence, or i-language, according to the terminology of the day. The usage speakers make of the system, their productions and utterances, is treated as a separate phenomenon, termed *parole*, performance, or e-language. Performance and production are argued to bear an uncertain relationship to the system, subject to non-linguistic constraints such as errors, interruptions, or memory lapses. And the system says nothing about prevalence. Grammars are postulated to be essentially non-quantitative; in one formulation, "grammars can't count." Properly speaking, such models are weakly quantified: the grammar permits the specification of at most three levels of quantification: "always" (anything obligatory), "never" (anything the grammar doesn't generate), and "sometimes" (anything optional or occurring in free variation). But any more quantitative detail than this is, by definition, a consequence of usage, and hence need not and cannot be modeled by the grammar. Statement (2a) is a possible element of competence, but (2b) is, by this definition, a statement about performance.

The position just described lies on one side of a major fault-line in phonological thought. On the other side lie several theoretical frameworks that take a more expanded view of the empirical responsibilities of grammar, and a different view of the capabilities of grammar. As we see in the next section, these frameworks undertake to quantify prevalence, and admit statements like (2b) as elements of grammar.

Quantified prevalence. Opposed to the models just described are theories in which phonological analysis explicitly engages with the quantitative facts of variable prevalence. Some examples of these are the "quantitative paradigm" arising from sociolinguistic research (exemplified by the variable rule model; Labov 1969; Cedergren and Sankoff 1974), the variable OT models associated with scholars such as Anttila (1997, 2009) and Nagy and Reynolds (1997), the Stochastic OT model (Boersma and Hayes 2001), and "usage-based" models such as Exemplar Theory (Bybee 2001; Pierrehumbert 2001, 2006). The point of departure for these frameworks lies in the empirical evidence demonstrating that many variants have distinctive quantitative tendencies: some forms are recognizably rare, while others are common, frequent, even highly preferred. For example, in American English, final voiceless stops are rarely aspirated, while in other dialects, such as Irish English, aspiration is common (Kirke 2005). Speakers appear to be aware of these quantitative facts in the sense that they faithfully reproduce the aspiration rates typical of their speech communities, and are capable of recognizing speech that shows a different rate as distinctive.

Empirical studies of variation massively document the systematic nature of such quantitative patterns (Labov 1966, 1969; Cedergren 1973; Poplack 1979; Guy 1981). Every language has phonological variables that systematically occur at certain frequencies. This is the quantitative form of structure, and the quantified models we will consider seek to account for this structure. The typical kinds of patterning and systematicity are illustrated in Table 92.1. These data are drawn from a study of CSD in early New Zealand English (conducted by the author with Jen Hay and Abby Walker using the ONZE corpus; cf. Guy *et al.* 2008; Hay *et al.* 2008). As many studies have demonstrated, coronal stop deletion rates are sensitive to following context, here classified as consonants, glides, vowels, or zero (i.e. utterance-final position). The study sample included speakers from several national backgrounds; the table separates those with Scottish backgrounds, English backgrounds, and mixed backgrounds (including Scottish, English, and/or Australian parents and settlement histories).

These data show several highly regular patterns. First, they are not random. Randomness does occur in linguistic production: some phonetic variability in articulation, such as the scatter observed in vowel articulations, is a consequence of random variability in physical gestures; random ordering of constraints is postulated to be a basic feature of the selection mechanism in variable and

Table 92.1 Probabilities of coronal stop deletion in Early New Zealand English, by following context and national background of speakers

	__C	__G	__V	__Ø
Scots	0.84	0.57	0.22	0.12
English	0.77	0.59	0.29	0.18
mixed	0.77	0.71	0.32	0.15
mean	0.793	0.620	0.275	0.151
range	±0.047	±0.09	±0.055	±0.031

Stochastic OT models. But random is not the same as non-categorical (CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY). The statistical meaning of randomness is that all possible outcomes are equally likely: when there are two alternants, each should occur 50 percent of the time, like coin flips. Since the possibilities in this case are deletion or non-deletion, a grammar that randomly generated forms would yield deletion probabilities of 0.5. Second, the results show a systematic effect of following context: consonantal contexts promote deletion (glides somewhat less than obstruents), while vowels disfavor deletion and null contexts are associated with the least deletion of all. Third, these contextual effects are regularly observed in all the speaker groups. The rank order of favorability to deletion is C > G > V > Ø for all three groups, and the actual numerical values for each context fall into non-overlapping ranges, tightly clustered around a mean. Each of these speaker groups, of course, constitutes an independent experiment, being made up of separate individuals who had little or no lifetime contact with the other speakers. If they were behaving randomly, it would be essentially impossible for them to converge on common values.

Such findings, echoed repeatedly in studies of variation, are a principal motive for the quantified models considered here. Given quantitative properties of language that are non-random, systematic, and linguistically conditioned, these theories all seek to account for them by grammatical means, which implies statements of prevalence along the lines of (2b). However, the various quantified theories differ substantially in their assumptions and formal models; in the rest of this section we consider the principal models in turn.

Variable rule model. The earliest formally quantified approach was the "variable rule" model (VR), developed by Labov (1969) and Cedergren and Sankoff (1974). This model was conceived as a straightforward extension of generative grammar, in which the rules are all quantified by probability of occurrence. "Obligatory" rules receive the same quantitative treatment as in conventional models: they have a probability of 1. But the distinctiveness of VR lies in the treatment of "optional" processes, here termed variable processes: for these, the probability can be any real number between 0 and 1. This permits a VR model to make specific quantitative predictions for any phonological variable. The values in Table 92.1 are taken from such an analysis, in which the likelihood of deletion was related to following context; VR incorporates a treatment of contextual constraints on a process. Context-sensitivity is of course an essential feature of an adequate account of variability, just as it is essential for adequate accounts of invariant outcomes, as we saw in the Germanic sound-shift case. Contextual limits on generality are discussed further in §4, below.

The probabilistic quantification of VR illustrated in Table 92.1 lends the model all the interpretive significance of real numbers. First, different numbers mean differing effects. A model that only defines options could only say that deletion is optional in all four contexts, but a quantitative model shows relationships of more and less: in this case, that consonants and glides favor deletion most, followed by vowels and then by zero. Second, difference is scalar, distinguishing proximal values from distant values; here following glides are closer in effect to consonants than they are to vowels: comparing means for C (0.793), G (0.620), and V (0.275), the distance from G to C (0.173) is just half of the distance from G to V (0.345). Third, the model is quantitatively comparable and falsifiable: we compared across the speaker groups and found that they all had similar values for these contexts. We could look at other speakers to see if they have similar values, or look at lexical or morphological subsets of the data to see if the context effects remain constant. If they do not, we might explore other predictors of these differences. Finally, precise quantification lets us use the standard apparatus of statistics: tests of significance, central tendencies, dispersion, etc. For example, the values reported in Table 92.1 are significant predictors at the 0.05 level or better.

An important characteristic of VR is that it uses a multivariate analysis to partial out the effects of various predictors or contexts. This has an important practical consequence: it controls for differing distributions of the data across the independent variables. In CSD, for example, the morphology of a word has a substantial effect on deletion rates: past tense forms like *missed, packed* are deleted less than monomorphemes like *mist, pact*. This factor was controlled for in the analysis shown in Table 92.1. But an analysis that failed to do this could yield numbers like those in Table 92.1 purely as an epiphenomenon. Thus if most of the words with following vowels happened to be verbs (in phrases like *messed up, baked it*), while most words with following consonants were monomorphemes (in phrases like *best friend, old man*), a univariate analysis would be open to doubt: do monomorphemes have high deletion rates because they usually occur before consonants, or do following consonants show high deletion because they are most often preceded by monomorphemes? Multivariate analysis, used in VR (and in Exemplar Theory studies; Jannedy and Hay 2006), routinely controls for such problems.

As noted above, the "variable rule" model was originally designed for a rule-based grammar, but it is not wedded to such a formalism. The conceptual framework of VR treats linguistic production as set of choices in which each alternative is associated with a probability. The model is agnostic with respect to whether those choices are modeled as rules, constraint orders, branching graphs (as in systemic functional grammar, cf. Halliday 1985), selections among allophonic inventories (CHAPTER 11: THE PHONEME), or other formal devices.

Stochastic OT. A prominent quantified model in a constraint-based framework is Stochastic Optimality Theory (Boersma and Hayes 2001). As we have noted, conventional OT can generate optional outcomes by means of variable constraint ordering, but says nothing about relative frequencies of particular constraint orders, nor about relative frequencies of occurrence of alternative outputs. This arises from the fact that constraint ranking in conventional OT is purely ordinal:

one can say that A outranks B, but not by how much. There is no concept of A being slightly higher than B, or a lot higher. Stochastic OT adds the concept of proximity. Ranking values in Stochastic OT are real numbers, not sequential positions – values like 92.7, 94, and 101.3, not A >> B >> C. An analogy can be drawn with the results of a horse race: conventional OT tells us which horse finished first, second, third, etc., but doesn't say if the first place finisher won by a nose or a mile. Stochastic OT, however, metaphorically gives the finishing times, from which we can deduce whether A beat B by a tenth of a second, or by several minutes. This concept of proximity is then translated into a likelihood of ranking orders being reversed: if the race is re-run with a certain amount of randomness affecting the outcomes, horses or constraints that are separated by very little might well end up in a reverse order, but a horse that trails by a mile is unlikely to ever make up the distance and come out ahead.

Formally, Stochastic OT postulates that each constraint varies randomly within a probability envelope around its central ranking value whenever it is called upon to evaluate a form; at some points in this range it may outrank other nearby constraints, but at other points it will fall below them in ranking value. By suitably adjusting the ranking values of constraints, this model can match frequency distributions observed in natural data. In the English CSD example, a speaker who deletes 50 percent of the time gets equal ranking values for the markedness constraint (*ComplexCoda) and the faithfulness constraint (Faith), so their order of dominance is completely random. But a speaker who deletes at a low rate like 10 percent would have a ranking value for Faith that is appreciably higher than the value for *ComplexCoda, and the distance between them is such that their probability envelopes overlap in just 10 percent of their total area, representing the region in which the markedness constraint prevails, selecting for deletion.

Variable OT. An interesting variant of Optimality Theory that generates quantitative predictions without an explicit quantitative apparatus within the grammar is found in the works of scholars such as Anttila (1997, 2007) and Nagy (1996; see also Nagy and Reynolds 1997). These approaches treat quantitative patterning in the data as a structural consequence of variable constraint ordering. As noted above, two randomly ordered constraints would, in the long run, occur in the order A >> B at 50 percent of the time, and B >> A the remaining 50 percent. If these orders select different candidate outputs, a random model predicts the two variant realizations should occur in a 50 : 50 ratio. Anttila, Nagy, and their associates have extended this simple observation to larger sets of randomly or partially ordered constraints, sometimes with remarkable results.

The basic device of these approaches depends on the identification of a number of interacting constraints which, if randomly ordered, would generate the appropriate proportions of observed variants. These proportions depend crucially on the number of relevant constraints. Two constraints have just two possible orders, but the number of possible orders (O) increases as the factorial of the number (n) of constraints ($O = n!$). Thus three constraints have six possible orders (ABC, ACB, BAC, BCA, CAB, CBA), four constraints have 24 possible orders, and so on. If these orders occur at random, the variants they select should occur in the ratios of the number of orders that select them.

This can be illustrated with an extension of our simplified example for CSD. If, in addition to the Faith and *ComplexCoda constraints, the more general markedness constraint NoCoda (disfavoring all coda consonants) were included in the variably ordered set, there would be six possible orders, of which only two – those with Faith highest ranked – would select an undeleted form like *best* or *old*. Hence such a system would generate full forms one-third of the time, and deleted forms two-thirds.

The works mentioned above differ in the details of the model. Anttila mostly relies on fully random subsets of constraints, as in the example just presented, while Nagy and Reynolds work with orders that are fixed for some constraints while others "float" within a certain range. This is illustrated in their study of variable word-truncation in Faetar (a Franco-Provençal language in Southern Italy). A word like *brokele* 'fork' has alternate realizations ['brokələ] ~ ['brokl̩] ~ ['brokə] ~ [brok] (plus a few other rarer forms not considered here). Nagy and Reynolds treat this variation as a consequence of variable ordering of the floating constraint Align-PrWd with respect to the members of a specified set of constraints; the ones relevant to this word are the ranked set *Coda >> Parse >> {HNuc, *Schwa} (HNuc = HarmonicNucleus, disfavoring syllabic consonants). When Align-PrWd (which aligns the right edges of the stressed syllable and the prosodic word) is highest ranked, the form [brok] is selected, because the post-tonic syllables in the other alternatives violate Align-PrWd. When Align-PrWd ranks below *Coda, either ['brokl̩] or ['brokə] is selected; the choice between these two depends on the relative order of HNuc and *Schwa (disfavoring reduced vowels), which are variably ordered with respect to each other. Finally, when Align-PrWd falls below Parse (which requires segments in the input to be maintained on the surface), the full form ['brokələ] surfaces.

Counting up all possible tableaux resulting from the variable orders of the constraints they consider, Nagy and Reynolds find 28 possible orders, of which 16 (57 percent) select ['brokələ], 6 (21 percent) select ['brok], while ['brokə] and ['brokl̩] are selected by 3 tableaux (11 percent) apiece. These are the percentages that Nagy and Reynolds's model predicts should be observed in actual data, assuming they have correctly identified the right constraints and orderings. Table 92.2 shows that the data they observed match these predictions fairly closely.

As in this example, variable OT studies have successfully modeled the empirical distribution of phonological variants in several cases (e.g. Anttila 1997, 2009 on case inflections and vocalic phonotactics in Finnish). But the strict linkage they require between number of varying constraints and predicted frequencies of prevalence raises questions about their general applicability. As we shall see, social variability commonly involves differences in prevalence: class

Table 92.2 Variable OT model of word-truncation in Faetar (from Nagy and Reynolds 1997)

variants	*% predicted*	*% observed*
'bro.kə.lə	57	55
'bro.ka	11	15
'bro.kl̩	11	14
'brok	21	10

differences, gender differences, even stylistic differences in the same individual are typically realized as higher or lower overall rates of use of some form. For example, for the English (*-ing*) variable, seen in alternations like *running/runnin'*, higher rates of the velar variant (*-ing*) are found in more careful speech styles and in higher status speakers and females. But studies rarely report any differences in constraints among these social groups. To generate quantitative differences, variable OT models would require different sets of variably ranked constraints for women and men, for different social classes, and even for a speaker's casual and careful speech styles. Nagy and Reynolds suggest that this is true for some of their speakers:

> women, particularly younger women [favor] full forms of the words. For those speakers, ALIGN-PRWD floats at the lower end of its domain, below PARSE, so that the optimal candidate more frequently has all its segments surface . . . In the grammar of the males and oldest women . . . ALIGN-PRWD has a greater tendency to float at the higher end of its domain, above PARSE. (1997: 47)

But such results are atypical in the sociolinguistic literature on variation; indeed, the empirical evidence suggests that speakers in a community tend to share important grammatical properties like constraint rankings (cf. Labov 1969; Guy 1980).

Exemplar Theory. The usage-based models that have achieved prominence in recent years begin with a very different set of assumptions about mental representations and grammatical processes (Johnson 1997; Bybee 2001, 2002; Pierrehumbert 2001, 2006; Hay and Sudbury 2005; see also CHAPTER 1: UNDERLYING REPRESENTATIONS). Instead of the conventional abstract mental representations of words and speech sounds, these approaches postulate that lexical and phonological units are stored in the mind as concrete memories ("exemplars") of the tokens that a speaker has previously encountered. This set of memories is potentially vast – in principle, the exemplar set for a given word may include all the utterances of that word that one has heard in one's lifetime.

Consequently, variability is directly represented in memory. If a speaker has heard a sound or a word pronounced in variant forms, those exemplars are available. Hence the exemplar set provides each speaker with direct knowledge of variation, of the quantitative ratios at which each variant occurs, and of the contextual facts about which contexts favor or disfavor a variant. Each speaker therefore "knows" precisely the values of parameters like those in (2b), quantifying prevalence. (They also know all the details of scope – social, contextual, and lexical, but these are treated in subsequent sections.)

This knowledge is used in production: when speakers compose utterances, they select production targets from the relevant exemplar clouds. Since these targets vary in the same proportions that the speaker has encountered in the input, the speaker faithfully reproduces this variability in production (subject to limitations by certain other factors discussed in §5). Given so rich a set of mental representations, Exemplar Theory relies very little on the abstract processes that are so prominent in other schools of phonology. Indeed, some versions of the theory deny that such processes exist. Consequently, this framework is little concerned with many of the theoretical debates we have mentioned, such as the domain of linguistic description, competence *vs.* performance, etc.

3 The social limits of generality: Sociolinguistic variation

The most obvious limitation on generality in phonology is that so many phonological phenomena are language-specific. Most of the phonology of Chinese has little resemblance to a phonological account of English, Arabic, or Seneca. The features that differ among these languages are thus, in effect, variable elements of universal grammar: different languages vary in terms of whether they have lexical tone, triliteral roots, nasal vowels, and voicing contrasts in the stops. Linguists often take it for granted that the domain of phonological analysis and generalization is a language, but a moment's reflection shows that there are substantive issues at stake here. A "language" is not necessarily a well-defined domain of description. Some things commonly considered languages, like English and Arabic, clearly encompass a wide range of dialects, social varieties, registers, and speech styles, and the differences between any two language varieties can occupy any point on a continuum from near-identity to complete disparity. Some varieties commonly treated as different languages, like Serbian and Croatian, occupy very similar points on such a continuum. Hence phonology faces a systematic issue of how to define the social limits of a generalization. The universal statement must be modified as follows:

(3) *Quantifying social scope*

For speakers in some social domain *i* . . .

(At this level of abstraction we leave open the question of whether the domain *i* defines a language, dialect, speech community, idiolect, speech style, social class, ethnic group, etc., but we return to this matter below.)

This issue has often gone unaddressed by phonologists. When a phonological analysis refers to the social limits of its range of applicability, these are often vaguely defined or deliberately restricted. Some approaches that are commonly encountered are: (i) informal definition of the social domain, using popular labels for languages ("The sound pattern of English"), localities or dialects ("Juchitán Zapotec") or social identities ("upper middle class white speech"); (ii) description of a defined social set ("my informants," "my idiolect"); (iii) definition by linguistic means, such as shared intuitions (speakers who judge a given form to be grammatical). This approach is of course tautological – the grammar accounts for the speakers who use the grammar.

Such approaches leave open questions. Vague definitions make it difficult to identify which speakers are included or excluded; narrowly restricted ones leave it unclear what relevance the analysis has for speakers outside the limits (How might your idiolect differ from mine? How representative of a broader social universe are the speakers who were studied, or who shared a given intuition?) Of course, in some theoretical schools these are not considered linguistic questions. Vague or narrow social scopes may serve a theoretical end: they externalize diversity, facilitating a more homogeneous description.

These issues reflect an unresolved theoretical debate about the domain of linguistic description. In the generative tradition, this domain is narrowly focused on the concept of a mental grammar containing the knowledge required to produce grammatical utterances. Each speaker possesses such a mental grammar, acquired through the interaction between innate capacities (the language faculty or universal grammar) and individual experience. Since the experiences of each individual are unique, it follows that each individual can potentially possess a unique grammar. In this tradition, the scope of a grammar is the speech of the individual who possesses it, the "idiolect." Social variability is therefore extragrammatical, reflecting only the variability between individuals in linguistic experience. The theory and methods of linguistics per se have little to say about sociolinguistic variation.

The principal alternative to the idiolect treats language as a social construct, with an existence independent of any given individual; indeed, the knowledge of the language that exists in a community of speakers is more comprehensive and arguably more systematic than the partial subset available to a single individual (Labov 1966). Linguistic description therefore takes some larger social entity as its object: a language, dialect, or speech community. Some level of social variability therefore lies within the domain of the grammar of the speech community; indeed, some knowledge of social variation in the community is possessed by each individual, and utilized in commonplace linguistic activities such as accommodation to context and interlocutors.

This social focus has deep roots in linguistic theory. Saussure sees linguistics as principally the science of *langue* – an object that characterizes the systematic and general properties of a language as a whole, rather than of *parole* – in which individual and idiosyncratic properties are located. Similarly, the methodological strictures of American structuralism typically define the appropriate domain of linguistic description as a corpus of observations, which could as easily be drawn from a speech community as from a single informant. Since the 1960s this position has been especially associated with sociolinguistic research (see for example, the discussion of "communicative competence" in Hymes 1972). It is from this tradition that most of what is known about social variability has emerged. Two important theoretical issues arise from this work: the nature of linguistic similarity and difference, and the content of grammar.

A central finding of the research on social variability in phonology is what Weinreich *et al.* (1968) term "orderly heterogeneity": social diversity in language use is neither random nor highly idiosyncratic; rather, it shows great consistency and order. The simplest summary of this orderliness is that social proximity correlates with linguistic similarity: you talk like the people you talk to.

The theoretical problem this presents for phonology is to define what it means for the usage of one speaker to be "like" another. Formal phonology has a good account of identity and non-identity; having the "same" grammar means being identical in all respects, while any differences in what speakers accept as grammatical means that they are "not the same" in grammar. But the theory lacks an account of linguistic similarity. My phonology is nearly identical to my brother's, very similar to those of the friends I grew up with, broadly like other speakers of American English whose ethnic and social backgrounds are comparable to my own, quite different from but recognizably related to that of speakers of Australian English, and drastically different from that of a speaker of Vietnamese. How does phonology capture this scale?

Finally, studies of social variability in language shed important light on the content of grammar – what level of diversity is one grammar capable of modeling?

Table 92.3 Overall rates of coronal stop deletion in early New Zealand English, by national background of speakers

	Probability of deletion
Scots	0.18
English	0.33
mixed	0.22

It seems uncontroversial that Kimbundu and Portuguese should be modeled by two different grammars, and that bilingual Angolans who speak both are therefore possessed of two mental grammars. But what about stylistic variation in one language in an individual: does the switch from a casual style, chatting with one's friends in a bar, to careful style, being interviewed for a job, involve different grammars, or modest adjustments to some stylistic parameters within a single grammar?

Research on social variability in language has yielded some answers to these questions. Consider for example the above-mentioned study of final coronal stop deletion in early New Zealand English. As noted, this corpus includes speakers from several national backgrounds. Table 92.1 showed the three nationality groups had very similar effects of following context on deletion. But other aspects of their usage were not so similar.

The first is difference in prevalence: the groups deleted at different rates. Table 92.3 shows overall rates of deletion by nationality. The differences among the groups are significant. The English delete the most, Scots the least, and the mixed group falls in between.

These results lead to an immediate conclusion: prevalence and constraint effects are independent and orthogonal. The nationality groups are significantly different in prevalence of deletion, but nearly identical in following context effects. This is a result that shows up repeatedly in studies of social variation. For example, speakers of different social classes in a speech community may have markedly different levels of prevalence of a socially significant variant, but show the same constraints: thus Labov's (1966) New Yorkers varied wildly in rates of rhotic articulations, but all produced more /r/ in final than in internal position. And speakers vary their styles by adjusting prevalence, using more or less of prestige variants, but without changing constraint effects.

This leaves the question of whether and when contextual effects can differ. In the New Zealand study, the nationality groups showed some differences with respect to several other linguistic contexts, as can be seen in Table 92.4.

Beside following context, all three groups showed a significant effect of morphology, but only two showed an effect of whether the target stop was /t/ or /d/, a different two showed a lexical frequency effect, and only one showed a preceding context effect.

These results demonstrate that different speech communities often differ in constraint effects. A variety of research results show the same finding: the following zero effect on CSD, although constant for the three national groups in Table 92.1,

Table 92.4 Significant constraints on coronal stop deletion in early New Zealand English, by national background of speakers

	Following context	*Morphology*	/t/ vs. /d/	*Lexical frequency*	*Preceding context*
Scots	✓	✓	✓		✓
English	✓	✓	✓	✓	
mixed	✓	✓		✓	

differs significantly in many speech communities (Guy 1980); the contexts for tensing of /æ/ vary substantially in American English dialects (Labov *et al.* 2006; Labov 2007); the contexts affecting subject pronoun expression vary among Spanish dialects (Cameron 1993; Otheguy *et al.* 2007); metrical constraints on Portuguese /-s/ deletion vary among speech communities in Brazil (Guy 2002). However, these and other studies show that, within speech communities, speakers are mostly similar or identical in contextual effects.

Such findings provide an answer to the question posed above about the content of grammar. Grammatical similarity is measured by shared constraints, not by prevalence; within a grammar, prevalence may vary, but not contextual constraints. Speakers within a speech community share a grammar; therefore, stylistic and social class differences in a community consist of differences in prevalence, but not in constraints. Differences in constraint effects imply different grammars, and different speech communities.

What are the implications for phonological theory? The unquantified models considered above, such as generative phonology and conventional OT, do not engage with such facts. Among the quantified models, these results are naturally accommodated in the VR model, which makes a basic distinction between contextual constraints and prevalence, expressed by an overall probability associated with each rule or grammatical choice-point. They are also easily modeled in Exemplar Theory, since all relevant data on both prevalence and context are stored in each speaker's memory to guide their production; indeed, Foulkes and Docherty (2006) argue that an exemplar account "offers the most productive means of modeling sociophonetic variation." But such results are awkward for OT-based models, because they lack an independent representation of prevalence.

This difficulty can be illustrated with the data on overall prevalence in Table 92.3. In a variable OT model like Anttila's (1997), the differences in deletion rates between the three NZ speaker groups can be modeled only by postulating different sets of variably ordered constraints. The 0.33 deletion probability for the English group could be generated by three constraints (P, Q, R) that select for deletion only when P is ranked highest. But to approximate the deletion rate of 0.22 for the mixed nationality group would require adding an additional constraint S to the variably ordered mix, which also prevents deletion if it outranks P. Then the model predicts deletion in the 25 percent of cases in which P was highest ranked of these four. And modeling the 0.18 rate for the Scots group would require still another mix of five or six constraints. This approach thus implies substantively different grammars for any differences in surface ratios, making no distinction between constant grammar and varying prevalence. A Stochastic OT model has a somewhat easier time, since it can predict different prevalence levels in a constant set of constraints, by varying the distances among them on the ranking scale. But how does Stochastic OT model different prevalence for speech styles or social classes in a community where everybody has the same constraint effects? It is a complex, perhaps impossible, mathematical task to generate constant constraint effects like those illustrated in Table 92.1, while simultaneously generating differences in overall prevalence, like those in Table 92.3, simply by sliding the same constraints up or down the ranking value scale.

4 The contextual limits of generality: Context-sensitivity and variability

Phonological theory commonly makes a distinction between statements, processes, properties, etc. that are context-free ("all syllables have a nucleus") and context-sensitive ("English vowels are ordinarily oral, but may be nasalized before a nasal consonant"). In the latter case, context-sensitive operations involve variability: English vowels vary in their realization on the nasal–oral dimension. This implies that the scope of many generalizations must be defined in some statement such as (4):

(4) *Quantifying contextual scope*

. . . in some linguistic context *j* . . .

It is an interesting fact about the discipline that such limitations on contextual scope are a routine part of phonological analysis, while limitations on social or lexical scope are less often examined and are typically seen as problematic. In fact, phonological theory doesn't even consider contextual limitations as variability, provided prevalence is categorical within the context. Following the Grimm/Verner example, a clearly defined contextual scope in which a unique outcome occurs is not a violation of exceptionlessness. The analytical heuristic that seeks to partial out variable outcomes to categorical contexts holds a privileged status in the conceptual armory of phonology. Nevertheless, context-sensitivity raises several issues associated with variability.

As we have seen, contexts are often associated not with categorical prevalence, but with probabilistic effects. Often a linguistic context favors one outcome without precluding others. This is clear in Table 92.1: each following context has a distinct quantitative effect on CSD, but none categorically demands or prohibits deletion.

Importantly, the following segment effects in Table 92.1 have a phonological explanation consistent with a broad spectrum of work on phonotactics and syllable structure. The words that undergo this deletion process end in consonant clusters: . . . CC#. When they occur in running speech with a following word, sequences like . . . CCC . . . (e.g. *west side*), . . . CCG . . . (*west wing*) and . . . CCV . . . (*west end*) result. Universal principles of markedness, sonority sequencing, etc. all agree that CCC is more marked and less acceptable than CCV (CHAPTER 49: SONORITY; CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS). OT constraints like NoCoda, *ComplexCoda, and Onset are designed to capture such generalizations. Sometimes these principles have categorical effect: many languages prohibit CCC sequences completely. The quantitative findings in English reflect the same principles, but with probabilistic effect.

Such results, echoed in numerous studies of variation, have been termed "stochastic generalizations" (Bresnan *et al.* 2001; Clark 2005): generalizations that are categorically true in one language but are probabilistic constraints in another language or social variety. This constitutes crucial evidence against the more extreme theoretical positions favoring categorical prevalence discussed in §2, such as denying the validity of non-categorical generalizations, and exiling all quantitative facts to the grammatically irrelevant terrain of usage and performance. If phonotactic markedness is a continuum (CHAPTER 4: MARKEDNESS), generating categorical effects in some social or contextual domains and probabilistic effects in others, phonological theory should treat it as an integrated phenomenon.

These results also shed light on another area of variability mentioned in the introduction: phonetic gradience. The intermediate effect of following glides in Table 92.1 (promoting deletion more than vowels but less than consonants) reflects an intermediate status for glides on phonetic/phonological scales. Many phonologists treat this as a sonority scale: glides are more sonorous than obstruents but less sonorous than vowels (CHAPTER 15: GLIDES; CHAPTER 49: SONORITY). Syllable structure and the markedness of consonantal sequences also depend on sonority sequencing: for example, many languages prohibit obstruent-obstruent sequences (CC) but permit obstruent–glide sequences (GC or CG). Viewed in this light, the results of Table 92.1 reflect gradient sonority effects: deletion is inversely correlated with sonority of following context. Phonetic properties thus parallel the quantitative behavior of generalizations: in some social and linguistic domains they are involved in discrete (categorical) phenomena, but in other domains they have probabilistic effects. (For an insightful discussion of gradient variability in an OT framework, see Anttila 2008.)

Finally, as seen in §3, context is also a defining element of linguistic similarity: differences in contextual effects define a difference in grammar, while differences in overall prevalence do not. Members of a speech community converge to a remarkable degree on the contextual limits of phonological generalizations, whether categorical or probabilistic, while they vary considerably in prevalence. Collectively, these results suggest that the typical heuristic of seeking linguistic structure by pursuing Neogrammarianesque categorical contexts is misguided, if the analyst ignores the possibility that such types of contextual conditioning are just one point on a continuum. Context is indeed a fundamental element of linguistic structure, but it includes probabilistic as well as categorical conditioning.

5 The lexical limits on generality: Lexical exceptions

The lexical scope of phonological generalizations has been a recurring topic of debate for over a century. The central question is lexical variability in phonology: does a statement or process apply to all relevant lexical items, or are there words that exceptionally fail to show some generalization (CHAPTER 106: EXCEPTIONALITY)? If so, how does the theory account for lexical limits on generality? A generalization that varies across the lexicon, applying to some words, but not all, requires a specification of lexical scope, as in (5):

(5) *Quantifying lexical scope*

. . . in some lexical domain *k* . . .

The debate over the lexical limits of phonology first achieves prominence in the nineteenth century, in connection with the emergence of the Neogrammarians. The Neogrammarian hypothesis specifically asserts lexical universality of phonological change: "exceptionlessness" means no lexical exceptions to a generalization (Osthoff and Brugmann 1878). In fact, the Neogrammarians propose a phonemic model that rules out lexical variability (Paul 1880): words are constituted as sequences of phoneme-like units; phonologically, they have no independent existence apart from the string of phones of which they are composed. Sound change operates on these phones, so when one of them changes, all words containing them necessarily change as well.

At the same time an anti-Neogrammarian critique emerged, arguing that the historical record actually does show lexical variability (Schuchardt 1885). One slogan for this position was "each word has its own history" – a position that affirms lexical variability as forcefully as the Neogrammarians denied it. In the twentieth century, this position re-emerged under the label "lexical diffusion," in work by Wang and his associates (Wang 1969, 1997; Chen and Wang 1975). These scholars argued against the "lexically abrupt" application of phonological change (i.e. categorical processes applying to all relevant words simultaneously), and in favor of a "lexically gradual" model of phonological process, which spreads across the lexicon word by word, in a manner reminiscent of analogical change (see also CHAPTER 93: SOUND CHANGE).

The focus on synchronic phonology in linguistics after Saussure continued to confront lexical generality. As we have noted, the dominant formal theories emphasized the pursuit of invariant generalizations, and largely assumed, with the Neogrammarians, that words are merely assemblages of the phonological units (such as phonemes, feature arrays, autosegments) on which phonological processes operate. Hence these theories give short shrift to lexical limits on generality other than those that can be given a segmental, prosodic, or morphological formulation. They assume that phonological statements normally apply without lexical limits.

But synchronic phonology has witnessed the reappearance of theories that give primacy to the word, and envisage significant lexical variability. Thus Exemplar Theory argues that the word is the primary unit that speakers recognize, remember, and manipulate, and, indeed, that speakers rely on their massive inventories of remembered auditory images (exemplars) of words they have heard as their primary mental database for most of phonology. Phonemes, features, and the like are emergent abstractions or generalizations across those exemplars, rather than primary units of perception and production. For some versions of the theory, abstract phonological operations do not exist; all of phonology is reduced to "phonetic" processes (neuromechanical events such as gestural overlap, gestural weakening, etc.) and "generalizations" that function more like analogies than like phonological operations.

Such a model consequently assumes that lexical variability is a normal state of affairs. Each word has its own phonological identity, and the mental representation of a word, since it incorporates a broad array of remembered exemplars, necessarily includes all the variability that a speaker has encountered hearing people say that word. The theory permits, even predicts, that any statement that might be made about the phonology of a language might have to be lexically qualified word-by-word across the vocabulary.

Empirical evidence. The issues in this debate are partly empirical. Do we observe lexical limitations on phonology? In the historical record, there are quite a few cases that suggest lexical exceptionality. For example, Latin onset clusters containing /l/ changed in several directions in Old Portuguese (Williams 1938; CHAPTER 30: THE REPRESENTATION OF RHOTICS; CHAPTER 31: LATERAL CONSONANTS). Sometimes /l/ becomes /r/, thus *branco, praia, praça, fraco, cravo, regra* (compare cognates in Spanish, which lacks this change: *blanco, playa, plaza, flaco, clavo, regla*). Other words show the entire cluster changing to the palatal fricative /ʃ/: *clavem* > *chave, plenum* > *cheio, flammam* > *chama*. And in numerous cases Portuguese retains the historical /l/: *flor, flamengo, claro, classe, planta*. There are no obvious contexts that predict one or the other outcome, nor do the usual Neogrammarian tactics for resolving lexical effects appear to offer solutions: the exceptional cases do not obviously arise from dialect borrowing, paradigm leveling, neologisms or classicisms, etc. But it should be emphasized that history records many cases where Neogrammarian regularity does prevail. The Grimm's Law changes in Germanic included a shift from IE voiced aspirates ([bh dh gh]) to Germanic simple voiced stops ([b d g]). This change left no lexical residue whatsoever in the Germanic languages, no words that retain a voiced aspirate. Similarly the loss of the voiced velar fricative in Middle English left behind not a single lexical item in which it was preserved, despite the fact that this sound continues to be spelled in English orthography as <gh>, in words like *cough, though, night*. The empirical evidence thus indicates that both patterns occur, although the literature suggests there are far more cases of regular sound change than of lexical diffusion.

In synchronic phonology the situation is similar: the literature attests many processes that apply without lexical limitations, but a substantial number of cases have been reported showing lexical limits; some of these involve just a few words that are exceptions to a general pattern, while others define substantial subsets of the lexicon that show distinctive phonology, such as the Chinese-origin loanwords in Japanese (CHAPTER 95: LOANWORD PHONOLOGY). These cases of lexical variability have attracted considerable attention in phonological theory.

Theoretical solutions. For a survey of theoretical thought on lexical exceptionality, readers are referred to CHAPTER 106: EXCEPTIONALITY. The principal approaches fall into two camps: lexical strategies, which represent exceptionality in the lexical entries, and phonological strategies, which use the phonological apparatus (features, rules, constraints, etc.) to generate distinctive outcomes for exceptional words. Although these are sometimes treated as technical questions about the workings of a theory, they actually raise substantive questions about the workings of the human mind. The lexical approach – encoding exceptionality in underlying representations – effectively assumes that speakers use minimalist, local, memory-based strategies: they simply remember that a word is anomalous. But the phonological approach implies that speakers strive to extract maximum generalities, and are willing to rejig the whole system if there are any efficiencies to be gained. These are matters worthy of direct investigation.

Lexical frequency. The emergence of Exemplar Theory has brought forward additional questions about lexical variability, and stimulated a range of research

Table 92.5 Lexical frequency effect on coronal stop deletion in early New Zealand English

	Log lexical frequency		
	0	1	2
English	0.42	0.50	0.58
mixed	0.41	0.49	0.56

that has revealed important evidence bearing on lexical variability. Since Exemplar Theory postulates that speakers retain a rich set of memories of utterances of words, they have available a great deal of evidence on which to base their linguistic behavior that other theories ignore. One prominent prediction of the theory is that phonological processes should be sensitive to lexical frequency – how often a speaker encounters a word (CHAPTER 90: FREQUENCY EFFECTS). The exemplar set obviously contains frequency information: frequent words have more exemplars. Scholars in this framework have argued that lenition processes should correlate directly with lexical frequency, because lenition is an articulatory process favored by repetition. This prediction has been confirmed in a range of research in the ET framework. For example, in the New Zealand study previously cited, significant frequency effects were found for CSD (Guy *et al.* 2008). Some relevant findings appear in Table 92.5.

For both groups shown, increasing lexical frequency (measured here by the logarithm of the number of occurrences of the item in the ONZE database) is associated with increasing rates of deletion, consistent with the predictions of Bybee (2001, 2002). Although this effect is treated discretely here, by dividing the frequency range into blocks, Guy *et al.* (2008) demonstrate a continuous correlation.

Lexical frequency effects have been investigated in a variety of studies in usage-based models and elsewhere (e.g. Phillips 1984). Many find significant effects, but the evidence is mixed: some studies show no frequency effects (e.g. the Scots nationality group in the NZ study showed no frequency effect). But the theoretical implications of frequency are extensive. Such facts are difficult to accommodate in any model that assumes economical (i.e. impoverished) lexical representation – in other words, in most theoretical currents from the Neogrammarians to Optimality Theory. Traditional abstract lexical representations provide no place to record how often and in what form the word has been heard, activated, or articulated. If the lexical frequency of items turns out to regularly affect their phonological treatment, richer representations that can incorporate frequency information will be required.

6 Conclusions

One historical task of phonology reflects the perceptual task that hearers face when listening to a speech signal: the signal contains many components, from which the meaningful elements need to be extracted, and the noise and other non-linguistic elements removed. Consequently, phonologists have sought to maximize generality and predictability, to find constancies and regularities, to tell the signal from the noise. Variability in its various forms presents a challenge to this undertaking – irregularities and limitations on generality. Not unreasonably, one response to variability has therefore been to exclude or minimize it, to treat it as part of the noise.

But in taking this approach, phonology runs the risk of reifying the task as the object. Since our method involves seeking generalities, we risk assuming that generalities are all there is to be sought, that phonology is exceptionless, always regular, invariant. This is a perilous course: if we predefine what we seek, we may not be able to find anything else. If we look only for forests, we may not see the trees. A careful scrutiny of the empirical evidence shows that language does contain "irregularities" that are not noise, variability that is generated by and part of the linguistic system. All the schools of thought we have considered recognize this at some level, at least by acknowledging optionality and contextual variability, and some level of lexical variability. Some schools additionally seek to model social variability and quantitative aspects of variable prevalence. A comprehensive account, which departs from the ideal generality – the phonological universal – in each relevant dimension, will have the following form:

(6) *Quantified generality*

For speakers in some social domain i,
in some linguistic context j,
in some lexical domain k,
x is true with a probability p.

The evidence so far available further suggests that the value of p will be some function of i, j, and k. Phonology has made great progress on the task of seeking maximal generality, even universals. It is now beginning to confront the task of exploring the limits of generality, and the linguistic uses of variability.

REFERENCES

Anttila, Arto. 1997. Deriving variation from grammar. In Frans Hinskens, Roeland van Hout & W. Leo Wetzels (eds.) *Variation, change and phonological theory*, 35–68. Amsterdam & Philadelphia: John Benjamins.

Anttila, Arto. 2007. Variation and optionality. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 519–536. Cambridge: Cambridge University Press.

Anttila, Arto. 2008. Gradient phonotactics and the Complexity Hypothesis. *Natural Language and Linguistic Theory* 26. 695–729.

Anttila, Arto. 2009. Derived environment effects in colloquial Helsinki Finnish. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Essays in honor of Paul Kiparsky*, 433–460. Cambridge, MA: MIT Press.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45–86.

Bresnan, Joan, Shipra Dingare & Christopher Manning. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. In Miriam Butt & Tracy Holloway King (eds.) *Proceedings of the LFG 01 Conference, University of Hong Kong, online proceedings*. Stanford: CSLI. Available (July 2010) at http://csli-publications.stanford.edu/LFG/6/lfg01.pdf.

Bybee, Joan. 2001. *Phonology and language use.* Cambridge: Cambridge University Press.
Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14. 261–290.
Cameron, Richard. 1993. Ambiguous agreement, functional compensation, and non-specific *tú* in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change* 5. 305–334.
Cedergren, Henrietta J. 1973. The interplay of social and linguistic factors in Panama. Ph.D. dissertation, Cornell University.
Cedergren, Henrietta J. & David Sankoff. 1974. Variable rules: Performance as a statistical reflection of competence. *Language* 50. 333–355.
Chen, Matthew & William S.-Y. Wang. 1975. Sound change: Actuation and implementation. *Language* 51. 255–281.
Chomsky, Noam. 1965. *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.
Clark, Brady. 2005. On stochastic grammar. *Language* 81. 207–217.
Foulkes, Paul & Gerard J. Docherty. 2006. The social life of phonetics and phonology. *Journal of Phonetics* 34. 409–438.
Grimm, Jacob. 1819. *Deutsche Grammatik.* Göttingen: Dieterich.
Guy, Gregory R. 1980. Variation in the group and the individual: The case of final stop deletion. In William Labov (ed.) *Locating language in time and space*, 1–36. New York: Academic Press.
Guy, Gregory R. 1981. Linguistic variation in Brazilian Portuguese: Aspects of the phonology, syntax, and language history. Ph.D. dissertation, University of Pennsylvania.
Guy, Gregory R. 2002. A identidade lingüística da comunidade de fala: Paralelismo interdialetal nos padrões de variação lingüística. *Organon* 14. 17–32.
Guy, Gregory R., Jennifer Hay & Abby Walker. 2008. Phonological, lexical and frequency factors in coronal stop deletion in early New Zealand English. Poster presented at Laboratory Phonology 11, Wellington.
Halliday, M. A. K. 1985. *An introduction to functional grammar*. London: Edward Arnold.
Hay, Jennifer & Andrea Sudbury. 2005. How rhoticity became /r/-sandhi. *Language* 81. 799–823.
Hay, Jennifer, Margaret A. Maclagan & Elizabeth Gordon. 2008. *New Zealand English.* Edinburgh: Edinburgh University Press.
Hockett, Charles F. 1942. A system of descriptive phonology. *Language* 18. 3–21.
Hymes, Dell H. 1972. On communicative competence. In J. B. Pride & J. Holmes (eds.) *Sociolinguistics*, 269–293. Harmondsworth: Penguin.
Jannedy, Stefanie & Jennifer Hay (eds.) 2006. *Modelling sociophonetic variation.* Special issue, *Journal of Phonetics* 34. 405–530.
Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson & John W. Mullenix (eds.) *Talker variability in speech processing*, 145–165. San Diego: Academic Press.
Kirke, Karen D. 2005. When there's more than one norm-enforcement mechanism: Accommodation and shift among Irish immigrants to New York City. *Penn Working Papers in Linguistics* 11.
Labov, William. 1966. *The social stratification of English in New York City.* Washington, DC: Center for Applied Linguistics.
Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45. 715–762.
Labov, William. 2007. Transmission and diffusion. *Language* 83. 344–387.
Labov, William, Sharon Ash & Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology and sound change.* Berlin & New York: Mouton de Gruyter.

Nagy, Naomi. 1996. Language contact and language change in the Faetar speech community. Ph.D. dissertation, University of Pennsylvania.

Nagy, Naomi & Bill Reynolds. 1997. Optimality Theory and variable word-final deletion in Faetar. *Language Variation and Change* 9. 37–55.

Osthoff, Hermann & Karl Brugmann. 1878. *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*. Leipzig: Hirzel.

Otheguy, Ricardo, Ana Celia Zentella & David Livert. 2007. Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language* 83. 770–802.

Paul, Hermann. 1880. *Prinzipien der Sprachgeschichte*. Halle: Niemeyer. English translation published 1889, *Principles of the history of language*. New York: McMillan.

Peterson, Gordon E. & Harold L. Barney. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24. 175–184.

Phillips, Betty S. 1984. Word frequency and the actuation of sound change. *Language* 60. 320–342.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam & Philadelphia: John Benjamins.

Pierrehumbert, Janet B. 2006. The next toolkit. *Journal of Phonetics* 34. 516–530.

Poplack, Shana. 1979. Function and process in a variable phonology. Ph.D. dissertation, University of Pennsylvania.

Rask, Rasmus. 1818. *Undersögelse om det gamle Norske eller Islandske Sprogs Oprindelse*. Copenhagen: Gyldendal.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Lausanne and Paris: Payot.

Schuchardt, Hugo. 1885. *Über die Lautgesetze: Gegen die Junggrammatiker*. Berlin: Oppenheim.

Swadesh, Morris. 1934. The phonemic principle. *Language* 10. 117–129.

Verner, Karl. 1877. An exception to Grimm's Law. Published 1978 in Philip Baldi & Ronald N. Werth (eds.) *Readings in historical phonology: Chapters in the theory of sound change*, 3–22. University Park: The Pennsylvania State University Press. Translation by R. Stanley from *Eine Ausnahme der ersten Lautverschiebung*.

Wang, William S.-Y. 1969. Competing changes as a cause of residue. *Language* 45. 9–25.

Wang, William S.-Y. (ed.) 1997. *The lexicon in phonological change*. The Hague: Mouton.

Weinreich, Uriel, William Labov & Marvin Herzog. 1968. Empirical foundations for a theory of language change. In Winfred P. Lehmann & Yakov Malkiel (eds.) *Directions for historical linguistics*, 95–195. Austin: University of Texas Press.

Williams, Edwin B. 1938. *From Latin to Portuguese*. Philadelphia: University of Pennsylvania Press.

93 Sound Change

José Ignacio Hualde

1 Sound change and synchronic phonology

This being a chapter on sound change in a series of volumes primarily concerned with synchronic phonology, it seems appropriate to begin by stating the relationship between synchronic and diachronic phonology that will be assumed here. As is well known, for the Neogrammarians a scientific investigation of language could only be undertaken from a historical perspective (Paul 1880). More recently some scholars of phonology have expressed somewhat similar views (see Good 2008a: 11–15 for discussion). Thus Blevins (2004) has argued that historical explanations must be given priority in accounts of synchronic phonological patterns. One of Blevins's main points, also expressed by other authors (e.g. Bybee 2008), is that the phonetic naturalness of phonological patterns is a consequence of their origin in common mechanisms of sound change.

The view that typological tendencies observable in synchronic sound patterns and (morpho)phonological alternations follow primarily from the relative frequency of different sound changes and paths of development (so that "naturalness" resides in diachrony) is not universally held. The alternative is that (in addition) there are principles of UG that dictate the shape of synchronic "phonological grammars." Kiparsky (2008: 52) proposes "a principled separation between true universals, which constrain both synchronic grammars and language change, and typological generalizations, which are the results of typical paths of change." In this view, "true universals" are due to UG.

I believe that, regardless of one's opinion on this matter, the study of synchronic sound patterns should not be reduced to an account of their historical evolution. There is more to synchronic phonology than accounting for the relative frequency or rarity of patterns and alternations. Synchronic and diachronic phonology have different goals. The goal of the diachronic analysis of a phonological pattern is to discover how it developed through time. A synchronic analysis of the same phenomenon, on the other hand, may be concerned with providing a succinct and precise statement of the facts and/or with modeling speakers' knowledge of the phenomenon in question.

Consider, for example, /e/-epenthesis before word-initial consonant clusters starting with /s-/ in Spanish (CHAPTER 67: VOWEL EPENTHESIS). Spanish, unlike

English, French, or Italian, lacks sC- clusters. In the adaptation of borrowings an initial /e/ is automatically inserted, as in *eslogan* 'slogan', *estrés* 'stress', etc. This insertion phenomenon is obligatory, and also pervasive in the second-language pronunciation of native speakers of Spanish. Spanish speakers are typically unaware that there can be a difference between [sC-] and [esC-]. A synchronic phonological account would presumably explain these facts by making reference to the syllable structure of the language.

When we consider the historical origin of the phenomenon, it turns out that this is a very old process in Romance, which originally had a much wider dialectal distribution (see Sampson 2010). Under one hypothesis, a short /i/ was initially inserted when a word starting with sC- was preceded by a consonant-final word, i.e. / C# __ sC, a context where epenthesis is still found in formal or conservative styles in Italian, as in *scritto* 'written' but *per iscritto* 'in writing', *strada* 'road', *in istrada* 'on the road'. In Old French, the context for vowel insertion was expanded and an epenthetical vowel is found both after a consonant and after pause, but not after a vowel: *espose* 'wife' ~ *ta spose* 'your wife', *espée* 'sword' ~ *la spée* 'the sword' (Price 1984). In both Gallo-Romance and Ibero-Romance, epenthesis with these words was then generalized to all contexts, as we can see by comparing the following examples in Latin, Italian, French, and Spanish, among many others that could be given (in French syllable-final /s/ was later lost: sC > esC > ehC > eC).

(1)	*Latin*	*Italian*	*French*	*Spanish*	
	SCRIPTUM	*scritto*	*écrit*	*escrito*	'written'
	SCHOLA	*scola*	*école*	*escuela*	'school'
	STATUM	*stato*	*été*	*estado*	'been'
	STUDIĀRE	*studiare*	*étudier*	*estudiar*	'study'
	SPONSA	*sposa*	*épouse*	*esposa*	'wife'
	SPATHA	*spada*	*épée*	*espada*	'sword'

In Spanish, as mentioned, the rule has remained fully productive as a phonotactic constraint on pronunciation up to the present day. In French, on the other hand, this rule lost its vitality at some point in its history, after which the language started accepting unmodified sC- words, as shown by the fact that we find words like *style*, *spatule*, etc. (cf. Sp. *estilo*, *espátula*). A diachronic analysis would need to account for all these facts: what is the source of the inserted vowel? How did the process become generalized from the phrasal C# __ sC- context to the lexical __ sC- context in Gallo- and Ibero-Romance? How was the insertion rule lost in Gallo-Romance after centuries of productivity? Why didn't the same thing happen in Ibero-Romance? These questions are different from those that arise in a synchronic analysis of present-day Spanish or French phonology.

Frequently, an adequate account of a synchronic alternation or sound pattern may need to differ substantially from its diachronic explanation. It is perhaps useful to briefly make the point that synchronic and diachronic phonology do not need to provide converging accounts. Knowing how an alternation came about does not necessarily tell us how it should be analyzed synchronically. There is, in fact, evidence that speakers sometimes interpret the facts in a way that is not consistent with their diachronic origin. Let us consider an example.

Regular plural formation in English involves a phonologically conditioned alternation (CHAPTER 99: PHONOLOGICALLY CONDITIONED ALLOMORPH SELECTION) among three allomorphs, /-əz/, /-z/, and /-s/. Diachronically, this allomorphy was brought about by a process of vowel deletion, followed by voice assimilation in resulting clusters of obstruents. Comparison of modern English *stone/stones* with Old English *stān* (NOM/ACC SG), *stānas* (NOM/ACC PL) shows us that, historically, a vowel has been lost in the plural form *stones*: Old English /sta:nas/ > Middle English /stɔ:nəz/ > Modern English /sto:nz/. The vowel was not deleted when the stem ended in a strident fricative or affricate, so as to avoid a cluster of strident consonants, as in *churches, kisses*. Historically, then, the ending /-əz/ is primary. It does not follow from this, however, that, from a synchronic point of view, /-əz/, as in *churches* and *kisses*, is the most basic plural allomorph and that a synchronic process of vowel deletion, mirroring the historical sound change, should be postulated to account for *stones* or *cats*. Depending on the analyst's theoretical persuasion, the choice of synchronic analysis may be based on psycholinguistic evidence such as "wug" tests (Berko 1958; CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY), simplicity of the statements, or other criteria. In this case, perhaps the most adequate account of the morphophonological alternation that we find in regular English plurals would posit /-z/, as in *stones*, as the most basic allomorph, and would derive /-əz/ as in *churches, kisses* by an epenthesis rule and /-s/ as in *cats* by a voice assimilation rule or constraint (tautosyllabic clusters of obstruents always agree in voice in English).

A well-known phenomenon, which English plural allomorphy exemplifies in part, is rule inversion (Vennemann 1972). This is the case where a synchronic rule takes the opposite direction from the sound change that caused the alternation. That is, a sound change of the type $x > y$ in context z is phonologized as a synchronic alternation best captured as $y \rightarrow x$ in other contexts.

In Basque, for instance, in the compositional form of a set of nouns a final vowel is deleted and, if the preceding consonant is *-r*, it is changed to *-l*, as schematized in (2a) and exemplified in (2b).[1]

(2) -r/-l *alternation in Basque*

a. *-r*V → *-l* in compositional forms

b.

basic forms		*compounds*	
gari	'wheat'	*gal-buru*	'head of wheat'
atari	'doorway'	*atal-zain*	'doorman'
euskara	'Basque language'	*euskal-dun*	'Basque speaker'
abere	'cattle'	*abel-gorri*	'free-roaming cattle'
merkatari	'merchant'	*merkatal-*	'commercial'

Since this rule is fairly productive, and can be applied to new formations, we may speculate that from knowledge of pairs like those in (2b), Basque speakers can extract something like the rule in (2a).

If we now consider the diachronic origin of the alternation, it turns out that it is to be found in a sound change in the opposite direction. We can see this from

[1] This Basque alternation is also discussed in Hualde (1991: 84) and is used to exemplify rule inversion in Trask (1996).

the treatment of early borrowings from Latin. Borrowings such as *zeru* 'sky' < Lat. CAELU(M), *gura* 'desire' < Lat. GULA, *goru* 'distaff' < Lat. COLU(M), etc. show a change from intervocalic *-l-* to *-r-*. It seems reasonable to assume that the same process affected the native lexicon at the time. The original form of, for instance, 'wheat' may have been **gali*, which became *gari* by rhotacism. In compositional forms intervocalic rhotacism was bled by the deletion of the final vowel (CHAPTER 74: RULE ORDERING). The original lateral of **gali* was thus preserved stem-finally in the compositional form *gal-*. This resulted in pairs with a basic form in *-r*V and a compositional form in *-l* (e.g. *gari* ~ *gal-* 'wheat'), which created a pattern for analogical formations. The compositional form *abel-* in (2b), for instance, is the result of such an analogy, since the source of Basque *abere* 'cattle' is Lat. HABĒRE. In other words, to repeat, the sound change *-l-* > *-r-* (**gali* > *gari*) has given rise to a synchronic rule *-r*V → *-l* (2a) (*gari* → *gal-* in compounds).

To give another example, Stockwell and Minkova (2001: 125) analyze the alternation in cases like English *table* ~ *tabulate*, *single* ~ *singular*, etc., as instances of synchronic "*u*-epenthesis," although from a historical point of view what we have is vowel deletion (in the evolution from Latin to French): Lat. TABULA > Fr. *table*.

The accumulation of sound changes can also give rise to synchronic alternations that lack any phonetic motivation, as in the alternation in English *critic* ~ *criticize*, where /k/ is replaced by /s/ before a low vowel (on this general topic, see Anderson 1981). Such cases of "telescoping" may be capturable by means of more or less natural synchronic rules by positing derivations with a series of intermediate stages, as in classical generative models. Consider for instance the examples in (3) for Ondarroa Basque:

(3) *Ondarroa Basque*

uninflected	*absolutive* sg	
gixon	*gixona*	'man'
sagar	*sagarra*	'apple'
neska	*neski*	'girl'
alaba	*alabi*	'daughter'

As the first two examples in (3) show, the absolutive singular form is normally created by adding the suffix /-a/. In words whose stem ends in /-a/, however, this vowel is replaced by /-i/. This is the result of the accumulation of four distinct sound changes affecting the sequence /a–a/ in this inflectional context: *neskaa* > *neskea* > *neskia* > *neskie* > *neski* 'the girl'. All the intermediate forms are attested in other Basque dialects. In a generative analysis it would be possible to formulate a series of ordered natural synchronic rules that mirror the (necessarily ordered) sound changes that we have illustrated (see Hualde 1991). In a less abstract analysis, different processes could be postulated to account for the pairings of uninflected and singular forms, depending on the last segment of the stem, without synchronically deriving *neski* from /neska + a/ (see Hualde 1999). Knowledge of the historical evolution does not determine the choice of synchronic analysis.

To summarize this section, synchronic and diachronic phonological analysis have different goals. Furthermore, since synchronic phonological patterns may reflect their history only in an indirect way, knowledge of how a given pattern originated does not necessarily provide information on the best synchronic analysis of the facts. The choice of synchronic analysis will depend in part on the analyst's criteria for selecting among competing accounts. A phonologist's goal may be, for instance, to provide an accurate and elegant description of the sound patterns of a given language; or it may be to discover what generalizations regarding the sound structure of the language its speakers actually make, as revealed by psycholinguistic experimentation, and perhaps by modeling them.[2] The diachronic origin of the phonological facts is not directly relevant for either of these purposes.

Sound changes have an effect on the synchronic "phonological grammar" of a language when they give rise to (morpho)phonological alternations or to new phonotactic constraints. But the same type of change that creates robust patterns of alternations in one language may not create any alternations in another language, depending on the structure of the lexicon of the language in question. In the rest of this chapter we will be concerned exclusively with changes in pronunciation, either complete or in progress, but will leave the synchronic phonological analysis of the facts aside for the most part. The issues to be addressed have been topics of debate for many decades, some since the origins of our discipline. Some of these issues will be presented here in a somewhat different light than they are in other contemporary discussions.

2 Phonemes or words as objects of phonological change

A classic but still current debate in diachronic phonology concerns the question of whether sound change affects phonemes (CHAPTER 11: THE PHONEME) in specific phonetic environments, without regard to lexical identity or, alternatively, whether sound change affects words, so that different words will undergo sound changes at different speeds. In recent years, some authors have defended the view that phonetic change always operates on words and that more frequent words will change more rapidly than less frequent ones (CHAPTER 90: FREQUENCY EFFECTS). The possibility (or necessity) of "word-specific phonetics" would appear to follow from the architecture of exemplar models of lexical encoding. Different words will be used in different contexts and with different frequencies, and their reduction patterns will be part of their mental representation (see Pierrehumbert 2002; Bybee 2003).

On the other hand, Labov (1981, 1994, 2006, 2007) has argued that, even though some changes do show lexical effects, the various vowel shifts currently taking place in North American English dialects are affecting all instances of the respective phoneme in the relevant phonological contexts in the same manner and simultaneously, regardless of whether the words containing them are frequent or infrequent. Whatever pronunciation a given speaker has for the vowel of high-frequency *pin* will also be his/her pronunciation of the vowel in the first syllable of low-frequency *pinafore*.

[2] See Hayes *et al.* (2009) for a recent application of the "wug" test methodology, with computational modelling of the results.

This debate goes back to the original formulation of the notion of sound change by the Neogrammarians.[3] The Neogrammarians distinguished two types of process, "(regular) sound change" *stricto sensu* and "analogy" (see Kiparsky 2003 for discussion). Regular sound change is claimed to be phonetically gradual and lexically abrupt. It is lexically abrupt because all words containing the same sound in the relevant phonetic environment are affected in the same manner and simultaneously (see, for instance, Paul 1889: 58–59). Analogical change, on the other hand, is lexically gradual and phonetically abrupt. For instance, pre-Latin **honōsis* 'honor (GEN SG)', **flōsis* 'flower (GEN SG)' became Latin *honōris, flōris* by a regular sound change that, by hypothesis, simultaneously affected all instances of intervocalic /s/, gradually modifying their articulation, s > z > r / V __ V (with phonetically intermediate realizations). Later on, -*s* also became -*r* in the nominative singular of some words, where it was word-final and therefore not in the context of the sound change, as in *honos* > *honor*. The latter was an analogical change that abruptly replaced a phoneme with another phoneme under the influence of other words in the inflectional paradigm. This analogical phenomenon is lexically gradual, because only some words were affected. For instance, *flōs* 'flower (NOM SG)' did not undergo the change (or not until much later).

(4)	Stage I		Stage II (s > *r* / V __ V)		Stage III	
	honos	**honōsis*	*honos*	*honōris*	*honor*	*honōris* (by analogy)
	flōs	**flōsis*	*flōs*	*flōris*	*flōs*	*flōris*

The distinct operation of regular (Neogrammarian) sound change and analogy can also be observed in phonological change in progress. As mentioned above, Labov has claimed that changes like those involved in the Northern Cities Vowel Shift operate with Neogrammarian regularity.[4] On the other hand, the shortening (or laxing) of long /u:/ to /ʊ/ is spreading throughout the lexicon, so that a given speaker may pronounce *roof* with a short vowel and *proof* with a long one. Gradual lexical diffusion involves the abrupt replacement of one phoneme with another. The spread to a new word may be due to analogy with similarly sounding words: e.g. if a speaker knows that *roof* can be either /ru:f/ or /rʊf/, s/he may also start accepting and producing both pronunciations for *proof*, by analogy.

In one type of process, "regular sound change," the change affects sounds or phonemes. In the other type, "analogy," the objects are words (see Labov 1981 for discussion). Another way to interpret this dichotomy is that "regular sound change" and "analogy" differ in the structural level at which they apply. "Regular sound change" operates at the level of meaningless sound units, phonemes and allophones, and it is thus purely conditioned by mechanical, physical aspects of speech. "Analogy," on the other hand, takes into account meaningful sound units, morphemes and words, and may thus be conditioned by both physical and mental aspects.

[3] For an overview of the development of Neogrammarian thinking, see Pedersen (1931). An early dissenting voice was Schuchardt (1885).

[4] The Northern Cities Vowel Shift involves a number of changes affecting different stressed vowels at different stages. First /æ/ rises and diphthongizes; then /a/ and /ɔ/ front; finally, /ɪ/ lowers, /ɛ/ lowers and retracts, and /ʌ/ retracts (Labov 1994: 195).

A third logical possibility is that sound change may be both lexically and phonetically gradual. In this view, the objects of sound change are always words, which undergo gradual, essentially reductive, change (see e.g. Bybee 2008).

It could obviously be the case that different types of sound changes show different paths of development and that only some types of sound changes show Neogrammarian regularity, but what the "words as objects of phonological change" hypothesis denies is that sound change ever operates as the Neogrammarians envisioned. Under this hypothesis it should be the case that those sound changes that have been claimed to operate with Neogrammarian regularity, affecting all lexical items with the relevant phonological environment simultaneously and at the same rate, actually show lexical effects when more carefully examined. This remains to be proven (see Dinkin 2008).

It is useful to make a distinction between reductive and non-reductive sound changes, since they appear to differ in fundamental ways regarding both their origin and their spread (Phillips 1984). Reductive sound changes are those that have their origin in the reduction of magnitude and temporal overlap of articulatory gestures, as typically found in relaxed styles and especially in less informative parts of utterances (see Browman and Goldstein 1991; also CHAPTER 79: REDUCTION).

Not all sound changes appear to be reductive (although most of them certainly are).[5] It should be noted that vowel shifts, on whose evidence Labov has relied to maintain the validity of the Neogrammarian hypothesis, are rather special types of sound changes. Whereas the most frequent sound changes involve reduction of segments in prosodically weak positions, the vowel shifts that Labov has studied target (prosodically strong) stressed vowels. There is no a priori reason to expect that lexical frequency or lexical identity should play the same role in reductive and non-reductive sound changes.

In the next sections we will consider the phonetic and phonological nature of reductive sound change, before turning to the less common non-reductive sound changes. Regarding regularity, the specific view that will be presented here is that many sound changes, including regular reductive phenomena with a well-understood mechanical origin, start as across-the-board phonetically gradual processes, as the Neogrammarians claimed, but involve the operation of something very close to analogy at some point in their phonologization. In other words, regular change initially affects sounds in specific phonetic contexts, but, at a later stage, the sound change is lexicalized. This happens when sounds are phonologically recategorized. The observation that word and morpheme boundaries may condition phonological processes by either preventing their application or allowing a less restricted application than within single morphological domains is relevant for establishing this point. The distinction between regular sound change and analogy may not be as straightforward as the Neogrammarians envisioned.

In principle, the same stages that I am proposing for lenitions would apply to fortitions, although the types of words that would lead in the change at the stage of phonological recategorization would be expected to be different. In this

5 The hypothesis that sound change (when properly defined) is always reductive is put forward in Mowrey and Pagliuca (1995). I do not believe this hypothesis can be maintained, but sound changes that have their origin in articulatory reductions are clearly much more frequent than all other sound changes.

view, the lack of lexical effects in the Northern Cities Vowel Shifts (if Labov is correct) would follow from the fact that there has not been actual phonological recategorization, in spite of changes in vowel quality. At the phonological level there has not yet been a change.

3 The phonetic seeds of reductive sound change

It is fairly well established that common processes such as lenition (CHAPTER 66: LENITION) and assimilation (CHAPTER 81: LOCAL ASSIMILATION) have their ultimate origin in the reduction and co-articulation that we find in casual speech, as opposed to more careful styles. Articulatory phonology provides an explicit model of these phenomena as arising from synchronic variation; in particular, from undershooting and overlap of targets (Browman and Goldstein 1990, 1991 and other work by these authors; see also CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATIONS). In this framework, segmental weakening, including deletion and assimilation, are shown to be the consequence of the online reduction in magnitude and overlap of articulatory gestures. For instance, the reduction in the magnitude of the gesture of the active articulator needed to produce complete occlusion, under time pressure or in relaxed speech, would result in an approximant realization. Assimilation between adjacent targets may likewise result from overlap and reduction of gestures under temporal compression.

In addition to weakenings and assimilations, quite a few other phenomena that are traditionally given very different labels can also result from gesture retiming and reduction. An example would be consonant epenthesis in sonorant-obstruent groups, as in Eng. *som*[p]*thing*, *el*[t]*se*, Lat. TENERA > **tenra* > Fr. *tendre*, etc. In the last example, for instance, the epenthesis is the result of a retiming of the closure of the velic-nasal port with respect to the oral gestures: the cessation of airflow through the nasal cavity while the apico-alveolar contact of /n/ is still maintained will result in the production of a segment identifiable as /d/.

Beddor (2009) shows that American English speakers perceive pronunciations such as [bɛnt] and [bɛ̃t] as being equivalent, the crucial thing being the duration of the nasal gesture, rather than its specific alignment with respect to the dorsal gesture of the vowel and the apical closing gesture at the end of the word. This equivalence could lead to the development of nasal vowels by the progressive favoring of the latter gestural alignment in this context.

Another common diachronic change, word-final devoicing (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION), which at first glance appears very different from lenition and assimilation, can also be understood in reductive terms. Hock (1991: 239–241) argues that word-final devoicing has its origin in prepausal devoicing and is a sort of assimilation (to silence), in spite of the apparent differences.

At least some types of metathesis (CHAPTER 59: METATHESIS) may also be understood as originating in gestural co-articulation. Metathesis would arise from the reinterpretation or resegmentation on the part of listeners of an ambiguous or indeterminate signal created by gestural co-articulation (Blevins and Garrett 1998; Hume 2004). To give an example from sound change in progress, whereas many Spanish dialects have aspiration of /s/ in syllable-final position, a recent development in Andalusian Spanish is the postaspiration of stops following an etymological /s/: *pasta* [pasta] > [pahta] > [patʰa] 'paste'. Torreira (2007) shows that this development, which mirrors historical metathesis in other languages, can be modeled as a re-adjustment in the timing of laryngeal and oral gestures.

Even some dissimilations (CHAPTER 60: DISSIMILATION) may be attributed to gestural mistiming/retiming, to the extent that they involve the coordination of "stretched-out" gestures with other gestures (see Ohala 1993).

It is thus clear that unintended online retiming and reduction of gestures, which are pervasive in unmonitored speech styles, may produce new articulations, as casual speech variants. These online reductions may mirror many of the most common changes that we can observe when we study the history of languages.

An understanding of these phenomena, however, is not sufficient to explain sound change. Casual speech phenomena need to be conventionalized and phonological recategorization must take place for sound change to occur.

4 Conventionalization and recategorization

Sound change involves both phonetics and phonology. It seems fair to say that the phonetic mechanisms that give rise to different reductive sound changes, which we have just briefly considered in the previous section, are better understood than the psychological and social processes that lead to their conventionalization in specific environments and to the recategorization of sounds.

Ohala (1993, 2003) has argued that the operation of sound change requires a mistake on the part of listeners, who misinterpret the sound signal as something different from what the speaker had intended (CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). For instance, the unintended voicing of /p/ between vowels, due to temporal reduction of the labial gesture and incomplete abduction of the vocal folds, is reinterpreted by listeners as intended /b/. As Ohala (2003) points out, however, the ambiguity of speech signals and the consequent potential for errors of interpretation is a pervasive phenomenon. To continue with our example, examination of acoustic records shows that sporadic voicing of intervocalic stops is found in many languages without leading to recategorization or even to the conventionalization of allophonic rules. One would expect sound change to operate at a much higher rate than it actually does if ambiguity systematically led to errors in perception and sound categorization. There must be factors that slow down the progress of sound change.

What may explain the relatively slow pace of sound change is that, at the phonological level, two things need to happen for /p/ to become /b/, as in, for instance, Lat. SAPĚRE > Port. *saber* 'to know', Lat. LUPU(M) > Port. *lobo* 'wolf', etc. First, at some point in time, [b] must become a conventionalized realization of the phoneme /p/ in intervocalic position. At a later point, [b] must be reinterpreted as /b/, a different phoneme, so that perhaps [p] is no longer an option in words that used to have that sound:

(5) *Sound change* p > b

a. Phonetic change	[p]	> [b] / V __ V	/apa/ [apa] > /apa/ [aba]
b. Phonological change (recategorization)	/p/ [b] > /b/ [b]		/apa/ [aba] > /aba/ [aba]

In fact, since [b] is only one of several ways in which [p] may be reduced in this context, the other common one being its reduction to a voiceless bilabial fricative or approximant, a prior phenomenon must be the conventionalization of the reduction process. That is, before recategorization takes place, a specific way of reducing targets must have already been conventionalized in the language. This may involve a choice among competing reductive mechanisms. For instance, in our example, among the various ways to reduce /p/ between vowels, its voicing to [b] may be chosen. Conventionalization of the phonetics involves the recognition of a distinct and acceptable articulatory target, first as an optional allophone and later perhaps as the normal realization of the phoneme in a given context; for example the recognition of [b] as an acceptable allophone of /p/ in intervocalic position.

Phonological recategorization consists of the recognition of a different phonological category. To continue with our example, recategorization takes place when [b] is no longer an allophone of /p/ but rather something different, /b/.

In Portuguese and the other Western Romance languages, Latin word-internal intervocalic /p t k/ were recategorized as /b d g/, becoming identified with the original voiced plosives in other contexts. Thus, for instance, the intervocalic consonant of LUPU(M) 'wolf' > *lobo* was at some point no longer identified with the initial consonant of PORTA 'door', but was now a member of the same phoneme as the initial consonant of BUCCA 'cheek' (> 'mouth').

(6) *Recategorization in Western Romance*

Latin		*Portuguese*	
/p/	PORTA	/p/	*porta*
/p/	LUPU	/b/	*lobo*
/b/	BUCCA	/b/	*boca*

In our example, there has been recategorization because *[lopo] is no longer a possible pronunciation for Portuguese /lobo/ (whereas in Italian dialects with optional voicing of intervocalic stops, completely voiceless realizations are still found; Cravens 2002). I would thus like to suggest that, in the case of reductive sound changes, phonological recategorization is typically preceded by a conventionalization of a specific reductive process. In our example, *p* > *b*, we would have the three stages in (7):

(7) *Stages in sound change* p > b / V __ V

a. *Online gestural reduction*
   Variable voicing and/or incomplete labial closure in casual speech
   /lopo/ [lopo] ~ [lop̬o] ~ [lobo] ~ [loβo] ~ [loɸo]
b. *Conventionalization of phonetics*
   Voicing becomes the conventionalized casual speech variant
   /lopo/ [lobo]
c. *Phonological recategorization*
   Realizations are no longer attributed to underlying /p/
   /lobo/ [lobo]

Conventionalization affects the phonetic realization of phonemes in specific contexts and it is thus lexically abrupt. Recategorization, on the other hand, may occur in a word-by-word fashion. Its subsequent spread across the lexicon may again produce regularity. Whereas recategorization of intervocalic /p t k/ operated quite regularly in word-internal position in Western Romance (but see §6 below for the word-initial position), in Italian it is a lexically very irregular process; compare, for instance, Lat. STRĀTA > Ital. *strada* 'road' and Lat. CANTĀTA > Ital. *cantata* 'sung (FEM SG)' (see Cravens 2006).

To give another example, in present-day Spanish, the voiced obstruents /b d g/ are systematically realized as the approximants [β ð ɣ], with a variable degree of constriction, intervocalically and in some postconsonantal contexts, without regard to lexical identity. This is a fully conventionalized allophonic process. The recategorization of the weakest degree of lenition, which involves the deletion of the segment, on the other hand, is operating in a lexically conditioned manner. Some words have acquired two distinct pronunciations, with and without an intervocalic consonant, but most have not; e.g. *lado* 'side', with two categorically distinct pronunciations (/lado/ [laðo] and /lao/), *recado* 'errand' (/r:ekado/ [r:ekaðo] and /r:ekao/) *vs. enfado* 'anger' (only /enfado/ [eɱfaðo], not */enfao/), *invado* 'I invade' (only /inbado/ [imbaðo], not */inbao/).

Browman and Goldstein (1991) point out that the change from Lat. HABĒRE to It. *avere*, Fr. *avoir* involves a reduction in degree of constriction, from stop to fricative. This is the phonetic aspect. But we also need to explain the conventionalization of this reduction whereby intervocalic [b] came to be regularly pronounced as [v] and subsequently recategorized as /v/, but only in word-internal position.

This sound change is slightly more complex than the voicing of intervocalic /p t k/ in Western Romance. We may note that from the reduction of [b] we would not directly expect a labio-dental fricative [v], but rather an approximant [β] (see Lavoie 2000: 163–164): HABĒRE [abe:re] > [aβe:re]. There is some evidence that this was in fact the initial result. The eventual outcome [v] may have been due in part to merger between intervocalic realizations of /b/ and the original phoneme /w/, as in LAVĀRE [lawa:re] 'to wash'. In all Romance languages there was a merger between -B- [β] and -V- [w]. The result of this merger is eventually [v] in most of the Romance area (Herman 2000: 39). We may assume that first [w] became a labio-dental [v], a relatively common sound change, and that this was followed by the merger between -B- [β] and -V- [v]. We will come back to the restriction of the merger to the word-internal position.

To give another example, long-distance dissimilation, as in Grassman's Law in Sanskrit and Ancient Greek, may have its roots in listeners' "hypercorrection," as claimed by Ohala (1993): in domains containing more than one segment bearing a certain "stretched-out" feature, the listener attributes all the effect to one of the segments. In order for something like Grassman's Law to arise, however, the phenomenon must be conventionalized. "Hypercorrection" does not determine directionality in the dissimilation. It could in principle result in asystematic assignment of the marked feature to only the first potential feature-bearing segment in one word and to only the last one in another. What we find in Grassman's Law is that the last aspiration is systematically retained: $C^h \ldots C^h > C \ldots C^h$ (also in Basque, e.g. *hil* 'dead' + *herri* 'town' > *ilherri* 'cemetery'; Michelena 1985: 211–212). In Quechua, under the hypothesis that present-day Aymara represents the older situation (which is not undisputed; see Landerman 1994), we find the opposite: only the first instance of glottalization or aspiration is retained, cf. Aymara *t'ant'a* 'bread', Quechua *t'anta* 'bread'.

In this section, I have made the point that an important aspect of sound change, beyond its phonetic origins, is phonologization/recategorization. I have suggested that this process may involve the conventionalization of (across-the-board, optional) phonetically conditioned processes followed by (lexically gradual) recategorization.

5 How does recategorization happen in sound change?

As mentioned, in Ohala's model (1993, 2003), sound change takes place when unintended articulatory reductions are reinterpreted by listeners as new intended targets. Intended or underlying forms are subject to online reductions, and, eventually, these online reductions become lexicalized, i.e. intended or underlying. This recategorization, in Ohala's model of sound change, involves a "mistake" on the part of the listener.

For instance, Nolan (1992) demonstrates that coronal assimilation in English, as in *red car*, is a gradual process that allows for a continuum of reduction of the alveolar gesture (CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY). In his response to this paper, Ohala (1992) points out that, even if this is the case, there could be instances where the assimilation is categorical, phonologized, intended as such. In his words, "there may be a huge gap between the faintest version of an alveolar stop in *red car* and the fully assimilated form [rɛgkɑː]" (Ohala 1992: 286). The point is that for sound change to take place, the products of online phonetic reductions must become lexicalized as categorically different. The question is how and when this lexicalization takes place. How can we know whether a given token of complete assimilation was intended as such, and whether it is not just an extreme instance within a continuum of reduced articulations? In the case of coronal assimilation in English, besides the gradual reduction and co-articulation that Nolan documents, could it be that some speakers already optionally have a different target gestural score to produce these sequences, which does not involve an apical gesture? In any event, coronal assimilation appears to be a conventionalized process in English, to the extent that this is something that affects specific sequences in a specific way in this particular language. The phonetic gradual process is a conventionalized phenomenon of English which may apply in a Neogrammarian fashion, without regard to the identity of the lexical items involved. On the other hand, phonological recategorization as something different from the historical sequence may operate on an item-by-item basis, showing frequency effects.

The traditional demonstration that re-analysis has taken place is provided by the occurrence of the new target in contexts beyond the one where it arose. For instance, we can be sure that listeners have interpreted the acoustically overlapped /t/ in *perfec(t) memory* (to use one of Browman and Goldstein's 1990 examples) as being intentionally absent when, as speakers, they start producing sequences such as *perfe*[k] *art*, where the absence of the word-final [t] cannot be attributed to acoustic masking. In the case of word-internal segments, however, this evidence may not be available, as the phonetic context is constant.

Recategorization clearly takes place earlier than has sometimes been assumed. Janda (2003) discusses Twaddell's explanation of the phonologization of umlaut. As Janda points out, at the time it was proposed, Twaddell's account was seen as real progress. Essentially the idea is the following. Umlaut was a predictable allophonic process as long as the triggering high front vowel was present (Stage I in (8a); I illustrate with a familiar pre-English example). At this stage [y] would be a predictable allophone of the phoneme /u/ in the environment of a following /i/. We may postulate the synchronic rule in (8b) to capture the distribution of allophones. Later, the reduction of final unstressed /i/ to schwa made the presence of [y] unpredictable from the context, thus triggering its categorization as an independent phoneme (Stage II):

(8) *Umlaut*

a. Stage I / Stage II
/mu:si/ [my:si] > [my:sə] /my:sə/ 'mice'

b. /u:/ → [y:] / __ (C) i

Janda (2003), following several other critics of Twaddell's analysis, calls our attention to a conceptual problem with this account. If umlaut was a predictable effect at Stage I, triggered by the phonological environment, it should have gone away once the conditioning factor disappeared. That is, if the synchronic allophonic rule was as in (8b), the centralization of the conditioning final vowel should have resulted in [my:si] changing to *[mu:sə]. That is, once the final vowel could no longer condition its fronting, the phoneme /u/ should surface as [u]. The conclusion that Janda draws is that the umlauted [y] vowel became a distinct phoneme while the original conditioning environment was still present. Whether or not we want to use the term "phoneme" in this sense (CHAPTER 11: THE PHONEME), the lesson clearly is that [y] was an intended, conventionalized target, part of the gestural score of the word (in Articulatory Phonology terms) at some historical point when final /i/ was also part of the gestural score.

If phonologization must precede loss of the conditioning environment, the question is, then, what causes it. Janda makes the reasonable assumption that when two allophones are phonetically distinct enough they will be learned as distinct categories, even if they are in complementary distribution (for discussion see also Kiparsky 1995: 656–657). It is unclear, however, how one can determine the relevant degree of dissimilarity that would warrant independent phonemic categorization (although this could, in theory, be empirically investigated through psycholinguistic experimentation). We must also ask the question of how the fronting of /u/, caused by co-articulation, was exaggerated to the point that speakers of the language reinterpreted it as a different target. One possibility is that this was an abrupt change: at some point, Old English listeners reinterpreted very co-articulated, fronted productions of /u/ as a distinct target /y/, maybe on a word-by-word basis. In principle, Ohala's listener-induced abrupt recategorization seems more compatible with changes that show lexical diffusion or are sensitive to the presence of morpheme or word boundaries than with sound changes that apply across the board. Umlaut and vowel harmony rules are typically limited to (prosodic) word domains and tend to have exceptions and/or be limited to certain morphological contexts. Co-articulation of the type that may give rise to umlaut has been observed across words sequences (Cole *et al.* 2010). If this co-articulation is not phonologized (as vowel assimilation) across word boundaries (e.g. hypothetical *bl*/u/ *boat vs. bl*/y/ *sea*), it must be because analogical pressure among tokens of the same word prevents this from happening.

To give another example, consider the difference in the duration of vowels preceding voiced and voiceless coda consonants in present-day American English. This difference goes considerably beyond what may be expected from merely phonetic co-articulatory effects and is sufficiently salient to serve as a cue for the voicing of the following consonant (Peterson and Lehiste 1960; Rafael 1972; Rafael *et al.* 1975). Final obstruents in English show a tendency to devoice (Flege and Brown 1982; Smith 1997). We may thus envision a future scenario where word-final devoicing of obstruents becomes an obligatory rule of English, as has happened in German (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). It would not be surprising in this hypothetical scenario if pairs such as *bet* and *bed* were still distinguished by the duration of the vowel: /bɛt/ *vs.* /bɛːt/. In this scenario, final devoicing would have caused the phonologization of a formerly predictable difference in vowel duration, as has happened, for instance, in Friulian (Vanelli 1979). Arguably, however, this vowel duration contrast has already been phonologized in present-day English. That is why we wouldn't expect it to disappear by the systematic devoicing of word-final obstruents. On the other hand, to the extent that vowels preceding both voiced and voiceless obstruents are still treated in a similar manner in phenomena such as the various vowel shifts that are taking place in English, it must be concluded that speakers still categorize over both contexts. What appears to be needed to understand sound change is a richer theory of categorization than is traditionally assumed (see e.g. Hualde 2004 and Ladd 2006 for proposals). Unlike the example that we considered above, umlaut in German and in Old English, which is restricted to certain morphological contexts (Janda 1998), durational differences in English vowels conditioned by the phonological voicing of a following consonant apply without exception. Vowel duration has been conventionalized, but no phonemic recategorization has occurred.

6 Word boundary effects

Unless word boundaries can be demonstrated to have phonetic correlates in the language, they should be irrelevant for "regular sound change." For instance in Spanish, /b d g/ are spirantized both within words and across word boundaries, e.g. *Maria bebe agua* [maɾiaβeβeaɣwa] 'Maria drinks water'. The spirantization of /b/ in Amharic is similarly an across-the-board phenomenon: [bet] 'house', [kaβet] 'from the house'. In some historical lenitions, however, we observe that rephonologization is restricted to the word domain.

In the Western Romance voicing of intervocalic stops, which we mentioned in the preceding section, only word-internal intervocalic instances of /p t k/ were recategorized as /b d g/; e.g. Lat. LATU(M) > Sp./Port. *lado* 'side' *vs.* Lat. ILLA TERRA > Sp. *la tierra*, Port. *a terra* 'the land'. On the other hand, outside of the Western Romance area, in central and southern Italian dialects, in Corsican and Sardinian, the voicing of these segments has an allophonic or optional character and applies both inside words and across word-boundaries, e.g. Lat. LATU(M) > Sard. [laðu], Lat. TERRA > Sard. [tɛrːa] 'land', Lat. IPSA TERRA > Sard. [saðɛrːa] 'the land' (Jones 1997: 377). As Weinrich (1958) pointed out, the area where Lat. /p t k/ underwent systematic recategorization coincides with those languages where the voicing is restricted to the word-internal domain. The question that arises is how to account for this difference in the domain of application of the sound change.

Examination of synchronic variation and possible sound change in progress appears to show that at the initial stages intervocalic lenition applies across the board. In some Spanish dialects, for instance, we are witnessing an incipient "second round" of lenition of intervocalic /p t k/. In a recent study, Hualde *et al.* (2010) found about 30 percent of partially or completely voiced tokens of intervocalic /p t k/ in a sample of spontaneous speech. Other studies have found roughly comparable percentages, sometimes higher, depending on the specific dialect, the speaking style, and the speaker (Machuca Ayuso 1997; Lewis 2001). Leniting voicing is found both word-internally and across word boundaries.

Both the state of affairs in other areas, such as Sardinia and southern and central Italian areas with "phonetic" voicing, and the examination of the incipient second round of voicing witnessed in some Spanish dialects lead us to the conclusion that voicing must initially also have operated across word boundaries in Western Romance (as Weinrich 1958, Hall 1964, and Cravens 2002 also conclude). As voicing became obligatory and categorical, it became restricted to the word-internal context. The elimination of voicing across word boundaries must be attributed to the operation of analogy among tokens of the same word in different syntactic contexts. In a phrase like ILLA TERRA > hypothetical Western Romance [latɛrːa] ~ [ladɛrːa] 'the land', the voiceless realization would eventually prevail under the influence of phrase-initial and postconsonantal tokens of the word /tɛrːa/, including contexts of word-initial gemination (from consonant assimilation): AD TERRĀ > [atːɛrːa] 'to land'. Sporadically, however, we find generalization of the voiced variant instead, as in Lat. COLAPHUS > Sp. *golpe* 'strike'.

We find the same restriction to the word-internal context in Browman and Goldstein's (1991) example of lenition as reduction in constriction degree: Lat. HABĒRE > It. *avere*. Whereas Latin word-internal intervocalic B and V have merged in all the Romance languages, the intervocalic [b] of, for instance, ILLA BUCCA did not become [v] in It. *la bocca* [labokːa] 'the mouth', Fr. *la bouche*.

As mentioned above, Latin -B- /-b-/ and -V- /-w-/ have merged in all Romance languages. In French, Italian, and standard Portuguese, the modern result is /v/:

(9)

	Italian	*French*	*Portuguese*	
HABĒRE	*avere*	*avoir*	*haver*	'have'
DEBĒRE	*dovere*	*devoir*	*dever*	'must'
LAVĀRE	*lavare*	*laver*	*lavar*	'wash'

In word-initial position on the other hand, these languages preserve an etymological contrast between /v/ and /b/:[6]

(10)

	Italian	*French*	*Portuguese*	
BUCCA	*bocca*	*bouche*	*boca*	'mouth'
BONA	*buona*	*bonne*	*boa*	'good (FEM SG)'
VACCA	*vacca*	*vache*	*vaca*	'cow'
VŌCE	*voce*	*voix*	*voz*	'voice'
VŌS	*voi*	*vous*	*vos*	'you (PL)'

[6] In Old Castilian Spanish the contrast may have been between /β/ and /b/. This contrast was subsequently lost, in post-medieval times, through the lenition of /b/.

As mentioned in the previous section, in these Romance evolutions, the first relevant change is the lenition of Latin intervocalic /b/, which subsequently results in its merger with /w/.[7] Again, the merger would also be expected to affect word-initial B- and V- when intervocalic; that is, ILLA BUCCA, for instance, should have undergone lenition to [ilːaβukːa], and its B should have acquired the same articulation as the V in ILLA VACCA [ilːavakːa] when the contrast between the medial labials of HABĒRE [aβeːre] > [aveːre] and LAVĒRE [lawaːre] > [lavaːre] was lost:

(11) *Hypothesized evolution of intervocalic* -B- *and* -V-

HABĒRE	>	aβere	>	avere
LAVĀRE	>	lavare	>	lavare
BUCCA	>	bukːa	>	bukːa
ILLA BUCCA	>	ilːa βukːa	>	ilːa vukːa
VACCA	>	vakːa	>	vakːa
ILLA VACCA	>	ilːa vakːa	>	ilːa vakːa

At the last stage represented in (11), the phonemes /b/ and /v/ are in contrast in word-initial position only if not intervocalic (e.g. after pause or consonant). There is no phonemic contrast in intervocalic position, either word-internally or phrase-internally. The result is that words that have different initial consonants in some contexts do not in other contexts.

No Romance language has preserved the situation depicted in (11). We may conclude that languages like Italian, French, and Portuguese have analogically re-established the contrast between /v-/ and /b-/ also after a vowel, so that a given word either begins with [v-] or with [b-] in all phrasal contexts (Weinrich 1958).

The frequent cases of confusion between initial *b*- and *v*- in old texts from central Italy and even in graffiti from Pompei (see Tekavčić 1972: 142–144) provide quite strong evidence for the hypothesis that the phonemic contrast was indeed analogically re-established in word-initial postvocalic position, after a period where the two phonemes were contextually neutralized, as proposed by Weinrich (1958).

In the Romance lenition processes that we have just considered, a sound change was prevented from operating across word boundaries as it became phonologized. In other cases, the opposite seems to have happened, with extension of the process to further contexts across word boundaries.

Although in American English the flap /ɾ/ originally arose from the weakening of /t/ and /d/ in certain contexts, it is now, arguably, a distinct articulatory target (CHAPTER 113: FLAPPING IN AMERICAN ENGLISH). Speakers show awareness of its existence as a different sound. A word like *better*, for instance, has a flap in its normal pronunciation in the relevant dialects, but speakers may replace it with /t/ for stylistic purposes. Flaps are found both word-internally and word-finally. All words ending in /t/ in their citation form can be pronounced

[7] Whereas we cannot be sure when /b/ started to admit approximant realizations or even when these approximant realizations of intervocalic /b/ became the usual articulatory target, we know that -B- /b/ merged with -V- /w/ by the first century of our era. This is because around this time the graphemes B and V start getting confused (Weinrich 1958: 87; Allen 1978: 41).

with /ɾ/ before a vowel in the next word, as in *but again, at all, forget about, eat oats*. Interestingly, the phonetic conditions for the occurrence of the flap are different word-internally and across word boundaries, as shown by the impossibility of flapping in *atoll* and its possibility in *at all*. Inside words, the segment must be preceded by the stress and followed by an unstressed vowel. Across word boundaries, however, flapping is not conditioned by stress and may occur immediately before a stressed vowel, as in *ate apples*; this is also the case in a compound such as *whatever* (cf. *potato*).

Given the importance of word stress in English as a conditioning factor in reductive processes, I assume that the more restrictive phonetic conditioning that is still found word-internally was the original one and that, across word-boundaries, the change was generalized to all prevocalic contexts, not only unstressed ones.

Flapping is a conventionalized reduction of /t/, since it is only one of the possible phonetic reductions that it may undergo. Other possibilities, all of them found in the same contexts in other English dialects, are the replacement of [t] with a glottal stop, its affrication to [ts], and its spirantization to [θ]. It seems reasonable to assume that after [ɾ] was conventionalized as a replacement for [t] before unstressed vowels it was analogically extended to the context before stressed vowels when word-final.

Bybee (2000, 2003) discusses a similar example, the extension of the aspiration of /s/ in Spanish dialects from the preconsonantal context, where it is found word-internally, as in *este* [ehte] 'this (MASC SG)', to the prevocalic context, where one may have aspiration across word boundaries in the relevant dialects, as in *los animales* [lohanimale] 'the animals'. In Bybee's account, the extension of aspiration to the intervocalic context across word boundaries involves a sort of analogy. In a first stage of development, the plural masculine article *los*, to give an example, would have aspirated realizations in phrases such as *los toros* 'the bulls', where the final /s/ is preconsonantal, but not in *los animales* 'the animals', where it is prevocalic and, thus, not in the context of aspiration. The fact that s#C is more common than s#V (i.e. there are more consonant-initial than vowel-initial words) implies that at this hypothetical stage [loh] ~ [los] would be more frequent than invariant [los], and the alternation would subsequently be extended from the more frequent to the less frequent context. Bybee shows that the distribution of aspiration in several Spanish dialects is consistent with this hypothesis. In the view that I am defending in this chapter, it is at this second stage where we are more likely to find lexical effects. An example would be intervocalic word-internal aspiration in *nosotros* /nos + otros/ [nohotro] 'we', where a morphological boundary is discernible (cf. *los otros* [lohotro] 'the others') *vs.* e.g. word-internal *ositos* [osito], **[ohito] 'little bears', in the relevant dialects. This is similar to flapping in *whatever*, as we saw above.

The phonologization of phrase-final devoicing at the word level provides another example of analogical extension. As mentioned above, a reasonable hypothesis is that word-final obstruent devoicing, as we find in German, starts out as phrase-final devoicing and then is extended to the word domain (Hock 1991: 80). The phonetic causes of phrase-final devoicing are reasonably clear, as this is in fact a phenomenon that is easily observable in many languages, affecting all kinds of word-final segments, not only obstruents. This phonetic effect may at some point become conventionalized for phrase-final obstruents. After that, it may be analogically generalized to the word-domain, so that alternations of the type -p## ~ -b#V are eliminated by extension of the word variant ending in a voiceless consonant.

7 Non-reductive sound change

As mentioned in §2, the main evidence that Labov adduces in favor of the Neogrammarian orthodoxy comes from the study of vowel shift in progress. These vowel shifts in English crucially differ from reductive processes in that they affect segments in prominent positions, i.e. stressed syllables. In fact, Labov points out that one finds more advanced vowel tokens in the direction of the shift in syllables with primary stress than in syllables with secondary stress, vowels in words with emphatic stress being especially advanced along the path of the vowel shift (Labov 1994: 122–123, 173, 195). This is quite different from reduction processes, where we expect to find the least reduced tokens under emphatic stress.

The changes involved in vowel shifts are essentially the same types of processes as other changes targeting stressed vowels, both conditioned (umlaut/metaphony) and unconditioned, including diphthongization, mora dissimilation within diphthongs and rising of (tense/long) vowels. In at least some of these changes it is clear that the added duration of vowels in certain positions led to their recategorization. Thus, in Spanish, the phonologically short mid vowels of Latin, which had acquired a more open quality /ɛ ɔ/, gave rise to rising diphthongs when lengthened under stress. In French they diphthongized in the same manner, but only in syllables that were stressed and open, cf. Lat. PEDE(M) > Sp. *pie*, Fr. *pied* 'foot' *vs.* Lat. PERDIT > Sp. *pierde*, Fr. *perd* 's/he loses'. The long mid vowels of Latin also diphthongized in French, in this case producing falling diphthongs, but again only in stressed open syllables, where the extra length would presumably favor the perception of a difference in quality between the beginning and the end of the nucleus. In this case, further developments show the sort of mora differentiation that we find in some of the vowels involved in the English vowel shifts; cf. Lat. TĒLA > [tei̯lə] > [təi̯lə] > [toi̯lə] > [twel] > [twal] *toile* 'cloth'. Notice that under the hypothesis that greater stress causes greater vowel duration and longer vowels are more likely to be recategorized as diphthongs (and diphthongs as containing a more extreme change), if anything, we would, expect the least frequent words to lead these changes, since they would be more likely to receive phrasal stress.

Besides changes targeting stressed vowels, the most obvious examples of non-reductive changes are consonant fortitions. Whereas experimental work has pointed out the existence of domain-initial strengthening (Fougeron and Keating 1997; Cho and Keating 2009), its phonologization as word-initial fortition is a rare phenomenon. When they happen, these changes for the most part seem to have non-phonetic, analogical origins. For instance, the strengthening of word-initial /r/, /l/, and /n/ in different Ibero-Romance languages, where they give the same result as word-internal geminates (Lat. ROSA > Sp. [rːosa], Lat. LŪNA > Cat. *[lːuna] > *lluna* [ʎuna] 'moon', Lat. NŪBES > Leonese [ɲuβes] 'clouds'), is most likely an extension from contexts where a word-initial geminate arose from consonant assimilation (AD LŪNA > [alːuna]; see Cravens 2002). To the extent that these changes are correctly analyzed as purely analogical, they operate on words, and we may expect lexical effects.

The fortition of word- and syllable-initial glides, as repeatedly found in the history of the Romance languages (e.g. Lat. IOCĀRE > It. *giocare*, Fr. *jouer*, Sp. *jugar* 'to play'; Sp. *huevo* [weβo] ~ [gweβo]; also in Basque **e-an* > *jan* [ɟan] ~ [ʤan] ~ [ʒan] ~ [ʃan] ~ [xan] 'to eat') remains ill understood. In the several successive processes of syllable-initial palatal glide fortition in the history of Spanish (and also of Basque), the recategorization of the palatal glide shows lexical effects, e.g. IAM > Sp. *ya* 'already' *vs.* IAM MAGIS > OSp. /ʒamas/ > Sp. *jamás* /xa'mas/ 'never'; Argentinian Sp. *yerba* /ʒerba/ *(mate)* 'maté leaves' *vs.* *hierba* /ierba/ 'grass', etc. The hypothesis would be that fortition of glides starts in prosodically strong phrase-initial contexts, and is later phonologized as a word-initial phenomenon, which may then spread to word-internal syllable-initial contexts.[8] On the other hand, the fortition that we observe in examples such as Lat. VACCA [wak:a] > OSp. [βaka] > Sp. *vaca* [baka] 'cow' is the indirect (analogical) result of the weakening of intervocalic /b/ at two distinct stages in the history of the language.

8 Summary

This chapter has focused to a large extent on the classic issue of the regularity of sound change (i.e. the Neogrammarian hypothesis) which, in the context of recent debate, may be termed the question of whether there are sound changes that operate purely on phonemes in specific phonological contexts, without regard to the lexical identity of the words containing them. I have proposed that in many common sound changes we should distinguish two stages. The first stage is the conventionalization of a phonetic process. Conventionalized phonetic processes operate without regard to lexical identity. At a second stage, there may be phonological recategorization, which will tend to operate on a word-by-word basis. The spreading throughout the lexicon of this recategorization (by analogy) will produce the effect of regularity. Phenomena such as word boundary effects may be an effect of analogical spreading.

We have studied reductive and non-reductive sound changes separately, since their paths of development can be reasonably expected to be different. Whereas significant progress has been made in our understanding of the phonetic origin of reductive sound changes, non-reductive changes, including those affecting stressed vowels and consonant and glide fortition, are less well understood.

ACKNOWLEDGMENTS

I am grateful to Jennifer Cole, the anonymous reviewers, and the editors for their valuable comments.

[8] Another ill-understood phenomenon is the gemination of consonants before glides, as found in Italian, cf. *avete* 'you (PL) have' *vs.* *abbiamo* /ab:jamo/ 'we have', Lat. AQUA > It. *acqua* /ak:wa/ 'water', etc.

REFERENCES

Allen, W. Sidney. 1978. *Vox Latina: The pronunciation of Classical Latin*, 2nd edn. Cambridge: Cambridge University Press.

Anderson, Stephen R. 1981. Why phonology isn't "natural." *Linguistic Inquiry* 12. 493–539.

Beddor, Patrice Speeter. 2009. A co-articulation path to sound change. *Language* 85. 785–821.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette & Andrew Garrett. 1998. The origins of consonant–vowel metathesis. *Language* 74. 508–556.

Browman, Catherine P. & Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341–376. Cambridge: Cambridge University Press.

Browman, Catherine P. & Louis Goldstein. 1991. Gestural structures: Distinctiveness, phonological processes, and historical change. In Ignatius G. Mattingly & Michael Studdert-Kennedy (eds.) *Modularity and the motor theory of speech perception*, 313–338. Hillsdale, NJ: Lawrence Erlbaum.

Bybee, Joan. 2000. Lexicalization of sound change and alternating environments. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, 250–268. Cambridge: Cambridge University Press.

Bybee, Joan. 2003. Mechanisms of change in grammaticization: The role of frequency. In Joseph & Janda (2003), 602–623.

Bybee, Joan. 2008. Formal universals as emergent phenomena: The origins of structure preservation. In Good (2008b), 108–121.

Cho, Taehong & Patricia Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37. 466–483.

Cole, Jennifer, Bob McMurray, Gary Linebaugh & Cheyenne Munson. 2010. Unmasking the acoustic effects of vowel-to-vowel co-articulation: A statistical modeling approach. *Journal of Phonetics* 38. 167–184.

Cravens, Thomas. 2002. *Comparative historical dialectology: Italo-Romance clues to Ibero-Romance sound change*. Amsterdam and Philadelphia: John Benjamins.

Cravens, Thomas. 2006. Microvariability in time and space: Reconstructing the past from the present. In Thomas Cravens (ed.) *Variation and reconstruction*, 17–36. Amsterdam & Philadelphia: John Benjamins.

Dinkin, Aaron. 2008. The real effect of word frequency on phonetic variation. *Penn Working Papers in Linguistics* 14. 97–106.

Docherty, Gerard J. & D. Robert Ladd (eds.) 1992. *Papers in laboratory phonology II: Gesture, segment, prosody*. Cambridge: Cambridge University Press.

Flege, James & W. S. Brown. 1982. The voicing contrast between English /p/ and /b/ as a function of stress and position-in-utterance. *Journal of Phonetics* 10. 335–345.

Fougeron, Cécile & Patricia Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101. 3728–3740.

Good, Jeff. 2008a. Introduction. In Good (2008b), 1–19.

Good, Jeff (ed.) 2008b. *Linguistic universals and language change*. Oxford: Oxford University Press.

Hall, Robert A., Jr. 1964. Initial consonants and syntactic doubling in West Romance. *Language* 40. 551–556.

Hayes, Bruce, Kie Zuraw, Péter Siptár & Zsuzsa Cziráky Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85. 822–863.

Herman, József. 2000. *Vulgar Latin*. Translated by Roger Wright. University Park, PA: Pennsylvania State University. Originally published (1967) as *Le latin vulgaire*. Paris: Presses Universitaires de France.

Hock, Hans Henrich. 1991. *Principles of historical linguistics*, 2nd edn. Berlin & New York: Mouton de Gruyter.

Hualde, José Ignacio. 1991. *Basque phonology*. London & New York: Routledge.

Hualde, José Ignacio. 1999. Vowel interaction and related phenomena in Basque and the nature of morphophonological knowledge. *Cognitive Linguistics* 10. 33–56.

Hualde, José Ignacio. 2004. Quasi-phonemic contrasts in Spanish. *Proceedings of the West Coast Conference on Formal Linguistics* 23. 374–398.

Hualde, José Ignacio, Miquel Simonet & Marianna Nadeu. 2010. Consonant lenition and phonological recategorization. Paper presented at the 12th Conference on Laboratory Phonology, Albuquerque, NM.

Hume, Elizabeth. 2004. The indeterminacy/attestation model of metathesis. *Language* 80. 203–237.

Janda, Richard. 1998. German umlaut: Morpholexical all the way down from OHG (two *Stützpunkte* for Romance metaphony). *Rivista di Linguistica* 10. 165–234.

Janda, Richard. 2003. "Phonologization" as the start of dephonetization – or, on sound change and its aftermath: Of extension, generalization, lexicalization, and morphologization. In Joseph & Janda (2003), 401–422.

Jones, Michael A. 1997. Sardinian. In Martin Maiden & Mair Parry (eds.) *The dialects of Italy*, 376–384. London: Routledge.

Joseph, Brian D. & Richard Janda (eds.) 2003. *The handbook of historical linguistics*. Malden, MA & Oxford: Blackwell.

Kiparsky, Paul. 1995. The phonological basis of sound change. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 640–670. Cambridge, MA & Oxford: Blackwell.

Kiparsky, Paul. 2003. The phonological basis of sound change. In Joseph & Janda (2003), 313–342.

Kiparsky, Paul. 2008. Universals constrain change; Change results in typological generalizations. In Good (2008b), 23–53.

Labov, William. 1981. Resolving the Neogrammarian controversy. *Language* 57. 267–308.

Labov, William. 1994. *Principles of linguistic change*, vol. 1: *Internal factors*. Oxford: Blackwell.

Labov, William. 2006. A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics* 34. 500–515.

Labov, William. 2007. Transmission and diffusion. *Language* 83. 344–387.

Ladd, D. Robert. 2006. "Distinctive phones" in surface representation. In Louis M. Goldstein, Douglas Whalen & Catherine T. Best (eds.) *Laboratory phonology* 8, 3–26. Berlin & New York: Mouton de Gruyter.

Landerman, Peter. 1994. Glottalization and aspiration in Quechua and Aymara reconsidered. In Peter Cole, Gabriella Hermon & Mario D. Martin (eds.) *Language in the Andes*, 332–378. Newark: University of Delaware.

Lavoie, Lisa. 2000. Phonological patterns and phonetic manifestations of consonant weakening. Ph.D. dissertation, Cornell University.

Lewis, Anthony. 2001. Weakening of intervocalic /p, t, k/ in two Spanish dialects: Toward the quantification of lenition processes. Ph.D. dissertation, University of Illinois at Urbana.

Machuca Ayuso, María Jesús. 1997. Las obstruyentes no continuas del español: Relación entre las categorías fonéticas y fonológicas en habla espontánea. Ph.D. dissertation, Universitat Autònoma, Barcelona.

Michelena, Luis. 1985. *Fonética histórica vasca*, 3rd edn. San Sebastián: Diputación de Guipúzcoa.

Mowrey, Richard & William Pagliuca. 1995. The reductive character of articulatory evolution. *Rivista di Linguistica* 7. 37–124.

Nolan, Francis. 1992. The descriptive role of segments: Evidence from assimilation. In Docherty & Ladd (1992), 261–280.

Ohala, John J. 1992. Comments on Chapter 10. In Docherty & Ladd (1992), 286–287.

Ohala, John J. 1993. The phonetics of sound change. In Charles Jones (ed.) *Historical linguistics: Problems and perspectives*, 237–278. Harlow: Longman.

Ohala, John J. 2003. Phonetics and historical phonology. In Joseph & Janda (2003), 669–686.

Paul, Hermann. 1880. *Prinzipien der Sprachgeschichte*. Halle: Niemeyer. English translation published 1889, *Principles of the history of language*. New York: Macmillan.

Pedersen, Holger. 1931. *The discovery of language: Linguistic science in the nineteenth century*. Translated 1962 by John W. Spargo. Bloomington: Indiana University Press.

Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32. 693–703.

Phillips, Betty S. 1984. Word frequency and the actuation of sound change. *Language* 60. 320–342.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology 7*, 101–139. Berlin & New York: Mouton de Gruyter.

Price, Glanville. 1984. *The French language: Present and past*, 2nd edn. London: Grant & Cutler.

Raphael, Lawrence J. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. *Journal of the Acoustical Society of America* 51. 1296–1303.

Raphael, Lawrence J., M. F. Dorman, Frances Freeman & Charles Tobin. 1975. Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies. *Journal of Speech and Hearing Research* 18. 389–400.

Sampson, Rodney. 2010. *Vowel prosthesis in Romance: A diachronic study*. Oxford: Oxford University Press.

Schuchardt, Hugo. 1885. *Über die Lautgesetze: Gegen die Junggrammatiker*. Berlin: Oppenheim. Available (July 2010) at http://schuchardt.uni-graz.at/.

Smith, Caroline L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics* 25. 471–500.

Stockwell, Robert & Donka Minkova. 2001. *English words: History and structure*. Cambridge: Cambridge University Press.

Tekavčić, Pavao. 1972. *Grammatica storica dell'italiano*, vol. 1: *Fonematica*. Bologna: Il Mulino.

Torreira, Francisco. 2007. Pre- and postaspirated stops in Andalusian Spanish. In Pilar Prieto, Joan Mascaró & Maria-Josep Solé (eds.) *Segmental and prosodic issues in Romance phonology*, 67–82. Amsterdam & Philadelphia: John Benjamins.

Trask, R. L. 1996. *A dictionary of phonetics and phonology*. London: Routledge.

Vanelli, Laura. 1979. L'allungamento delle vocali in friulano. *Ce fastu?* 40. 66–76.

Vennemann, Theo. 1972. Rule inversion. *Lingua* 29. 209–242.

Weinrich, Harald. 1958. *Phonologische Studien zur romanischen Sprachgeschichte*. Münster: Aschendorff.

94 Lexical Phonology and the Lexical Syndrome

Ellen M. Kaisse
April McMahon

1 Introduction

The theory of Lexical Phonology and Morphology (LPM) deals with the place of phonology within the larger grammar. It has much to say about the relation between morphology and phonology, and also provides a model for the integration of phonology with the material provided by syntax and the phrasing derived from syntax. It is also a theory of phonological typology, assigning one set of characteristics to processes that apply only within words (the lexical phonology) and another, complementary set to processes that apply within or between words (the post-lexical phonology). The theory was first developed in the early 1980s by Kiparsky (1982) and his colleagues and students (especially Mohanan 1982, 1986) at MIT. It rapidly attracted a great deal of interest among phonologists because of the new tools it supplied for attacking recalcitrant problems, the set of intriguing questions it allowed a researcher to ask about any phenomenon, and the organic way it grew out of many of the major trends in phonology and morphology that had occupied linguists since the publication of Chomsky and Halle (1968). LPM was the basis for much of the synchronic and diachronic work, both descriptive and theoretical, that went on in phonology for a decade or more following its birth. Classical LPM was probably also the last model of phonology–morphology interaction to enjoy a wide consensus (Noyer 2004). LPM remains influential today, in its legacy of ways of thinking about phonology and in new instantiations that marry it with Optimality Theory (Orgun 1996; Kiparsky 2000; Rubach 2000; Bermúdez-Otero, forthcoming; among many others). As is often the case with valuable but complex and inevitably imperfect theories, LPM did not simply collapse under its own weight. Rather, interest largely moved elsewhere. Starting in the early 1990s, many phonologists turned away from rules and derivations to Optimality Theory (OT). Until recently, most versions of OT have involved only one evaluative step – potential output candidates are evaluated simultaneously for their satisfaction of ranked constraints. The result of the large-scale turn to OT was that most people stopped writing about and within LPM before consensus had been reached on what to do about the difficulties that had been encountered within it. It is certainly true that there was not much from Kiparsky's (1982) proposals that did not have some qualifications, exceptions, or caveats attached by the mid-1990s, but this is not, as we reconstruct it, the main reason one no longer hears so much about LPM. On the contrary, many phonologists continued to recognize advantages to the theory and to adopt pieces of its legacy.

We will not be able to give a full introduction to LPM in this chapter, though we will certainly review some of its major features. For more detailed background, the original source, Kiparsky (1982), is invaluable. Kaisse and Shaw (1985) and Kenstowicz (1994: ch. 5) also provide introductions to the original model. Kaisse and Hargus (1993) summarize developments which the theory underwent in the decade after its first appearance, including those contained in the volume to which it is an introduction, Hargus and Kaisse (1993).

Lexical Phonology and Morphology's contributions center around two major areas: the segregation of characteristics typical of lexical *vs.* post-lexical rules, sometimes called the *lexical syndrome*, and the interaction of morphology and phonology, including affix ordering, stratal organization, and cyclicity. We will begin this chapter with a review of the characteristics that together made up the lexical syndrome and a discussion of how each of these has been modified or discarded in the light of empirical challenges. The succeeding section deals with the interaction of morphology and phonology in LPM, the difficulties which much of the original proposal has encountered, and alternatives that have been offered. The final section discusses the lasting influence of LPM and some of the directions it has taken in more recent years.

2 The lexical syndrome and its complications

Classical Lexical Phonology (Kiparsky 1982, 1983) proposes that a cluster of properties characterizes lexical rules, while a different, complementary cluster is found in post-lexical rules. The properties that have been used most often as diagnostics and which have attracted attention in subsequent literature are the following:

(1) a. Lexical rules are word-bounded. They find their target and structural description within the same word and do not apply between words. This is the easiest diagnostic to apply, and is often used as the starting point from which other characteristics are then predicted.

b. Lexical rules are cyclic. They have access to the internal morphological structure of a word and may apply more than once, first to the most deeply nested constituents, then again as more morphological material is brought into consideration (see also CHAPTER 85: CYCLICITY).

c. Lexical rules are subject to non-derived environment blocking. They apply to strings derived by the concatenation of morphemes or to strings derived by previous applications of phonological rules, but they do not apply within a single, underived morpheme (see also CHAPTER 88: DERIVED ENVIRONMENT EFFECTS).

d. Lexical rules are structure preserving. They are restricted to producing segments that are found in the underlying inventory of the language. Allophonic variants are not produced by lexical rules. However, lexical rules may assign structure not present underlyingly, in the form of tonal associations, metrical feet, and syllable structure (see also CHAPTER 76: STRUCTURE PRESERVATION: THE RESILIENCE OF DISTINCTIVE INFORMATION).

e. Lexical rules apply categorically, rather than gradiently. As a corollary of structure preservation, they do not produce partially assimilated sounds, and they do not apply more robustly to segments nearer to the trigger of the change (see also CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY).
f. Lexical rules may have exceptions (see also CHAPTER 106: EXCEPTIONALITY).
g. Lexical rules are subject to lexical diffusion (Kiparsky 1988). In diachronic sound changes, they apply to more and more individual lexical items over time (see also CHAPTER 93: SOUND CHANGE).

Post-lexical rules have radically opposed characteristics:

(2) a. They apply between as well as within words.
b. They apply once, non-cyclically, across the board, independent of morphological or syntactic structure.
c. As a corollary to (b), they are not subject to non-derived environment blocking.
d. They need not be structure preserving. They may produce allophones not present in the underlying segment inventory.
e. They may apply gradiently, producing partially assimilated segments or applying more robustly to segments temporally closer to the trigger.
f. They do not have exceptions.
g. In diachronic change, they apply in the fashion predicted by Neogrammarian views of sound change. Every string that meets their description undergoes the change, with no lexical conditioning. The changes can be small, allophonic, and non-neutralizing.

These characteristics were posited within an organization of grammar approximately like that in Figure 94.1. The stem level is so called because it contains affixes that can attach to bound bases (stems) that cannot appear unaffixed, while the word level deals in the formation of words from other words.

To illustrate how these segregated lists of characteristics play out, consider the simplified example of Trochaic Shortening *vs.* North American Flapping employed in Kaisse and Shaw (1985). (For a more nuanced and current discussion of the

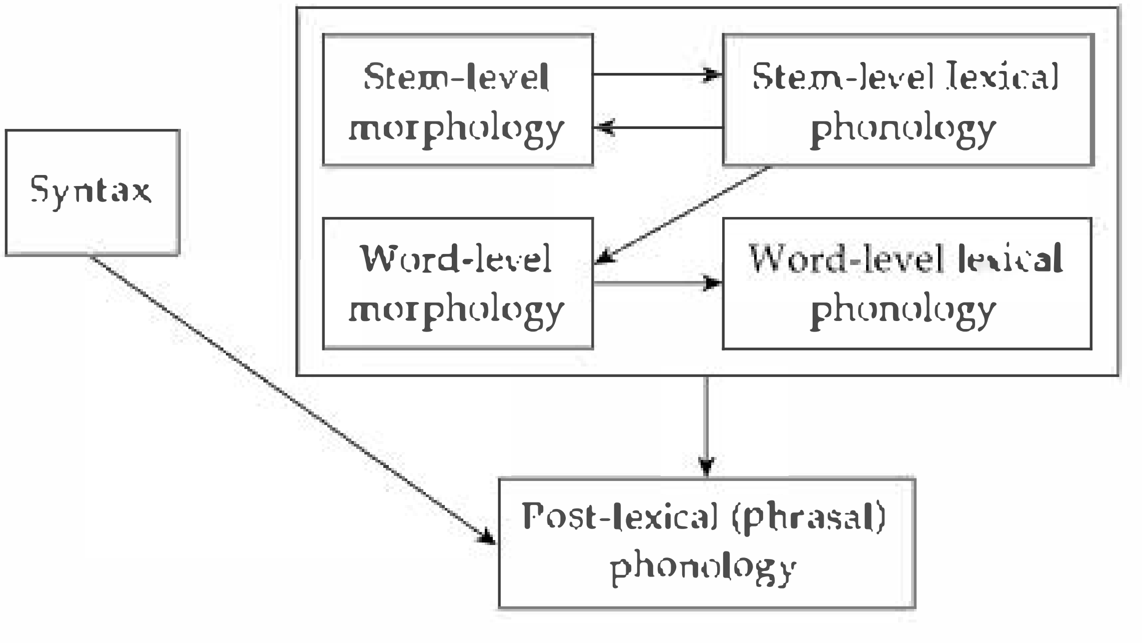


Figure 94.1 The organization of the grammar in LPM

theoretical implications and analysis of Flapping, see CHAPTER 113: FLAPPING IN AMERICAN ENGLISH and Bermúdez-Otero 2007b.) Flapping converts an underlying coronal stop to [ɾ] when it is intervocalic (roughly speaking) and ambisyllabic (Kahn 1976.) The rule can make up its intervocalic environment with vowels from the same word, or the second vowel can come from a following vowel-initial word. The word-internal environment can come from a single morpheme (*ci*[ɾ]*y*), from the combination of a stem plus stem-level suffix (*fa*[ɾ]+*al*), or from a word plus a word-level suffix (*chocola*[ɾ]-*y*, *edi*[ɾ]*ed*). The between-word environment is insensitive to the part of speech of either word and to their syntactic relationship, though phrasing into independent prosodic groups, which interferes with ambisyllabicity, does block Flapping (compare *It's the close*[ɾ] *Annette cares about* with *It's the close*[t], *Annette insisted*). The fact that Flapping can apply between words is the prime diagnostic for its being post-lexical. We have just seen its insensitivity to morphological or syntactic structure. It is also non-cyclic. A word like *atomic* does not contain a flap even though the word from which it is derived, *atom*, does. Continuing down the list, words like *city* and *butter*, where the coronal stop is frozen in an intervocalic environment within a single morpheme, show that Flapping is not subject to non-derived environment blocking. The segment [ɾ] is not part of the underlying inventory of English; the coronal tap or flap is not well distributed, but occurs only where the environment of Flapping is found. We are not aware of studies that show Flapping to be gradient, though this is certainly true of the distribution of aspirated and plain stops in English, to which Flapping is related – stops which are initial in large prosodic categories are more aspirated than those which are initial in internal syllables. However, Flapping is extremely common but nonetheless variable, in that the likelihood of its application is partly dependent on register and rate of speaking (see Patterson and Connine 2001, who find that American English Flapping is almost categorical in its application). And while Flapping is not completely obligatory, it does not have lexical exceptions – any word of English can undergo it if the phonological and stylistic prerequisites are present.

Compare the behavior of Trochaic Shortening, the rule responsible for alternations like that in *sane* ~ *sanity*, *meter* ~ *metric*(*al*), and *nation* ~ *national*. The rule shortens a vowel followed by an unstressed syllable – i.e. it shortens the stressed syllable of a trochaic foot. (The rule is often called Trisyllabic Shortening, because English has final extrametricality, placing the stressed syllable of many trochees three syllables from the end of the word; see CHAPTER 43: EXTRAMETRICALITY AND NON-FINALITY; CHAPTER 44: THE IAMBIC–TROCHAIC LAW.) Trochaic Shortening cannot make up its structural description out of adjacent words – the vowel of *sane* in the phrase *sane as a judge* must be long even though the prosodic phrasing can place its vowel before a string of unstressed syllables. Going down our list, we see that Trochaic Shortening is sensitive to the type of affix added: word-level affixes do not trigger shortening, as shown by *nationhood* or *metering*. The rule is also cyclic, since once the vowel in *metric* is shortened, it persists as short in derived words even when the stress falls on the vowel immediately after, as in *metrician*. As for non-derived environment blocking, Kiparsky (1982) makes the point that English contains many long vowels in the Trochaic Shortening environment, but almost all are found within a single morpheme (synchronically speaking), such as *nightingale* or *Rotenberg*. Thus Trochaic Shortening is blocked in non-derived environments. Trochaic Shortening is structure preserving, since all the shortened vowels are phonemes of English. It is non-gradient. And it has exceptions, such as the long vowels in *notional* and *obesity*.

Optimality-theoretic work has tried to deal with many of the phenomena of the lexical syndrome, but for the most part in piecemeal fashion, rather than as following from a particular organization of the grammar. One finds many publications framed in OT which aim to explain apparent cyclicity effects (Kenstowicz 1995; Benua 1997; Kiparsky 2000; Pater 2000; and many others); work by Łubowicz (2002) and Wolf (2008) attempts to recapture non-derived environment effects within OT; Steriade (1999), Burzio (2002), Raffelsiefen (2004), and Bermúdez-Otero and McMahon (2006), among others, have advanced explanations for why word-level morphology seems to have so little effect on phonology. And this piecemeal treatment may exist with good reason. We will shortly look into how the features of the lexical syndrome have fared since they were first proposed and will discover that the connections between strata and features are not nearly as tight as the original LPM theory proposed.

But before we get to the complications, let us first consider how the philosophical outlook imposed by the LPM model influenced the way in which phonologists worked. There are some key, if non-architectural, features that strongly affected the research strategies of those who considered themselves practitioners and which continue to affect the expectations we have of a phonological description to this day:

(a) LPM encourages interest in processes which are not exceptionless, transparent, or expressible in maximally general and elegant statements. Interest was directed equally to processes with exceptions and oddities either in the environment or in the list of items applied to, and researchers tried to understand the difference between these and the maximally general and transparent type of process. The concept of opacity, long of interest to Kiparsky, gained a new dimension, since lexical rules are often opaque, while post-lexical rules typically are not. And since morphology can make a phonological rule opaque by being invisible to that rule at the stratum in which it applies, or by allowing a phonological rule to apply before a new morphological process destroys the environment, understanding the interaction of phonology with morphology can lend insight into some of the sources of opacity (CHAPTER 103: PHONOLOGICAL SENSITIVITY TO MORPHOLOGICAL STRUCTURE).

(b) LPM encourages investigations of interactions between phonology and other aspects of the grammar, notably phonetics and syntax at the "bottom end" in the post-lexical component, and morphology and the lexicon at the "top end." The observation that post-lexical rules can be gradient was influential in the great expansion of our understanding of the relation between phonetics and phonology in the last two decades and in the realization that phonetics is a language-specific, rule-governed domain closely related to the phonology of the language.

(c) LPM encourages investigations of diachronic change. LPM provided a new architecture which made sense of the way sound changes progressed over time and thus aided in resolving the so-called "Neogrammarian controversy" (Kiparsky 1988). If there were lexical and post-lexical rules, and more than one level in the lexicon, and each was characterized by different properties, rules could apply differently throughout their history. A rule could start post-lexically and be very close to the phonetics, transparent and phonetically conditioned, applying everywhere that its context arose. Later it could come to apply lexically, be influenced by the morphology, gain lexical exceptions, and in time become more and more opaque. Eventually, it could become a matter for learning rather than online application, and color or alter underlying representations, while no longer existing as a rule per se. Given that Trochaic Shortening occurs in almost all dialects of English while Flapping is geographically restricted, it seems likely that Trochaic Shortening is an older rule and has moved deep into the lexicon – indeed perhaps it is no longer an online rule at all – while Flapping is still showing the post-lexical characteristics of its initial entry into the language.

We shall return in more detail to each of these points in §4. One moral to draw is that, even though the lexical syndrome is not as strong as it first appeared, the sorts of questions that the segregated lists of traits made one ask about a phonological process were valuable research tools. It is important to know whether a process has exceptions, applies only in derived environments, shows cyclic characteristics, refers to morphological structure, is bounded with the word, creates novel segments, and so forth. And it may still be the case that one can make a better-than-chance prediction of how these traits are going to be distributed on the basis of the theory of LPM. Our suspicion is that the lexical syndrome is like so many other things once thought to be universal but now realized to be violable – it is not an utterly exceptionless principle, but an insight into the unmarked state of affairs in phonologies. It is worth remembering that it took over a decade for research within the LPM model to come up with the list of exceptional phenomena we present below, and that some are single instances of problematic cases in the face of dozens of attested cases of the predicted behavior. These problematic cases were not easy to come by because, in fact, the predictions of the model were much more often right than not.

Where do we stand with respect to the health of the lexical syndrome? In the first two columns of Table 94.1, we repeat a list of the characteristics that make up the lexical syndrome (including for completeness a few of the less influential ones we did not mention above), adapted from Kaisse and Hargus (1993: 16–17). Where the characteristic was not part of the early and generally accepted LPM canon (Kiparsky 1982, 1985; Mohanan 1982, 1986), we give a source. The last two columns give a quick summary and citation of works that raised problems with each putative characteristic. Most of these difficulties are summarized in Kaisse and Hargus (1993), and the details can be found in papers from their volume or other sources in the table, so we will only outline their content here. What we see in general is a gradual and traceable retreat from some of the formal features of LPM. For instance, structure preservation is at best a feature of stem-level rules, more likely not predictable from stratum. By the time we reach OT, the question of structure preservation becomes unaskable, since Richness of the Base precludes any characterization of the underlying segment inventory of a language. Cyclicity also stops being a feature of all lexical rules, and becomes a feature only of some, again perhaps those of level I. Even cyclicity of the stem level becomes a non-issue in the most recent work of Bermúdez-Otero and McMahon (2006) or Collie (2008), being replaced with a type of lexical listing called "fake cyclicity," which we discuss below. And, of course, rules become non-standard currency as we shift toward OT and constraints.

Table 94.1 Characteristics of the lexical syndrome

	Lexical	*Post-lexical*	*Problems*	*Reference*
a.	Word-bounded	Not word-bounded	Apparent post-lexical rules with exceptions may be precompiled in lexicon; lexical rules may access output of syntax	Hayes (1990); Odden (1993)
b.	Access only to word-internal structure assigned at same level	Access to phrase structure only	Some post-lexical rules like English /l/-velarization need access to word-internal morphology	Sproat (1993); Bermúdez-Otero (2007b)
c.	Cyclic	Apply once	Stem level only; word level not cyclic; no true cycle in stem level, rather lexical listing and faithfulness	Borowsky (1993); Bermúdez-Otero & McMahon (2006); Collie (2008)
d.	Apply in derived environments	Apply across the board	Non-Derived Environment Blocking more likely associated with obligatory neutralization rules	Kiparsky (1993); Wolf (2008)
e.	Structure-preserving	Not necessarily structure-preserving	Word-level rules not structure-preserving; P1 rules often structure-preserving; post-lexical rules may differ w.r.t. structure-preservation. Sound changes may introduce non-structure-preserving segments at stem level	Harris (1989); Kaisse (1990); Borowsky (1993); Hyman (1993); Bermúdez-Otero & McMahon (2006)
f.	May have exceptions	Automatic	P1 rules like the English Rhythm Rule have exceptions	Kaisse (1990)
g.	Not transferred to a second language (Rubach 1984)	Transferable to L2		
h.	Output subject to lexical diffusion (Kiparsky 1988)	Change via Neogrammarian sound change		Calabrese (2009)
i.	Apply categorically	May have gradient outputs	Raises issues relevant to the link between gradience and phonetic implementation	

Of these characteristics, some were not really counterexemplified or rejected by the authors cited – rather they were refined or altered in interesting ways. For instance, Borowsky (1993) (row (c) in Table 94.1) did not reject the idea that lexical rules are cyclic. Rather, she claimed that cyclicity should be limited to the stem level, while both the word and the phrasal level are non-cyclic. She also followed Kiparsky (1985) in suggesting that structure preservation turns off at the stem level, so that word-level rules can create allophones (CHAPTER 11: THE PHONEME). And finally, in the most radical of her proposals, she argued that many word-level rules show a syndrome all their own. They are non-cyclic, do not obey the strict cycle condition (and thus can apply within a single morpheme), apply between stems and level I suffixes, but are unable to see word-level affixes. Finally, they do not apply between words, so that they cannot be phrase level. One of her more familiar examples uses Harris's (1989) Belfast Dentalization, a process where coronal consonants are realized as dentals before [r] or [ər]. The process is not structure preserving, as it produces dentals which alternate in variable fashion with alveolars and are not separate phonemes. These dentals, however, are only produced morpheme-internally (*butter*) or before [(ə)r] introduced by a stem-level suffix (*elementary*). The rule does not apply between an alveolar and an [r] in a word-level suffix (*fitter*), nor in the next word (*foot rest*), the latter fact indicating that Dentalization cannot be post-lexical. For Borowsky, Belfast Dentalization is a word-level rule that cannot see word-level morphology. It shows its word-level nature by applying in non-derived environments (*butter*) and by producing the novel dental segments. Because of work like Kiparsky and Borowsky cited here, Ricardo Bermúdez-Otero (personal communication) suggests that it is more sensible to talk about a stem syndrome than a lexical syndrome.

Such examples point to the difficulties the theory begins to encounter once the boundaries between lexical and post-lexical characteristics begin to break down. Is the non-structure-preserving nature of Dentalization enough to categorize it as a word-level rule (Borowsky), or is its inability to apply before word-level affixes enough to categorize it as stem level, requiring the rejection of structure preservation as criterial of the stem level? The latter is the tack taken by Bermúdez-Otero and McMahon (2006), who go on to argue that the whole notion of structure preservation was never well defined and that we are well rid of it within any theory that adopts Richness of the Base.

Another example of refinement rather than rejection is Kaisse (1990). Kaisse did not deny that the original class of post-lexical rules referred to by Kiparsky and Mohanan is exceptionless, i.e. that there are no listed lexical items that fail to undergo them. Rather, it drew attention to a different sort of syntax-sensitive post-lexical rule which was, in its own way, cyclic, exception-bearing, and subject to the lexical syndrome. One of Kaisse's examples is the Rhythm Rule of English, the process which repels clashing pitch accents in phrases like *ˌthirteen ˈpeople*. The rule is clearly post-lexical, as the stresses in the independent words *thirˈteen* and *ˈpeople* show. Yet it has many individual lexical exceptions in the forms of words that will not retract their accents in clash: *inˈtense*, *oˈbese*, and many other words do not participate in the Rhythm Rule. Kaisse attempted to divide post-lexical rules into two strata, the first of which (P1) contained more grammaticalized rules like the Rhythm Rule, while the second (P2) was the home of the classical post-lexical rule. Hayes (1990) was an alternative proposal for dealing with non-word-bounded rules that exhibited some of the characteristics of lexical rules, again with the intention of supporting the general notion of a lexical rule syndrome.

On the other hand, Kiparsky's (1993) attack on the Strict Cycle Condition (which he terms "Non-Derived Environment Blocking"; NDEB) as a characteristic of only lexical rules is quite devastating, and leads us to conclude that we still have not really understood this phenomenon, which has vexed phonologists since Kiparsky discovered it in the early 1970s. Perhaps the best-known example of NDEB, first introduced by Kiparsky (1973), is assibilation in Finnish. The rule turns /t/ to [s] before [i] either across morpheme boundaries (/tilat+i/ → [tilasi] 'ordered') or when the [i] is derived by another phonological rule, such as raising of final /e/ to [i] /vete/ → veti → [vesi] 'water (NOM SG)'. It is blocked within a morpheme, such as [tila] 'order'. This example is already problematical, Kiparsky (1993) argues, since assibilation is fed by the word-level rule of raising and thus must itself be word level. NDEB had otherwise been thought to affect only cyclic processes, and cyclic processes are supposedly restricted to the stem level. Even worse, Kiparsky shows that the Vedic Sanskrit rule that retroflexes /s/ after /r/, /u/, /k/, and /i/ shows strong NDEB effects, but applies between words. Kiparsky concludes that NDEB is not a property of any particular level – it does not lend support to the idea of a lexical syndrome. Most recently, Wolf (2008) has argued that NDEB effects are best modeled in an optimality-theoretic model with candidate chains that permits the interleaving of phonological and morphological operations.

Hyman's (1993) discovery of two tone rules in Dagbani, which are both post-lexical and have the same domain of application, yet which differ in whether they are structure preserving – in this case, in whether they are or are not permitted to create a contour tone – seems to us to be a clear demonstration that structure preservation cannot be absolutely predicted on the basis of the component in which a rule applies. But this is one of those cases which were so difficult to find; it took a decade of research within the model to uncover a counterexample of this type. Hyman's example could simply be taken as evidence that structure preservation, though generally predictable on the basis of a process's domain, may sometimes need to be stipulated. Note also that the claim of LPM was that lexical (in particular, stem-level) rules could not create novel segments and structures. The claim was never that post-lexical rules could *not*, accidentally, happen to create structures that obeyed structure preservation. Hyman's example is therefore more a counterexample to Kaisse's (1990) claim that structure preservation is to be found in syntax-sensitive (P1) rules and not in across-the-board (P2) rules.

Odden's (1993) discovery of a demonstrably lexical glide formation rule in Matumbi that must have access to the syntactic structure and phonological characteristics of surrounding words is highly problematical for the general organization of phonology proposed within LPM. But the case is extremely unusual and complex; again, it is not the sort of thing phonologists run into more than once or twice in a career.

Cyclicity has perhaps received the most attention from later theoreticians and is the subject of a full chapter in this *Companion* (CHAPTER 85: CYCLICITY). Within monostratal Optimality Theory, cyclic application has been rejected in favor of various sorts of output–output faithfulness constraints, paradigm uniformity constraints, or constraints favoring "uniform exponence," where the allomorphs of a morpheme differ as little as possible. And, as mentioned earlier, even within contemporary stratal theories, some authors deny even that level I shows true cyclic derivations. Rather, the concept of fake cyclicity (see Collie 2008 for a review) ensures that all stem-level outputs are listed in long-term memory rather than derived online. Stem-level phonological generalizations simply express typical but not exceptionless static relations between a base and its derivative. Collie argues that this is the only way to handle the fact that the preservation of relative prominence under embedding in English is gradient and less likely to occur the more frequent the derivative is in relation to its base. So for instance, the prominence-preserving position of pre-tonic stress in *E,liza'bethan* is a consequence of faithfulness to the listed form *E'lizabeth*. The variable lack of faithfulness between the possible pronunciation *,antici'pation* (with secondary stress on the first rather than the second syllable) and the related form *an'ticipate* results from the possibility that *anticipation* is accessed whole rather than decomposed into component morphemes and derived via cyclic application of stress. We will again encounter this relationship between the frequency of a derived word and the likelihood that it is accessed whole rather than having its affixes stripped off when we discuss affix ordering in the next section.

3 The interaction of morphology and phonology in LPM

We think it is safe to say that the morphological side of the LPM model was always weaker than the phonological side. One of the most attractive, if doomed, features of the LPM model was an attempt to unify apparent generalizations about the phonology of cohering (also known as stem level, level I, or +-boundary) affixes, their morphological behavior (especially their triggering of cyclic rule application), and their linear order. Such affixes were opposed to the non-cohering (word level, level II, or #-boundary) affixes. Unfortunately, the strong position adopted by LPM has turned out to be readily falsifiable. LPM adopts and elaborates the level ordering (or "affix ordering") hypothesis of Siegel (1979). This hypothesis, first made on the basis of English, claims that cohering affixes will occur close to the root, while non-cohering affixes will always occur outside the cohering ones. In other words, level I affixes cannot attach to a word to which a level II affix has already been attached. Thus, according to Siegel, Kiparsky (1982), and other LPM work, the non-existence of words like **happi#ness+al* or **sing#er+ous* comes from the fact that they contain a word formed with a non-cohering affix (*#ness*, *#er*) to which a cohering affix (*+al*, *+ous*) has been affixed. Affix ordering permits words like *person+al+ity*, since both *+al* and *+ity* are level I suffixes. The word *danger+ous#ness* is also fine, since it contains a level I suffix followed by a level II suffix. And some words with strings of level II suffixes are also fine: *seam#less#ly*, *seam#less#ness*. The correlation with phonology is that all the affixes starting from the stem outward to the first non-cohering affix should form part of the visible input to level I lexical phonological rules, while all of the affixes starting from the first non-cohering affix outward will not be the trigger or target of any such rule, and will only undergo post-lexical rules.

Table 94.2 shows the workings of the lexical module as it was devised for English, modified slightly from the diagram found in Kiparsky (1982: 133). The double-headed arrow symbolizes the potential passing of forms back and forth from the morphology to the phonology as more morphemes are added.

Table 94.2 The lexical module for English

	Morphology		*Phonology*
Level I (stem level)	+-boundary inflection and derivation, zero derivation of nouns from verbs	⇔	stress rules, trochaic shortening, velar softening, sonorant syllabification, etc.
Level II (word level)	#-boundary derivation and compounding, zero derivation of verbs from nouns, most inflection		compound stress

In this model, morphological operations occur one affix at a time. Each time an affix is added at level I, the form is passed to the level I phonology, which applies to the string as it is currently concatenated. The form is then passed back to the morphology at that same level, over to the phonology again, and so on until all level I affixes for that word are added. Cyclic application (Chomsky and Halle 1968) results from this interleaving of phonological and morphological operations at level I. Then the form passes on to level II. In English, it appears that all the cyclic phonological rules apply at level I. To illustrate the segregation of morphological operations into levels and the concomitant applicability of certain phonological rules only to strings created at that level, consider a point about the zero derivation of nouns from verbs *vs.* that of verbs from nouns made in Kiparsky (1982). When nouns are derived from verbs with final stress, they often preserve final stress as a secondary prominence but add a penultimate, primary stress, as befits a noun. Thus we find noun/verb pairs such as *'con,vict* from *con'vict*, and *'tor,ment* from *,tor'ment*. In a derivational theory such as LPM, one would say that nouns are zero-derived from verbs at level I, and are thus subject to another cycle of the stress rules after that affixation occurs. However, Kiparsky argues, the zero derivation of verbs from nouns takes place at level II. Since the stress rules of English do not apply at level II, the stress does not change in denominal verbs like *to 'pattern* (**to ,pat'tern*), even though primary verbs ending in two consonants receive final stress. Thus we could say that deverbal -∅ is a level I suffix while denominal -∅ is level II.[1]

But the affix ordering hypothesis makes incorrect predictions, certainly for English. Problems with the level ordering of affixes had been recognized as early as Aronoff (1976). Best known from the beginning were cases where affix ordering proves to be too strong a theory, ruling out combinations that actually occur, such as *#ment+al* (*governmental*) and *#iz+ation* (*neutralization*). The suffixes *-ment* and *-ize* are stress-neutral, and hence non-cohering, as witnessed by forms like *'government*, with stress in the same position as its source verb and no stress on its heavy penult, and *'marginalize*, with stress four syllables from the end of the word. But *-al* and *-ation* are stress affecting, and hence cohering. Bracketing paradoxes like *un#grammatical+ity*, *re#organiz+ation*, and many others have demonstrated that the affix ordering hypothesis undergenerates. For instance,

[1] Kiparsky (1983) modified this analysis on the basis of a wider dataset, making it more complex in the process; we use the possibly oversimplified version for illustration.

in *ungrammaticality*, level I *-ity* should attach before level II *un-*, and yet *un-* attaches to adjectives (like *grammatical*) and not to nouns.

Equally telling is Fabb's (1988) demonstration that the affix ordering hypothesis is too weak and overgenerates. Fabb notes that affix ordering restrictions account for only a small percentage of the sequences of affixes that don't occur in English. He lists 43 common affixes in English. If there were no stratal affix ordering, we might expect around 600 grammatical combinations of these affixes by simply making sure that affixes that select for a particular part of speech are combined only with that part of speech, and that the particular stress requirements of the affixes like deverbal *-al* are met. With affix ordering, we can pare the expected number down to 459 combinations. But English words actually contain only about 50 pairs of suffixes. The main reason is that 28 of the common suffixes never combine with another suffix. Six suffixes combine with only one suffix; for instance *-ic* only attaches to unsuffixed stems or words (*comic*, *metallic*) or to the suffix *-ist* (*modernistic*). Six other suffixes are semi-productive: noun-selecting *-al*, for instance, combines with three cohering and non-cohering affixes *-ion*, *-ment*, and *-or*. Only *-able*, deverbal *-er*, and *-ness* show no selectional restrictions beyond part of speech.

Plag (1996, 1999) shows that even Fabb's restrictions are inadequate. First, Fabb's treatment gives no obvious reasons why suffixes should fall into the combinatorial classes they do. And critically, there are more restrictions on the combinations of English suffixes than are covered by Fabb's observations. These restrictions extend to non-cohering suffixes as well. Plag demonstrates that semantic, phonological, morphological, and syntactic selectional restrictions must be stated on individual suffixes' combinations with stems and other affixes. Once these are stated, both level ordering and Fabb-style lists of combinatorial restrictions become superfluous. Consider for instance the verb-forming suffix *-en*. It can only attach to monosyllabic bases, as in *blacken* vs. **maroonen*. It immediately follows that *-en* will fall into the class of affixes that cannot attach to another affix, since an affixed base would already have at least two syllables. Its behavior is thus explained without additional need for level order or lists of what suffixes it can combine with. Additional insights can chip away further at the otherwise unexplained list of suffixes which cannot attach outside of any suffix. For instance, that list contains five deverbal noun-forming suffixes: *-age*, *-al*, *-ance*, *-ment*, and *-y*. If these were to attach to already suffixed words, they would have to attach to words ending in one of the verb-forming suffixes of English, namely *-ate*, *-ify*, *-ize*, and the already accounted for *-en*. But affixation onto these first three suffixes is "base-driven" in Plag's terminology: *-ify* requires the suffix-particular allomorph *-ication* (as in *classification*), *-ize* requires *-ation*, and *-ate* requires *-ion*. These very specific base-driven requirements again make superfluous any statements about the affixes that cannot attach to other affixes, let alone the stratal membership of those affixes.

Hay (2002) and Hay and Plag (2004) take one further step in the advancement of our understanding of what controls affix ordering. They criticize Plag's earlier work because of a certain residual arbitrariness. Are there not any underlying principles behind the individual suffix- and base-driven restrictions that he notes? For this they utilize Hay's Complexity-based Ordering, which states that affixes that are easily parsed out of derived words must occur outside affixes that are not easily parsed. The likelihood of a derived word being accessed as one unit or instead parsed into base and affix depends on the relative frequency of the whole word to its parts. For instance, *government* is more likely to be left unparsed into component parts than is *discernment*, because *government* is more frequent than *govern*, while *discern* is more frequent than *discernment*. Therefore there is nothing wrong with a formation like *governmental*, while *discernmental* is ill formed. This is because *-ment* is not easily parsed out of *government* and so can be followed by *-al*. But *-ment* is easily parsed out of *discernment* so it cannot be followed by *-al*.

The inevitable conclusion from the combined work of Aronoff, Fabb, Plag, and Hay is that the affix ordering hypothesis must be rejected, at least for English,[2] and with it, that part of LPM that rests upon it.

A strong trend in the mid-1980s to early 1990s was the attempt to replace phonology's direct access to morphological structure with access to prosodic structure based on morphology but not identical to it. Inkelas (1993) and Booij and Lieber (1993) argue that it is possible to maintain a version of LPM, substituting prosodic groupings that need not be isomorphic with morphology. For instance, in Korean (Kang 1992) and Polish (Booij and Rubach 1984) prefixes are independent morphological words. In Polish, no lexical phonological rules apply between prefix and stem, there is always a syllable boundary in that position, and the rules of Lowering and Vowel Deletion cannot be made to operate properly unless prefixes are treated as words. Truckenbrodt's work on the phonology–syntax interface (2002, 2007) is a logical outgrowth of the Inkelas (1993) and Booij and Lieber (1993) approaches, and shows how the variable phrasing of words into post-lexical prosodic groupings can be achieved using the tools of Optimality Theory. As with morphology, a strict translation of syntactic structure into phonological structure does not seem to be adequate.

In contrast to their rejection of the affix ordering hypothesis, both Fabb (1988) and Aronoff and Sridhar (1987) continued to believe in another tenet of level ordering, namely that the word boundary and morpheme boundary affixes could be sorted into two coherent groups on the basis of their phonological behavior, just as had been claimed in Chomsky and Halle (1968) and in LPM. The +-boundary suffixes were available to rules of stress assignment – both those assigning stress and those choosing which of these would be the primary stress; they were available for syllabification so that the vowel-initial ones could bleed Sonorant Syllabification (*hinder*, *hinder#ing*, but *hindrance*); and their content was visible to Trochaic Shortening. We are not aware of many discussions that place this claim under the scrutiny that the affix ordering hypothesis has received. And indeed, a recent instrumental study by Sugahara and Turk (2009) continues to assume this segregation, working upon prosodic structure rather than directly on morphological structure. Sugahara and Turk find support for a distinction among affix types. Their study investigated whether differences in word-internal morphology is reflected in durational differences of monomorphemic stems, stems before level I suffixes and stems before level II suffixes in Scottish English. They hypothesize that the base to which a level II suffix is attached must be a prosodic word and thus will be temporally longer than a level I base. In slow speech, they found stem-final rhymes followed by level II suffixes were longer

[2] We do not know if arguments like Fabb's go through for languages with highly productive agglutinative suffixation, such as Turkish.

than corresponding strings in monomorphemic words, and even longer than stem-final rhymes followed by level I suffixes. Their results are consistent with a model where stems preceding level II suffixes are mapped onto prosodic words. Of course, given the exigencies of a rigorous experimental setting, Sugahara and Turk were only able to test a small subset of affixes. To represent level II, they used only the inflectional suffixes -s, past tense *-ed*, and *-ing*. Level I was represented by *-ance*, *-ence*, *-ent*, and adjective-forming *-al*. Thus their results will need to be extended to many more affixes before it can be concluded that there are coherent sets of level I and of level II affixes in English.

The question is how far one should go between holding out for two well-segregated groups of affixes, and losing the generalizations and possible predictions that follow from some such grouping (albeit with exceptions). Raffelsiefen (2004), for instance, allows each affix to invoke its own constraint ranking, so that every affix is in effect its own class. Note, however, that this work does not necessarily cast doubt on the lexical syndrome, because it is not concerned with operations within the syntax, only within the morphology–phonology nexus.

4 The lasting legacy of lexicalism

In the first section of this chapter, we suggested that LPM had not been entirely disproved or abandoned, but on the contrary that many of its earlier adherents continue to recognize its advantages and achievements. However, rather than remaining as part of a single, self-contained theoretical approach, these positive aspects of LPM have typically become integrated into newer models, or have led to innovative ways of addressing phonological problems. This final section addresses the legacy of LPM in more detail; in particular, we have chosen to focus on its contributions to the understanding of phonological change, opacity, and the phonetics–phonology interface. We will also illustrate the operation of a version of Optimality Theory that incorporates levels, showing how it draws on aspects of LPM to handle some forms of opacity which otherwise create difficulties for OT.

4.1 Diachrony

While standard generative phonology did invoke an interaction between sound changes and phonological rules, this was not especially well motivated, and was difficult to integrate with the architecture of the model. While there was a mechanism for introducing new rules into the grammar through "rule addition" (King 1969), rules could only be added at the bottom of the list of ordered rules, and there was no clear way of allowing those then to percolate up into the grammar and hence alter the underlying representation. In turn, this led to the notorious difficulty that the order of rule applications tended to reflect the order of operation of sound changes, and hence to recapitulate history; this was one of the more powerful factors influencing the abstractness of underlying representations, or at least enabling them to be particularly abstract, and hence not a reason for recommending the model.

The stratal architecture of LPM and the properties of the lexical syndrome solve both these problems. For one thing (as argued by McMahon 2000), it is simply not possible both to adhere to the tenets of LPM and to construct maximally abstract representations and strings of rules recapitulating history. Instead, we must accept that the form of a sound change and of the phonological rule that derives from it may be radically different. McMahon (2000) suggests that the historical Great Vowel Shift, for instance, has been restructured into two modern English phonological rules, one for lax vowels and one for tense ones, both fed by a set of independently motivated tensing and laxing rules, and hence applying in derived environments only. Of course, it is an open question whether or not one might wish to maintain the existence of a rule or process or constraint to generate alternations which might equally well be lexically stored; the point is that if we are to do so, we cannot maintain that underlying forms are unchanged since Chaucer, and that the shape of a sound change precisely dictates the form of its derived and productive phonological rule.

Moreover, once LPM allowed phonologists to depart from the assumptions of identity between sound changes and phonological rules, and of the essentially static nature of underlying representations, there was a welcome freedom to explore the ways in which sound changes could in fact enter the grammar and continue to develop through time. Labov (1981), while strictly speaking pre-dating LPM, introduces the idea of two sorts of phonological rules, mapping on to two types of sound changes, which respectively follow Neogrammarian principles and spread gradually through lexical diffusion. Kiparsky (1988) and Harris (1989) develop this theme, making explicit the relationship between Neogrammarian changes and post-lexical phonological rules, and between lexically diffusing changes and lexical rules. However, Harris also develops the argument further, arguing that post-lexical rules may become lexical over time. Harris (1989) focuses on /æ/-tensing, which in a range of varieties of English gives a lax vowel in *tap, bath, manner, ladder*, but a tense one, albeit variably realized, in *pass, path, man, manning, man hours*. However, while /æ/-tensing applies before anterior nasals and anterior voiceless fricatives in Philadelphia, it also applies before voiced stops in New York, and additionally before /l/ and voiced fricatives in Belfast. Harris traces this phonological process to an earlier, automatic, phonetic change, operating in a hierarchy of environments which have been phonologized differently in different dialects. It is also clearly now a lexical rule, being lexically selective and sensitive to morphological information. This analysis, supported for example by McMahon's (2000) parallel account of the development of the Scottish Vowel Length Rule through phonologization of the familiar low-level English lengthening process before voiced consonants, establishes that post-lexical rules may begin to diffuse, and may then progressively acquire more properties of the lexical syndrome. In due course, as rules percolate up through the lexicon from level II to level I and become increasingly idiosyncratic, they may cease to be productive altogether, instead triggering changes in the underlying representations, which themselves become more dynamic and changeable, and hence also less abstract.

There are two important consequences of the dynamic architecture of LPM, and crucially the lexical syndrome. First, it sharpens the focus on different dialects, which become important repositories of data on intermediate stages in the cycle from sound change to post-lexical rule to lexical rule to lexical restructuring. Furthermore, it legitimates a new interest in sound change and in how changes become rules, which has had a wider impact on linguistics. In theoretical terms, this development raises the question of whether it is the job of a formal linguistic theory to explain or account for change, and to what extent that should be the case (see McMahon 2003, 2007; Bermúdez-Otero 2007a). If we are to evaluate linguistic models partly on the basis that good ones ought to allow us to model a trajectory from change to synchronic process, this raises questions both about the nature of change and about the nature of theory, and understandably these continue to be controversial issues. For example, Hale (2007) explicitly excludes these longer-term trajectories of change from the purview of formal linguistics, which is taken to be concerned only with the modeling of intergenerational changes in the grammar. On the other hand, modeling change has been taken very seriously in Optimality Theory, which also has a relevant mechanism in the shape of constraint reordering, and which in its stratal form shares architectural characteristics with LPM. It is interesting to note that the papers in Holt (2003), which despite the title are really almost all about phonology, quite often refer to LPM as well as to OT.

4.2 *Phonetics and phonology*

While the distinction between the lexicon and the post-lexicon is vitally important in modeling the two types of sound change originally envisaged by Labov (1981), and the possible diachronic transition between them, the bipartite structure of LPM also reconceptualizes the interplay between phonetics and phonology. McMahon *et al.* (1994) argue for a gestural model, as in Articulatory Phonology (Browman and Goldstein 1992; see also CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATIONS), which allows straightforward and perhaps explanatory modeling of phonetic processes and phonetically induced sound changes. They propose that this gestural account should be integrated into LPM, allowing different constraints to operate on the manipulation of gestures in the lexical component, and hence allowing some less restricted phonological processes than would otherwise be compatible with a gestural model. There is interesting support for this conclusion from studies of connected speech processes, where Holst and Nolan (1995) and Nolan *et al.* (1996), for instance, argue that not all such processes can be modeled simply as gestural overlap. However, this does not mean that lexical processes are phonological while post-lexical ones are phonetic; instead, it arguably begins to break down the distinction between phonetics and phonology, reintroducing the important category of language-specific phonetic processes, which form an important bridge between universal phonetics and language-specific phonology.

Again, this shift toward a more dynamic interaction between phonetics and phonology has two important consequences, this time partly methodological. If processes which might once have been relegated to phonetics are increasingly seen as language-specific and phonological, then methods previously restricted to phonetic investigation become extendable to the exploration of phonological issues. It might not be too far-fetched to see this as a contributory motivation (though obviously not the only one) for the establishment of laboratory phonology. Secondly, and again looking forward to the development of Optimality Theory, there is considerable ongoing debate about whether constraints are all, or should all be, essentially phonetically grounded. Flemming (2001, 2004) proposes a unified theory of phonetics and phonology, in which constraints are functionally motivated; this links the current section to the previous one, as it introduces a further option for explanation of changes and of their offspring, synchronic phonological consequences, this time in the form of constraints. Overall, this work causes us to re-evaluate the relationship between phonetics and phonology, or even to ask whether there is a distinction, and reintroduces concepts of gradience discussed far earlier by Kiparsky, for instance. Although the overall framework has moved on, and these discussions are no longer held within LPM as such, we again contend that the structure of LPM, and the dynamic possibilities it opened up, have enabled and encouraged such questions to be asked, and have sharpened our expectations of phonological (and phonetic) theory.

4.3 Stratal Optimality Theory

In this section so far, we have focused on aspects of LPM which do not translate directly into most current phonological models, though they have profoundly influenced the ways phonologists do their work and the expectations they have of their models and tools. However, there is one case in which direct structural influence from LPM feeds a current debate within Optimality Theory. In the last decade or so, several proposals have been offered for a theory of LPM-OT (Kiparsky 2000), Derivational OT (Rubach 2000), and Stratal OT (Bermúdez-Otero, forthcoming). While they may differ in their details, these theories have in common the adoption of the independent stem, word, and phrase levels of LPM. Otherwise they adopt most of the architecture of Optimality Theory – most notably, ranked violable constraints rather than language-specific ordered rules. The main justifications for a stratal OT are its ability to deal with opacity and cyclicity, and its adoption of the division of levels, which has found much support across languages and analytic frameworks. Proponents argue that a theory with the limited derivational capacity of levels is able to handle the very kinds of opacity that are attested, while opacity has been of persistent difficulty for purely parallel versions of Optimality Theory. (Capturing the lexical syndrome has not been a particular goal of these stratal OTs.) The reader is directed to McCarthy (2007) and references therein for discussion of how to describe opaque processes within OT. Generally, those OT theorists who support fully parallel evaluation argue that not all opacity can be described using strata, certainly not if the analysis is limited to strata that are independently motivated by the morphology.

A simple example from a dialect of Spanish will illustrate how stratal organization can be used to treat a certain kind of opacity and to capture a cyclic effect. Kaisse (2009) discusses a process in Spanish that is responsible for the realization of underlying /s/ as [h] in codas. In the Rio Negro dialect of Argentina, this process, known as "Aspiration," is transparent within morphemes and between morphemes, but it overapplies when one word ends in an underlying /s/ and the next word begins with a vowel.

(3)	/kas+a/	[ka.sa]	house+FEM SG	'house'
	/kasp+a/	[kah.pa]	dandruff+ FEM SG	'dandruff'
	/des+krem+ad+a/	[deh.kre.ma.ða]	NEG+cream+PPL+ FEM SG	'non-fat'
	/des+arm+ar/	[de.sar.mar]	NEG+arm+INF	'to disarm'
	/los##oxo+s/	[lo.ho.xoh]	the (MASC PL) eye+MASC PL	'the eyes'

In parallel OT, the opaque form [lo.ho.xoh] contains a gratuitous faithfulness violation, the first [h], which is not in a coda yet not faithful to the underlying /s/. The solution in a derivational framework would simply be that there is a cycle, with Aspiration following syllabification at the stem and word level but preceding (re)syllabification at the phrase level.

Kaisse shows that this example does not yield easily to a Sympathy analysis (the solution to most opacity in OT at the time the paper was written, in the early 2000s). However, independently motivated lexical *vs.* phrasal strata provide a limited type of serial derivation, which allows the output of the lexical stratum to form the input to the phrase level. We first derive [de.sar.mar],[3] [loh], and [o.xoh] as outputs of the word-level evaluation, shown in the first tableau in (4). The second tableau then shows how these outputs become the new inputs to the phrase level.

(4) *Word-level derivational OT tableau*

/des+armar/	Onset	*s]$_{CODA}$	Anchor(Rt,σ,L)	Ident(pl)
a. des.ar.mar	*!	*		
☞ b. de.sar.mar			*	
c. deh.ar.mar	*!			*
d. de.har.mar			*	*!

/los/	Onset	*s]$_{CODA}$	Anchor(Rt,σ,L)	Ident(pl)
a. los		*!		
☞ b. loh				*

/oxos/	Onset	*s]$_{CODA}$	Anchor(Rt,σ,L)	Ident(pl)
a. o.xos	*	*!		
☞ b. o.xoh	*			*

At the phrasal level, the outputs [loh] and [o.xoh] are strung together, and it becomes possible to fulfill the onset constraint by violating Anchor(Rt), the constraint which favors a coincidence of syllable boundaries with the edges of morphemes. The opacity emerges naturally from the tableau, because the input form for *los* already has /h/. There is no gratuitous faithfulness violation, as there was in the parallel OT derivation.

[3] Spanish prefixes, like those of Korean and Polish discussed in §3, act in some ways like independent words, with their consonants initially syllabifying into codas regardless of the identity of the following root-initial segment. Harris and Kaisse (1999) present derivational analyses of Rio Negro Argentinian Spanish that show that this separate syllabification occurs early, at the stem level, and is overridden by resyllabification between prefix and vowel-initial stem at the word level. Aspiration in Rio Negro is not a stem-level process – it turns on at the word level. Therefore, the /s/ in *desarmar* does not aspirate in Rio Negro.

(5) *Phrase-level derivational OT tableau*[4]

/loh##o.xoh/	Onset	*s]$_{\text{CODA}}$	Anchor(Rt,σ,L)	Ident(pl)
a. loh.o.xoh	*!			
☞ b. lo.ho.xoh			*	
c. los.o.xoh	*!	*		*
d. lo.so.xoh			*	*!

5 Conclusion

We have shown in this chapter that elements of Lexical Phonology and Morphology were either overstated or disproved by later data, and that not all aspects of the model can or should be integrated into later frameworks. Nonetheless, the perennial LPM concerns with the interaction of phonology with morphology and syntax, and with the concepts of opacity and cyclicity, may be recast in optimality-theoretic terms in current stratal, LPM-OT analyses. Much more importantly, the empirical phenomena which LPM highlighted and legitimized continue to form a central part of the debate.

REFERENCES

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Aronoff, Mark & Shikaripur N. Sridhar. 1987. Morphological levels in English and Kannada. In Edmund Gussmann (ed.) *Rules and the lexicon*, 9–22. Lublin: Katolickiego Uniwersytetu Lubelskiego.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Ph.D. dissertation, University of Massachusetts, Amherst.

Bermúdez-Otero, Ricardo. 2007a. Diachronic phonology. In de Lacy (2007), 497–517.

Bermúdez-Otero, Ricardo. 2007b. Word-final prevocalic consonants in English: Representation vs derivation. Paper presented at the Old World Conference in Phonology 4, Rhodes. Handout available (August 2010) at www.bermudez-otero.com/OCP4.pdf.

Bermúdez-Otero, Ricardo. Forthcoming. *Stratal Optimality Theory*. Oxford: Oxford University Press.

Bermúdez-Otero, Ricardo & April McMahon. 2006. English phonology and morphology. In Bas Aarts & April McMahon (eds.) *The handbook of English linguistics*, 382–410. Cambridge, MA & Oxford: Blackwell.

Booij, Geert & Rochelle Lieber. 1993. On the simultaneity of morphological and prosodic structure. In Hargus & Kaisse (1993), 23–44.

Booij, Geert & Jerzy Rubach. 1984. Morphological and prosodic domains in Lexical Phonology. *Phonology Yearbook* 1. 1–27.

Borowsky, Toni. 1993. On the word level. In Hargus & Kaisse (1993), 199–234.

[4] This particular example does not require the constraints to be differently ranked at the lexical and post-lexical level. However, the re-ranking option is also available, as seen in the example of linking and intrusive *r* in English in CHAPTER 85: CYCLICITY.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180.

Burzio, Luigi. 2002. Surface-to-surface morphology: When your representations turn into constraints. In Paul Boucher (ed.) *Many morphologies*, 142–177. Somerville, MA: Cascadilla Press.

Calabrese, Andrea. 2009. Markedness theory *vs.* phonological idiosyncrasies in a realistic model of language. In Eric Raimy & Charles Cairns (eds.) *Contemporary views on architecture and representations in phonology*, 261–304. Cambridge, MA: MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Collie, Sarah. 2008. English stress preservation: The case for "fake cyclicity." *English Language and Linguistics* 12. 505–532.

de Lacy, Paul (ed.) 2007. *The Cambridge handbook of phonology.* Cambridge: Cambridge University Press.

Fabb, Nigel. 1988. English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory* 6. 527–539.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18. 7–44.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 232–276. Cambridge: Cambridge University Press.

Hale, Mark. 2007. *Historical linguistics: Theory and method.* Oxford: Blackwell.

Hargus, Sharon & Ellen M. Kaisse (eds.) 1993. *Studies in Lexical Phonology.* San Diego: Academic Press.

Harris, James W. & Ellen M. Kaisse. 1999. Palatal vowels, glides and obstruents in Argentinian Spanish. *Phonology* 16. 117–190.

Harris, John. 1989. Towards a lexical analysis of sound change in progress. *Journal of Linguistics* 25. 35–56.

Hay, Jennifer. 2002. From speech perception to morphology: Affix-ordering revisited. *Language* 78. 527–555.

Hay, Jennifer & Ingo Plag. 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory* 22. 565–596.

Hayes, Bruce. 1990. Precompiled phrasal phonology. In Inkelas & Zec (1990), 85–108.

Holst, Tara & Francis Nolan. 1995. The influence of syntactic structure on [s] to [ʃ] assimilation. In Bruce Connell & Amalia Arvaniti (eds.) *Phonology and phonetic evidence: Papers in laboratory phonology IV*, 315–333. Cambridge: Cambridge University Press.

Holt, D. Eric (ed.) 2003. *Optimality Theory and language change.* Dordrecht: Kluwer.

Hyman, Larry M. 1993. Structure preservation and postlexical tonology in Dagbani. In Hargus & Kaisse (1993), 235–254.

Inkelas, Sharon. 1993. Deriving cyclicity. In Hargus & Kaisse (1993), 75–110.

Inkelas, Sharon & Draga Zec (eds.) 1990. *The phonology–syntax connection.* Chicago: University of Chicago Press.

Kahn, Daniel. 1976. Syllable-based generalizations in English phonology. Ph.D. dissertation, MIT.

Kaisse, Ellen M. 1990. Toward a typology of postlexical rules. In Inkelas & Zec (1990), 127–143.

Kaisse, Ellen M. 2009. Sympathy meets Argentinian Spanish. In Sharon Inkelas & Kristin Hanson (eds.) *The nature of the word: Studies in honor of Paul Kiparsky*, 199–214. Cambridge, MA: MIT Press.

Kaisse, Ellen M. & Sharon Hargus. 1993. Introduction. In Hargus & Kaisse (1993), 1–19.

Kaisse, Ellen M. & Patricia A. Shaw. 1985. On the theory of Lexical Phonology. *Phonology Yearbook* 2. 1–30.

Kang, Ongmi. 1992. Word-internal prosodic words in Korean. *Papers from the Annual Meeting of the North East Linguistic Society* 22. 243–257.
Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.
Kenstowicz, Michael. 1995. Cyclic *vs.* non-cyclic constraint evaluation. *Phonology* 12. 397–436.
King, Robert D. 1969. *Historical linguistics and generative grammar*. Englewood Cliffs, NJ: Prentice-Hall.
Kiparsky, Paul. 1973. Phonological representations: Abstractness, opacity, and global rules. In Osamu Fujimura (ed.) *Three dimensions of linguistic theory*, 57–86. Tokyo: Taikusha.
Kiparsky, Paul. 1982. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.
Kiparsky, Paul. 1983. Word formation and the lexicon. In Frances Ingemann (ed.) *Proceedings of the 1982 Mid-America Linguistics Conference*, 3–29. Lawrence: University of Kansas.
Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.
Kiparsky, Paul. 1988. Phonological change. In Frederick J. Newmeyer (ed.) *Linguistics: The Cambridge survey*, vol. 1: *Linguistic theory: Foundations*, 363–415. Cambridge: Cambridge University Press.
Kiparsky, Paul. 1993. Blocking in nonderived environments. In Hargus & Kaisse (1993), 277–313.
Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–365.
Labov, William. 1981. Resolving the Neogrammarian controversy. *Language* 57. 267–308.
Łubowicz, Anna. 2002. Derived environment effects in Optimality Theory. *Lingua* 112. 243–280.
McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.
McMahon, April. 2000. *Lexical Phonology and the history of English*. Cambridge: Cambridge University Press.
McMahon, April. 2003. On not explaining language change. In Ray Hickey (ed.) *Motives for language change*, 82–96. Cambridge: Cambridge University Press.
McMahon, April. 2007. Who's afraid of the Vowel Shift Rule? *Language Sciences* 29. 341–359.
McMahon, April, Paul Foulkes & Laura Tollfree. 1994. Gestural representation and Lexical Phonology. *Phonology* 11. 277–316.
Mohanan, K. P. 1982. Lexical Phonology. Ph.D. dissertation, MIT. Distributed by Indiana University Linguistics Club.
Mohanan, K. P. 1986. *The theory of Lexical Phonology*. Dordrecht: Reidel.
Nolan, Francis, Tara Holst & Barbara Kühnert. 1996. Modelling [s] to [ʃ] accommodation in English. *Journal of Phonetics* 24. 113–137.
Noyer, Rolf. 2004. Some remarks on the architecture of grammar with special attention to absolute neutralization. Paper presented at the symposium on Phonological Theory: Representations and Architecture, City University of New York.
Odden, David. 1993. Interaction between modules in Lexical Phonology. In Hargus & Kaisse (1993), 111–144.
Orgun, Cemil Orhan. 1996. Sign-based morphology and phonology with special attention to Optimality Theory. Ph.D. dissertation, University of California, Berkeley.
Pater, Joe. 2000. Non-uniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* 17. 237–274.
Patterson, David & Cynthia M. Connine. 2001. Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production. *Phonetica* 58. 254–275.
Plag, Ingo. 1996. Selectional restrictions in English suffixation revisited: A reply to Fabb (1988). *Linguistics* 34. 769–798.

Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation.* Berlin & New York: Mouton de Gruyter.

Raffelsiefen, Renate. 2004. Absolute ill-formedness and other morphophonological effects. *Phonology* 21. 91–142.

Rubach, Jerzy. 1984. Rule typology and post-lexical interference. In Stig Eliason (ed.) *Theoretical issues in contrastive phonology*, 37–50. Stuttgart: Julius Groos.

Rubach, Jerzy. 2000. Glide and glottal stop insertion in Slavic languages: A DOT analysis. *Linguistic Inquiry* 31. 271–317.

Siegel, Dorothy. 1979. *Topics in English morphology*. New York: Garland.

Sproat, Richard. 1993. Looking into words. In Hargus & Kaisse (1993), 173–195.

Steriade, Donca. 1999. Lexical conservatism in French adjectival liaison. In J.-Marc Authier, Barbara E. Bullock & Lisa A. Reed (eds.) *Formal perspectives on Romance linguistics*, 243–270. Amsterdam & Philadelphia: John Benjamins.

Sugahara, Mariko & Alice Turk. 2009. Durational correlates of English sublexical constituent structure. *Phonology* 26. 477–524.

Truckenbrodt, Hubert. 2002. Variation in p-phrasing in Bengali. *Linguistic Variation Yearbook* 2. 259–303.

Truckenbrodt, Hubert. 2007. The syntax–phonology interface. In de Lacy (2007), 435–456.

Wolf, Matthew. 2008. Optimal interleaving: Serial phonology–morphology interaction in a constraint-based model. Ph.D. dissertation, University of Massachusetts, Amherst.

95 Loanword Phonology

YOONJUNG KANG

1 Introduction

"Loanwords" are words borrowed from one language to another. These borrowed words usually undergo "adaptation" processes to conform to the structural constraints of the borrowing language phonology. Such adaptation affects all facets of phonological structure, reflecting the segmental, phonotactic, suprasegmental, and morphophonological restrictions of the borrowing language. The patterns that emerge in loanword adaptation often reveal aspects of native speakers' knowledge that are not necessarily obvious in data of the native language and, as a result, loanword data can inform our analysis of the native phonology (Hyman 1970; Holden 1976; Ahn and Iverson 2004; Kawahara 2008; Wetzels 2009; Chang, forthcoming, among others). In this respect, loanword adaptation can be considered a real-life *Wug test* (Berko 1958) which can enable us to probe into the grammatical knowledge of speakers in ways that native data alone cannot. Conversely, however, such emergent patterns in loanword adaptation present a learnability puzzle (cf. Broselow 2009): if a loanword pattern is underdetermined by the native phonology, where does the pattern come from? Also, what type of representation does the adaptation process refer to as it searches for licit forms in the borrowing language that most closely match the foreign language input? Is it an abstract phonological representation, a detailed phonetic representation, or a combination of the two? Are there any universal preferences for certain types of repair over others (e.g. epenthesis over deletion, or retention of a vocalic feature over a consonantal feature)? These are some of the major recurring questions in recent studies in loanword phonology and we will address them in this chapter.

This chapter is organized as follows. §2 examines various types of emergent patterns in loanword adaptation which pose a potential learnability puzzle, as well as examining proposed explanations for such emergent patterns. §3 provides a survey of instances of segmental adaptation and the levels of representation they refer to. §4 examines phonotactic adaptations.[1]

[1] There is a growing body of research on suprasegmental adaptation, which is not dealt with in this chapter due to limitations of space. Readers should refer to Y. Kang (2010) for a recent review of this topic.

2 Emergent patterns in loanwords

Research on loanword phonology intensified with the advent of constraint-based models of phonology, such as the theory of Constraints and Repair Strategies (Paradis 1988) and Optimality Theory (Prince and Smolensky 1993) in particular, largely due to their use of output constraints which could motivate the adaptation process even when particular processes themselves had no precedents in the native phonology (Yip 1993; Paradis and LaCharité 1997; Broselow 1999; Jacobs and Gussenhoven 2000). In addition, the principle of *Richness of the Base* in Optimality Theory naturally allows for, and perhaps even requires, an analysis of novel input forms which are not attested in native learning data (cf. Smolensky 1996), making the theory more amenable to the study of loanword adaptation phenomena. However, it has also been noted that constraint-based models only resolve one aspect of the puzzle (i.e. why an adaptation takes place at all), while leaving many other questions unanswered (Golston and Yang 2001; Peperkamp 2005; Broselow 2009). The puzzling emergent patterns in loanword adaptation identified in the literature can be classified into five categories.

1. THE TOO-MANY-SOLUTIONS PROBLEM: The *too-many-solutions* problem (Steriade 2001), or *differential faithfulness* (Broselow 2009), refers to how, given an offending structure in the foreign input, there is almost always more than one logically possible repair strategy, yet adaptation often converges on a specific strategy even when speakers have no apparent evidence for that process in their native language. For example, Hawaiian does not have a voiced stop /b/, and thus, as it is unattested in native phonology by definition, there is no direct evidence from the native phonology as to how such an illicit segment should be repaired. Yet English /b/ is systematically adapted as /p/ (*boulder* → [polu'ka:]) and not /m/, /w/, or any other segment of the Hawaiian inventory (Adler 2006). Similarly, a repair for a phonotactic constraint violation presents a wide range of logically possible choices. For example, when an onset cluster (C_1C_2V) is borrowed into a language which bans complex onsets, the structural requirements of the native language can be satisfied by the deletion of a consonant (> C_2V or C_1V), an option found in French loanwords in Vietnamese (*crème* → [kem]), or by the epenthesis of a vowel in front of the cluster (> vC_1C_2V) or inside the cluster (> C_1vC_2V). The epenthesis repair is found in Japanese (*Christmas* → [kurisumasu]) and Hawaiian (*Christmas* → [kalikimaki]), among other languages, and these examples also illustrate some of the different possibilities in the quality of the epenthetic vowel (Broselow 2006). How, given all these options, do adapters converge on a specific repair strategy?
2. DIVERGENT REPAIR: Even more puzzling is the fact that the repair chosen sometimes seems to contradict the native repair strategy – a situation referred to as *divergent repair* by Kenstowicz (2005) and as *ranking reversal* by Broselow (2009; see also CHAPTER 70: CONSPIRACIES). For example, Thai requires the final syllable of a word to be heavy. This requirement is satisfied by glottal stop insertion in native words, as in /pʰrá/ → [pʰráʔ] 'monk', but by vowel lengthening in English loans, as in *coma* → [kʰō:mā̂:]. In native Korean, a restriction against a sequence of an obstruent + nasal is repaired by nasalization, as

in /kuk-mul/ [kuŋmul] 'soup', but epenthesis is the dominant repair option employed for English loanwords, as in *picnic* → [pʰikʰɨnik] (see Peperkamp *et al.* 2008: 156 for more examples of divergent repairs).

3. Unnecessary repair: Moreover, there are cases where adaptation takes place even when there is no apparent illicit structure in need of repair – a situation referred to as *unnecessary repair* by Peperkamp (2005). For example, Korean allows voiceless stops in coda position, but English voiceless stops are variably adapted with vowel epenthesis, as in *cut* → [kʰʌtʰɨ] ~ [kʰʌt] (Y. Kang 2003). Also, an English cluster of a coronal stop + [w] is adapted with epenthesis in Korean (*twin* → [tʰɨwin], *[tʰwin]) (H. Kang 2006), even though such clusters are allowed in Korean. In Hmong, /ʒ/ in French loanwords is adapted as /j/, despite the fact that /ʒ/ is a phoneme in Hmong (/ʒo.zɛf/ *Joseph* → /jɔ̀.sè/, */ʒɔ̀.sè/) (Golston and Yang 2001). Also, in French loanwords in Japanese and Korean, an epenthetic vowel is added "unnecessarily" following word-final nasals (French [kan] *Cannes* → Japanese [kannu], *[kan] (Shinohara 1997; Peperkamp *et al.* 2008); French [kom] *comme* → Korean [k'omɨ], *[k'om] (H. Kang 1996)).
4. Differential importation: Yet another type of puzzling emergent pattern is *differential importation*. Importation refers to a situation where a structure not attested in native phonology is exceptionally allowed in loanwords. While such importation in and of itself is not a problem from a learnability perspective, the fact that only certain structures, but not others, are imported requires an explanation (Holden 1976; Itô and Mester 1995; 1999, 2001; Davidson and Noyer 1997; Broselow 2009). Given foreign input with two types of novel structures which are both equally unattested in the native data, why is one structure readily allowed into the language but not the other? For example, in Hawaiian, the fully nativized form of the English word *truck* is [kə'lakə]. Also possible is a "less Hawaiian" variant [tə'lakə], where English /t/ remains unadapted. But the variant *['krakə], where the complex onset is retained, but /t/ is adapted as /k/, is judged to be impossible (Adler 2006). In other words, the restriction against /t/ is more easily relaxed than the restriction against an onset cluster. In Russian, the requirement that only palatalized consonants occur before /e/ is often violated in adaptation, but the process of reducing unstressed /o/ and /e/ is more likely to be upheld (Holden 1976).
5. Retreat to the unmarked: While importation is a situation where native constraints are relaxed in the loanwords, allowing a wider range of output structure in the loanword stratum than in the native stratum, we also find the opposite situation, i.e. that loanwords conform to *stricter* structural requirements than the native phonology, such that the foreign input is transformed to an unmarked form, even when there is a seemingly more faithful licit form available in the language. Kenstowicz (2005) refers to such cases as *retreat to the unmarked*. Pitch accent assignment in Japanese and Korean exhibits an emergence of a default accent assignment pattern – the most general pattern being to accent the penultimate syllable in Kyungsang Korean (Kenstowicz and Sohn 2001; Lee 2009) and the antepenultimate mora or syllable in Tokyo Japanese (Shinohara 2000; Kubozono 2006). Similarly, tone assignment in loanwords in White Hmong (Golston and Yang 2001), Thai (Kenstowicz and Suchato 2006), Tibetan (Hsieh and Kenstowicz 2008), Taiwanese (Hsieh 2006), Mandarin (Wu 2006), and Vietnamese (Barker 1969) is based on the segmental

composition of the input. The "retreat to the unmarked" is also found in the segmental domain. In Hungarian, word-final voiceless obstruents in monosyllabic loanwords are geminated, in an apparent requirement for syllables to be heavy (e.g. *shock* → [sokk], *[sok]), but "Hungarian does not have a requirement for syllables to be heavy" and this is a case of "a peripheral stratum of the lexicon introducing a requirement which is not a part of the core-stratum" (Kertész 2003).[2] In Thai, English word-initial voiceless stops are generally adapted as aspirated stops, retaining the aspiration of the input (e.g. English [pʰ]*in* → Thai [pʰin], [tʰ]*eam* → [tʰiːm], [kʰ]*one* → [kʰoːn]). The only exceptions are when there is an unaspirated stop in the same word, as in [pʰ]*eg* → [pek], *[pʰek], and [kʰ]*ook* → [kuk], *[kʰuk], indicating a preference for a non-aspiration harmony – a generalization not present in the native phonology (Kenstowicz and Suchato 2006). The retreat to the unmarked in many of these cases is particularly puzzling, because there is no clear evidence for the "unmarked" status of the resulting structure in the native data.

We now turn to explanations for emergent patterns in loanword phonology in the literature, which can be grouped into five broad categories, which are not necessarily mutually exclusive.[3]

1. Native phonology: The first possibility is that the adaptation pattern is indeed a reflection of language-specific facts of the native phonology – the loanword pattern only appears to be novel. For example, it has been proposed that in segmental adaptation, the choice regarding which feature to preserve and which feature to sacrifice is informed by the status of the features in the native phonology (Hancin-Bhatt 1994; Clements 2001; Herd 2005, among others; see §3.1 for further discussion). Rose and Demuth (2006) propose that the choice of the epenthetic vowel quality in English and Afrikaans loanwords in Sesotho is predictable from the contrastive feature specifications of the native phonology. The native phonology generalizations that affect loanwords can also take the form of covert statistical generalizations. Zuraw (2000) demonstrates that the variable application of nasal substitution in loanwords in Tagalog is a direct reflection of statistical tendencies in the native lexicon. It has also been argued that the default accentuation in loanwords in Tokyo Japanese has a direct correlate in native phonology as a covert default in the lexicon as a whole (Kubozono 2006). Luke and Lau (2008) show that in recent English loans in Cantonese, verbs are generally truncated to become monosyllabic,

[2] Consonant gemination in loanwords is a widespread phenomenon, found in Japanese (Katayama 1998; Shinohara 2004; Kato 2006; Kubozono *et al.* 2008), Italian (Repetti 1993, 2006, 2009), Finnish (Karvonen 2005), Maltese and Egyptian Arabic (Hafez 1996), and Kannada (Sridhar 1990). While some writers analyze gemination in loanwords as the emergence of the unmarked (Repetti 1993; Shinohara 2004; Kubozono *et al.* 2008, among others), others argue that the gemination is motivated by the preservation of the input structure and does not necessarily result in less markedness of the output (Kato 2006; Repetti 2006). Repetti (2006), in fact, refers to the gemination in loanwords in Italian as a case of the emergence of "marked" structures. Under these alternative interpretations, consonant gemination can be categorized as a case of (seemingly) unnecessary repair ((3) above), rather than a retreat to the unmarked.

[3] Another possibility not listed here is to posit a separate set of principles or constraints governing loanword adaptation (Paradis and LaCharité 1997). Such approaches may be considered a type of UG approach if we assume that UG contains a separate set of principles for loan adaptation.

whereas nouns tend to be bisyllabic, and this pattern conforms to lexical statistics in native words. Walter (2006) argues that gender assignment in Arabic loanwords in Spanish also mirrors phonological generalizations and statistical tendencies in the native lexicon.

2. Default setting of UG: However, many emergent patterns in loanwords still elude explanations based on covert generalizations in native phonology. Some attribute these emergent patterns to default settings of Universal Grammar. Default accentuation in Northern Kyungsang Korean exhibits a case of this retreat to the unmarked – a pattern which does not appear otherwise to be motivated in the native phonology and which must thus be attributed to UG (Kenstowicz and Sohn 2001). Uffmann (2006) resorts to a universal markedness hierarchy to account for the epenthetic vowel quality in loanwords in Shona, Sranan, Samoan, and Kinyarwanda. In their influential work on the lexical stratification of Japanese, Itô and Mester (1995, 1999, 2001) argue that the differential importation of foreign features in loanwords reveals covert constraint rankings from the initial state of UG that lie latent in the native phonology (also see Shinohara 2000, 2004).[4] Similarly, Davidson *et al.* (2004) argue that the differing rates in the correct L2 production of foreign clusters reveal a covert ranking in the initial state of UG. Analyses that resort to the universal hierarchy of perceptual similarity (*P-map*) to account for adaptation patterns can also be grouped into this category (Fleischhacker 2005; Shinohara 2006; Kawahara 2008). We will return to this P-map-based approach below.
3. Adaptation as perception: The next possibility is that a seemingly puzzling adaptation in fact takes place during the perception of foreign input and not in the computation of the production grammar (Silverman 1992; Peperkamp and Dupoux 2003; Peperkamp *et al.* 2008; Boersma and Hamann 2009; Calabrese 2009). Based on findings that the perception of foreign sounds is constrained by the segmental and structural constraints of the native language (Massaro and Cohen 1983; Werker and Tees 1984; Dupoux *et al.* 1997; Dupoux *et al.* 1999), it is argued that most, if not all, of the adaptation in fact takes place during the perception of foreign input. This approach breaks away from the assumption that the input to the production grammar in loanword adaptation faithfully retains the phonetic and/or phonological structure of the source language input (cf. Jacobs and Gussenhoven 2000; LaCharité and Paradis 2005). This view provides a solution to many puzzling adaptations, such as *unnecessary repair* or *divergent repair*, where the adaptation pattern seems to contradict the production grammar of the borrowing language.[5]

   While some propose that perceptual adaptation is one of many steps in adaptation (e.g. Silverman 1992; Kenstowicz 2003; Broselow 2009), others propose that "*all* loanword adaptations are phonetically minimal transformations that

[4] Crawford (2007, 2008) argues that the different degree of nativization is grammar-external (i.e. the apparent differential adaptation is a reflection of differing transmissibility of foreign features as they spread from bilingual to monolingual populations). Davidson (2007) also shows how foreign structures can initially be introduced by bilingual speakers and evolve further during transmission to the monolingual population.

[5] Under this view, the burden of explanation for the adaptation pattern is passed on to the perception module and the question of learnability is still not fully resolved – i.e. why is the novel foreign input perceived the way it is? Some cases of perceptual adaptation have a comparable precedent in the native language, but not all do.

apply during speech perception" [emphasis original] (Peperkamp 2005; see also Peperkamp *et al.* 2008 and Boersma and Hamann 2009).[6] Peperkamp *et al.* (2008: 132) state:

> [D]ue to the automatic character of perceptual assimilation and the primacy of perception over production, [our psycholinguistic model] only allows for a limited number of loanword adaptations that are not due to distortions during speech perception.

Boersma and Hamann (2009) take the strict view that perception is largely bound by native output constraints, so that a structure that violates native constraints cannot be perceived faithfully. Peperkamp *et al.* (2008), on the other hand, allow for the possibility that foreign structures may be correctly perceived, leading to importation rather than adaptation. Overall, this approach makes the strong empirical prediction that adaptation is tightly correlated with perception. See CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY for further discussion.

4. PERCEPTUAL SIMILARITY: Steriade (2001) proposes that speakers possess knowledge of perceptual similarity between strings of sounds (the P-map), which is utilized in loanword adaptation (see Yip 2002, 2006; Y. Kang 2003; Kenstowicz 2003, 2007; Fleischhacker 2005; Kawahara 2006; Miao 2006; Shinohara 2006; Y. Kang *et al.* 2008, among others, for the application of P-map constraints to loanword adaptation).[7] The perceptual similarity approach, similar to the adaptation-as-perception approach, places an emphasis on perceptual factors and phonetic details in accounting for adaptation patterns. However, the former differs from the latter in that perceptual factors are incorporated into grammatical constraints which can be ranked with respect to other grammatical constraints. As a result, the P-map dictates that the repair be as perceptually minimal as possible, but does not necessarily dictate whether adaptation has to actually take place; this is determined by the relative ranking of P-map-based faithfulness constraints with respect to native structural constraints (Steriade 2007). In other words, this approach allows for the possibility that foreign input is faithfully perceived, yet can nevertheless be adapted to adhere to native phonotactic constraints. For example, Kabak and Idsardi (2007) examined Korean speakers' perception of English consonant clusters that violate phonotactic restrictions in Korean and found that some clusters that undergo adaptation in loanwords, such as [gm] and [km], were correctly perceived as distinct from their counterparts with an epenthetic vowel, i.e. [gɯm] and [kɯm] respectively. This contrasts with the adaptation-as-perception view, where the connection between perception and adaptation is much tighter. Yip (2002: 10) notes:

> [R]eference to perceptual salience within the phonology proper resolves the paradox that quite subtle non-native distinctions are clearly perceived, but nonetheless less salient segments are more likely to be sacrificed than highly salient ones.

[6] Peperkamp (2005) and Peperkamp *et al.* (2008) acknowledge the possibility that some adaptation processes may be due to non-perception-related factors (e.g. cases of "retreat to the unmarked").

[7] The knowledge of the P-map is largely universal, and therefore this approach can be categorized under the UG approach. However, Steriade (2001: 243) also leaves open the possibility that the specifics of the P-map can differ according to language-specific experiences, and Y. Kang (2003), for example, employs the P-map model, but assumes that some aspects of the P-map can be language-specific.

A strict interpretation of the P-map, as presented in Steriade (2009), predicts a strong correspondence between native repair and loan repair, since the P-map hierarchy affects both native and loanword phonology. However, cases of divergent repair, discussed above, contradict this strict interpretation. Y. Kang (2003), Kenstowicz (2005), and Yip (2006), on the other hand, propose loanword-specific mapping constraints, variously named Be Similar, Match, Mimic, or output–output faithfulness constraints, which are distinct from native input–output constraints (see chapter 63: markedness and faithfulness constraints). These constraints require preservation of the source language information, and – depending on the relative ranking of these constraints with respect to native markedness and faithfulness constraints – importation (F_{loan} >> M >> F_{native}), retreat to the unmarked (F_{native} >> M >> $F_{foreign}$), or divergent repair ($F1_{loan}$ >> $F2_{loan}$, while $F2_{native}$ >> $F1_{native}$) can occur (Kenstowicz 2005).

Yip (2006) also notes that adaptation by perception alone is too deterministic to account for the range of variation attested in data. For example, a single language, such as Mandarin, can show different adaptation strategies for illicit codas in English loanwords, namely epenthesis in Mainland Mandarin (*Friedman* → [fu.li.tə.man]) and deletion in Taiwanese Mandarin (*Friedman* → [fu.li.man]). Similarly, Adler (2006) reports variation between deletion and epenthesis in English loanwords in Hawaiian. In such cases, perception underdetermines adaptation and the ranking of grammatical constraints determines the final outcome. Broselow (2009) and Peperkamp *et al.* (2008), concerned with the unconstrained nature of these loanword-specific mapping constraints which can be freely ranked across languages, argue that adaptation is actually much more constrained by perceptual factors than is predicted by loanword-specific faithfulness constraints.

5. Grammar-external factors: Finally, it has been proposed that sociolinguistic or grammar-external factors affect the pattern of (non-)adaptation, especially where aspects of loanword phonology are underdetermined by grammatical factors. First of all, the rate of importation has been shown to positively correlate with the level of bilingualism in the community (Haugen 1950; Paradis and LaCharité 1997, 2008, 2009; Heffernan 2007; Friesner 2009a). The level of bilingualism has also been argued to determine the mode of adaptation: the higher the level of bilingualism, the more likely the adaptation will refer to phonological representations over phonetic representations of the input language (Heffernan 2007).[8] The channel of borrowing (i.e. spoken *vs.* written) and the related influence of orthography have also been proposed to affect the adaptation pattern (Dohlus 2005; Smith 2006; Vendelin and Peperkamp 2006; Detey and Nespoulous 2008; Friesner 2009a). It is also proposed that adapters look to orthography, especially when other factors underdetermine the adaptation pattern and the adapters are uncertain about the "correct" pattern (Y. Kang 2009). For example, in 1930s Korean, non-preconsonantal /s/ in English loanwords was variably adapted as lax /s/ or tense /s'/, the latter written as geminate <ss> in Korean orthography. Whether the English /s/ was written with a single or double <s> had a significant effect on the choice between the two adaptation patterns.

[8] See Paradis and LaCharité (2008, 2009) for arguments against this view that the mode of adaptation is dependent on the level of bilingualism.

3 Segmental adaptation

When the foreign input contains a non-native segment, the segment is replaced with the "closest" sound in the native language. The main problem here is how to define the "closest" sound. For example, when a front rounded vowel ([−back, +round]) is borrowed into a language that lacks such vowels, the vowel undergoes transformation. The French high rounded vowel [y] is adapted as [u] in White Hmong (Golston and Yang 2001), preserving the rounding feature of the input, as [i] in Mauritian Creole (Jacobs and Gussenhoven 2000), Fula, Kinyawaranda, and Lingala (Paradis 2006), preserving the backness feature of the input, and variably as [u] or [i] in Moroccan Arabic (Paradis 2006) and Egyptian Arabic (Hafez 1996). The same vowel is adapted as [ju] in Japanese (Dohlus 2005), preserving both the backness and rounding features of the input segment, but creating a bisegmental structure. Such diverse substitution indicates that there is no universal metric of similarity that all languages follow in segmental adaptation. How then is similarity determined in segmental adaptation? A related key issue in the literature on loanword phonology, and segmental adaptation in particular, is the level of representation that the similarity calculation refers to. This question can be examined in two parts: the nature of the *borrowing* language representation (§3.1) and the nature of the *input* language representation (§3.2).

3.1 *Borrowing language representation*

As for the level of representation of the *native* language which acts as a sieve for the foreign language input, it has been hypothesized that phonological features in the input form that are underlyingly contrastive in the native phonology are preferentially preserved over features that are redundant and non-contrastive in the native phonology (Clements 2001; Herd 2005; Dresher 2009, among others). For example, in all Indo-Aryan languages that maintain a contrast between dental and retroflex stops, English alveolar stops ([+anterior, −distributed]) are consistently adapted as retroflex stops ([−anterior, −distributed]), rather than as dental stops ([+anterior, +distributed]) (Lehiste 1988; Arsenault 2009). For example, English *taxi* and *soda* are adapted as /ʈæksi/ and /soɖa/, respectively, in Hindi. The observed adaptation preserves the [−distributed] feature of the English input while sacrificing the [+anterior] feature. Arsenault (2009) proposes that this is due to the fact that [distributed] is a phonologically active feature in these host languages and [anterior] is not.

Similarly, Clements (2001) discusses the adaptation of English consonants to Hawaiian and proposes that substitution preserves the contrastive feature specification of the native language, which is determined by the composition of the native inventory, as well as a universal hierarchy of feature accessibility. Herd (2005) applies Clements's (2001) model to the adaptation of English consonants in other Polynesian languages. For example, both Hawaiian and New Zealand Maori lack sibilants, and English sibilants /s z ʃ ʒ/ are adapted to /h/ in New Zealand Maori, but the same English sibilants map to /k/ in Hawaiian, despite the fact that both /k/ and /h/ are available in both languages.[9] According to Herd

[9] Adler's (2006) elicitation data, however, show that sibilants are variably adapted as /k/, /h/, or null (i.e. are deleted) in Hawaiian.

(2005), the crucial difference between Hawaiian and New Zealand Maori is that in Hawaiian /h/ contrasts with /ʔ/ and, as a result, /h/ is contrastively specified for [+spread glottis], thereby creating a mismatch with English sibilants, while in New Zealand Maori /h/ does not contrast with another glottal sound and therefore is unspecified for [+spread]. At the same time, in New Zealand Maori, /t/ and /k/ contrast and /k/ is contrastively specified for [Dorsal], which creates a mismatch with English sibilants. Similarly, Hancin-Bhatt (1994) and Brown (2000) claim that features that play a higher contrastive function in the L1 are more likely to be preserved in the modification of foreign sounds in L2 production.

An alternative view is that phonetically salient input characteristics are preferentially preserved over less salient characteristics and that the abstract phonological status of those characteristics – i.e. whether they are underlyingly contrastive or not – is not particularly relevant (Brannen 2002; Hsieh and Kenstowicz 2008; Lin 2008; Hsieh *et al.* 2009; Steriade 2009). Given that phonologically contrastive features tend to be phonetically salient and that there is also a degree of indeterminacy in phonological analyses of contrastiveness, the two views often converge on the same predictions, but not always (see Y. Kang 2008a for more discussion on this issue).

For example, Brannen (2002) proposes that the English interdental fricative /θ/ is adapted differently in European French and Quebec French, mapping to /s/ in European French and to /t/ in Quebec French. The crucial difference between the two dialects is that in European French the coronal fricative /s/ has a dental place of articulation, while in Quebec French /s/ is alveolar, a phonetic detail that is not contrastive in either of the dialects. On the basis of these findings, Brannen (2002) argues that adaptation is sensitive to non-contrastive, but phonetically salient, features, such as stridency and minor place of articulation features. The adaptation of English vowels in Mandarin (Lin 2008) is another case where non-contrastive features of the borrowing language are preferentially preserved at the cost of contrastive features. In Mandarin non-high vowels, height is contrastive but backness is not – it is predictable from the context (Duanmu 2000). If the phonological status of these features directly influences the segmental adaptation, we would expect height to be preserved and backness to be sacrificed when the input vowel has to be modified due to phonotactic or semantic restrictions. However, contrary to this prediction, Lin's (2008) survey finds that when English vowels are adapted in Mandarin, vowel height is routinely altered, but backness is fairly consistently preserved (on the relation between backness and vowel height see also CHAPTER 21: VOWEL HEIGHT). A similar tendency is also found in English loanwords in Cantonese (Yip 2002).

Proponents of perceptual similarity assume that there is a quasi-universal hierarchy of featural salience such that "certain features are inherently more salient than others" (Brannen 2002). For example, Steriade (2009) states that "stricture differences ([sonorant], [continuant], [consonantal]) play the major role in generating dissimilarity judgments, in contrast to voicing and place."[10] This hierarchy is compatible with the adaptation pattern found in Selayarese (Broselow 1999) and

[10] Clements's feature hierarchy (2001: 80, repeated below) also predicts quasi-universal tendencies in selective feature preservation, although in his model, only features that are contrastive in the native language are relevant in loanword adaptation: [coronal] > [sonorant] > [labial] > [dorsal] > [strident] > [nasal] > [posterior] > [lateral] > [voice].

Mandarin (Miao 2006). In Selayarese, the coda position is restricted to glottal stops and velar nasals, and potential violations of the coda restriction in loanwords from Bahasa Indonesian are repaired by changes in place of articulation, but never by changes in nasality or continuancy. Also, in a comprehensive examination of segmental adaptation in English, German and Italian loanwords in Mandarin, Miao (2006) concludes that segmental adaptation obeys the following hierarchy: IDENT(Manner) >> IDENT(Major Place) >> IDENT(Place) >> IDENT (Voice/Aspiration). Brannen (2002), on the other hand, provides a slightly different hierarchy to account for the adaptation of English dental fricative in various languages: Turbulence (strident/mellow) > Major Articulators (Labial /Coronal) > Airflow (stop/continuant) and Location (Lip/Dental/Alveolar) > Minor Articulators (Laminal/Apical). It remains to be seen to what extent the claim of universal hierarchy holds true, either as an absolute principle or as a general tendency, in view of the full body of data.

Finally, the adaptation-as-perception view claims that segmental adaptation is also a result of L1 speech perception applied to the foreign acoustic input (Silverman 1992; H. Kim 2008, 2009; Peperkamp *et al.* 2008; Boersma and Hamann 2009). H. Kim (2008, 2009) proposes that segmental adaptation is due to L1 feature-driven perception. Foreign input is mapped to a phonological category via perceptual mapping of the acoustic signal to relevant featural representations. The concept of "feature" in her model is somewhat more abstract than what is assumed in the other studies discussed above. For example, a [tense] feature can be signaled by a combination of acoustic correlates, such as duration and pitch of the adjacent vowel, and the same feature can be signaled by different combinations of acoustic correlates depending on the context within which it occurs.

As mentioned above, the adaptation-as-perception view predicts a very strong correlation between perception and adaptation. The strongest view, where all foreign segments are obligatorily transformed to native sounds (Silverman 1992; Boersma and Hamann 2009, among others) has difficulty accounting for how some foreign contrasts can be easily perceived (Best 1994), and some foreign segments are adopted without adaptation (i.e. imported). Also, the view that all adaptation occurs during perception is unable to explain cases where adapters can perceive the foreign contrast but nevertheless adapt it to a native segment. Also, there are cases where the perception results do not match the adaptation results. For example, Brannen (2002) shows that [f] is the segment that is most likely to be confused with [θ] by French speakers, yet [t] or [s] is the consonant of choice in the adaptation of English [θ].

3.2 *Input language representation*

With respect to the input to the adaptation process, some argue that the input is the phonological representation of the source language, devoid of redundant phonetic details (Paradis and LaCharité 1997; Shinohara 2004; LaCharité and Paradis 2005). This view, referred to here as the "phonological input" view, predicts uniform adaptation of a source language phoneme across different contexts (i.e. phonemic uniformity). Others assume that the input is the acoustic representation of the source language, including all subphonemic phonetic details of the source language sounds (Silverman 1992; Yip 1993; Peperkamp 2005; Iverson and Lee 2006; Peperkamp *et al.* 2008). The latter view, referred to here as the "phonetic input" view, predicts that a given phoneme of the source language can be adapted differently in different segmental contexts depending on its surface phonetic characteristics in the input language.

Empirical research accumulated over the years, however, shows that adaptations of both types are amply attested. The range of possibilities is well illustrated by the different adaptation of English voiceless stops in languages with aspiration contrasts in voiceless stops. English voiceless stops vary in aspiration depending on the context. The "phonological input" view predicts that the adaptation should be uniform, regardless of position, barring modifications due to phonotactic restrictions. The "phonetic input" view, on the other hand, predicts adaptation to aspirated and unaspirated stops, depending on the allophonic realization of the English input stops. An example of adaptation based on "phonological input" is found in Korean, where English voiceless stops are consistently adapted as aspirated, even in contexts where the English input is unaspirated, as in [tʰ]*oy* *s*[t]*ory* → [tʰoi sitʰori].[11] Examples of adaptation based on "phonetic input" are found in Cantonese (Silverman 1992; Yip 1993) and Thai (Kenstowicz and Suchato 2006). In Thai, English voiceless stops are adapted as aspirated in word-initial position, but as unaspirated following /s/. In word-medial position and elsewhere, the adaptation varies between aspirated and unaspirated stops, with the aspirated adaptation being more likely in pretonic position.

Burmese and Mandarin present a mixed picture. In Burmese, voiceless stops are adapted as unaspirated voiceless stops in most contexts, as in *Poland* → [pou.lã], except in English word-initial TR sequences, where the stop is adapted as aspirated, as in *cream* → [kʰə.jĩ] (Chang 2009). In Mandarin, the majority pattern is to adapt English voiceless stops as aspirated stops regardless of the input aspiration as in [pʰ]*izza* → [pʰitsa] ~ [pʰisa] and *hi*[p]*ies* → [sipʰiʂ] (Paradis and Tremblay 2009). However, the aspirated adaptation is proportionately much more likely to occur when the English input stop is aspirated (88.8 percent) than when it is unaspirated (78.8 percent), indicating some influence of the phonetic information.

Additional examples of segmental mapping based on phonological *vs.* phonetic input found in the literature are listed in (1).

(1) a. *Phonetic input*

Thai: English /v/ is mapped to /w/ in the onset and to /p/ in the coda (Kenstowicz and Suchato 2006).

Jahai: Malay /k/ is generally adapted as /k/, but as /ʔ/ word-finally, reflecting the allophonic realization of /k/ in the Malay input (Burenhult 2001).

Fon: French /r/ is mapped to /xl/ word-initially, to /l/ in non-initial prevocalic position, and is deleted in preconsonantal or word-final position (Gbéto 2000, as discussed in Kenstowicz 2003).

Korean: English /s/ is adapted as /s'/ in prevocalic or word-final position and as /s/ elsewhere (S. Kim and Curtis 2002; Ahn and Iverson 2004; Davis and Cho 2006; Y. Kang 2008c).

1930s Korean: English voiced stops /b d g/ are mapped to tense stops /p' t' k'/ word-initially and as lax stops /p t k/ elsewhere (Y. Kang 2008b).

[11] See Oh (2004) and Kenstowicz (2005) for alternative analyses.

Yanbian Korean: Mandarin unaspirated voiceless stops/affricates /p t k ts tɕ tʂ/ are mapped to tense stops/affricates /p' t' k' c'/ word-initially, but to lax stops/affricates /p t k c / elsewhere (Ito and Kenstowicz 2009).

Korean: Japanese voiceless stops are adapted as lax stops word-initially, but as lax or tense stops in word-medial position, depending on the place of articulation (Ito *et al.* 2006).

Hawaiian: English /t d/ is mapped to /k/ in general but the adaptation varies between /k/ and /ʔ/ word-finally (Adler 2006).

Palauan: Palauan lacks /g/ as a phoneme. Japanese /g/ is adapted as /k/ word-initially but as /ŋ/ word-medially, reflecting the allophonic variation of Japanese /g/ (Takahashi 2006).

b. *Phonological input*

Thai: English voiced stops are adapted as voiced word-initially, where they are significantly devoiced, and hence voiceless adaptation may be expected under the phonetic adaptation view (Kenstowicz and Suchato 2006).

Korean: English non-morphemic /z/ is consistently adapted as /c/, even in contexts where /z/ is significantly devoiced and is better matched by /s/ or /s'/ (Y. Kang 2009).

See LaCharité and Paradis (2005) for more examples.

The empirical evidence overwhelmingly suggests that loanword adaptation is not completely based on phonetic or phonological representation; rather, both types of adaptations are attested, oftentimes within a single contact situation, leading many to conclude that the adaptation process can refer to both phonological and phonetic (as well as morphological,[12] semantic, and orthographic) details of the source language (Y. Kang 2003, 2009; Adler 2006; Kenstowicz and Suchato 2006; Smith 2006; Friesner 2009a, 2009b; Chang, forthcoming, among others).

It seems reasonable to assume that adapters have some knowledge of the input language phonology, which is accessible in adaptation – contra the view that the input to adaptation is an unstructured acoustic signal (Silverman 1992). At the same time, simply the fact they know phonology does not mean that it is the only aspect they pay attention to (Y. Kang 2008a; Chang, forthcoming) – contra Paradis and LaCharité (2009). From this perspective, a more appropriate question is not whether loanword adaptation is phonological or phonetic, but rather what factors make one type of adaptation more likely than the other. For example, aspects of the contact situation, such as the level of bilingualism or the spoken *vs.* written channel of borrowing, have been proposed as factors that help shape the adaptation pattern (Dohlus 2005; Rose and Demuth 2006; Smith 2006; Heffernan 2007; Y. Kang 2008c, 2009).

4 Phonotactic adaptation

I now turn to adaptation due to phonotactic restrictions. Here again, we are faced with the "too-many-solutions" problem and many seemingly puzzling emergent adaptation patterns. In §4.1, I will examine the claim that epenthesis is the generally

[12] See Repetti (2006, 2009), Y. Kang (2009), and Friesner (2009b) for discussion on the role of morphological information in the input language in adaptation.

preferred adaptation over deletion (Paradis and LaCharité 1997; Shinohara 2006). In §4.2, I will examine co-occurrence restrictions where the adaptation forces the preservation of certain features over others.

4.1 Epenthesis vs. deletion

Paradis and LaCharité (1997) propose the *Preservation Principle*, which dictates that the input material be preserved as much as possible, unless the cost of preservation is too extreme (see also CHAPTER 76: STRUCTURE PRESERVATION and CHAPTER 67: VOWEL EPENTHESIS). In other words, epenthesis should generally be preferred over deletion. In this section, I will provide a survey of word-initial and word-final cluster adaptations to examine whether epenthesis is indeed the preferred option overall.

4.1.1 Word-initial clusters

As for word-initial onset clusters, a survey of available cases indeed shows that epenthesis is the predominant choice of repair. Epenthesis-only adaptation is found in Burmese (Chang, forthcoming), Egyptian Arabic (Hafez 1996), Farsi (Shademan 2003), Fijian (Schütz 1978), Fula (Paradis and LaCharité 1997), Hindi (Singh 1985), Huave (Davidson and Noyer 1996), Inuktitut (Pollard 2008), Japanese (Katayama 1998),[13] Kikuyu (Mwihaki 2001), Korean (H. Kang 1996, among others), Samoan (Cain 1986), Sesotho (Rose and Demuth 2006), and Shona (Uffmann 2006), among others. For example, Inuktitut does not allow consonant clusters, and onset clusters are repaired by epenthesis as in *(Santa) Claus* → [kalasi] and *Scotia* → [sikusa] (Pollard 2008).

There are many fewer languages that only employ a deletion strategy in word-initial clusters and many of them are creole languages where only /s/ + stop clusters are repaired (see also CHAPTER 38: THE REPRESENTATION OF SC CLUSTERS). In the English-based creole Sranan, /s/-nasal clusters are retained without adaptation, as in *smoke* → [smoko], but /s/-stop clusters are repaired by /s/-deletion: *speak* → [piki] (Alber and Plag 2001). A similar pattern is found in Dutch-based Negerhollands (e.g. *stop* → [tɔp] ~ [stop]; Sabino 1990, cited in Alber and Plag 2001), English-based Krio (e.g. *spoon* → [pun]) and Guyana (e.g. *story* → [tɔri]; Tinelli 1981, cited in Fleischhacker 2005), and an older stage of English-based Belizean Creole (e.g. *skirt* → [koːti]; Greene 1999, cited in Fleischhacker 2005). The only non-creole language where deletion repair is reported to the exclusion of epenthesis for word-initial clusters is Finnish, where all consonants, except for the one immediately preceding the vowel, are deleted (e.g. Swedish *strand* → [ranta] 'shore', Russian *gramatika* → [raːmattu] 'bible'). English loans found in "Finglish" (Finnish as spoken by the Finns who migrated to America) also employ the deletion strategy, as in *street* → [riːti] ~ [triːti] and *blanket* → [lɛnketti] ~ [plɛnketti] (Karttunen 1977). However, in most deletion-only cases, importation of clusters is the prevalent option along with deletion.[14]

[13] According to Smith (2006), deletion repair is marginally found even in word-initial clusters in Japanese, but these forms exist as a part of doublets with a variant with epenthesis.

[14] As Fleischhacker (2005) points out, the deletion of /s/ is potentially problematic for theories of loanword adaptation that appeal to perceptual factors (i.e. the adaptation-as-perception view and the perceptual similarity view), since /s/ is arguably one of the most salient consonants and is argued to be preferentially preserved in most other contexts (cf. Steriade 2001; Shinohara 2006).

Several languages use a deletion strategy in combination with epenthesis word-initially. Polynesian languages that lack /s/ in their inventory delete /s/ in /s/-initial clusters while all other clusters are repaired by epenthesis. Examples include Maori (*frock* → [poraka], *Scotsman* → [koːtimana]; Harlow 2004; Herd 2005) and Tahitian (*president* → [peretiteni], *stocking* → [totini]; Geraghty and Tent 2004a; Herd 2005). In Hawaiian, initial clusters are generally adapted with epenthesis, as in *blessing* → [pele'kine], but for /s/-initial clusters, there is variation between epenthesis and /s/-deletion, as in *speak* → [kə'pikə] ~ ['pikə] (Adler 2006). Adler (2006) attributes the Hawaiian pattern to the fact that /s/ is not a segment of Hawaiian and retaining it requires additional steps of repair in violation of the Threshold Principle of Paradis and LaCharité (1997). However, note that this explanation does not extend to the deletion of /s/ in the creole languages discussed above. For example, in Sranan, the language not only has /s/, but it preferentially preserves /s/ in non-initial position, as in *sister* → [sisa], *[sita], *haste* → [hesi], *[heti] (Alber and Plag 2001).

Other languages that show a combination of deletion and epenthesis strategies preferentially preserve /s/-initial clusters, while obstruent + liquid clusters may be simplified via deletion of C2 (liquid) (cf. Fleischhacker 2005). In Cantonese, deletion and epenthesis repairs are both attested, and the variation is conditioned by the bisyllabic word requirement, as shown in *freezer* → [fiisa] *vs. cream* → [kejlim] (Silverman 1992; Yip 1993). Interestingly, it is only the obstruent + liquid clusters that allow deletion (of the liquid); /s/-initial clusters undergo epenthesis rather than deletion. Mandarin shows a similar asymmetry between cluster types (Miao 2006). Vietnamese is another case where the repair varies between deletion and epenthesis for obstruent + liquid clusters (French *gramme* → [gam] ~ [garam]) (Barker 1969; Andrea Hoa Pham, personal communication).

In Thai, /s/-initial clusters are adapted with epenthesis (*style* → [sataːj]), but obstruent + liquid clusters are often adapted with deletion of the liquid (*plastic* → [pattik]) (Gandour 1979).[15] A similar asymmetry is found in Contemporary Saramaccan (Aceto 1996, cited in Fleischhacker 2005).

Finally, in Telugu, the repair varies between epenthesis and deletion, regardless of the type of cluster, and deletion targets a liquid in obstruent + liquid clusters (*glass* → [gasu] ~ [galasu]) and /s/ in /s/-initial clusters (*station* → [teʃənu] ~ [isteʃənu]) (Broselow 1992, cited in Fleischhacker 2005). Languages that employ deletion repair in word-initial clusters are summarized in Table 100.1.

4.1.2 Word-final coda consonants and clusters

Compared to word-initial position, repairs for coda clusters (e.g. NoCoda or CodaCondition violation) in word-final position are more variable and it is not clear whether epenthesis is cross-linguistically the preferred strategy over deletion in this position (see also chapter 36: final consonants). Examples of languages that choose epenthesis as the repair of choice for word-final coda restrictions include Japanese (Katayama 1998), Kikuyu (Mwihaki 2001), Korean (H. Kang 1996), and Sesotho (Rose and Demuth 2006). On the other hand, there are languages that systematically choose deletion for repair, such as Vietnamese (Barker 1969), Burmese (Chang 2009), Thai (Kenstowicz and Suchato 2006), and White Hmong (Colston

[15] Kenstowicz and Suchato (2006) report that obstruent + liquid clusters are allowed in Thai and therefore remain unadapted.

Table 100.1 Languages that show deletion repair in word-initial clusters

	/s/-C	*Obstruent + Liquid*	*Languages*
Deletion only	**/s/-deletion** ~ importation	*n/a*	Sranan, Negerhollands, Krio, Guyanese Creole, Belizean Creole
	C1 deletion ~ importation	**C1 deletion** ~ importation	(older) Swedish, Russian, and English loans in Finnish
Epenthesis and deletion	**/s/-deletion** ~ epenthesis	Epenthesis (CC > CvC)	Hawaiian, Maori, Tahitian
	Epenthesis (sC > svC)	Epenthesis ~ **C2 deletion**	Cantonese, Mandarin, Vietnamese
	Epenthesis (sC > svC)	Importation ~ **C2 deletion**	Contemporary Sarammacan, Thai
	Epenthesis (sC > vsC) ~ **/s/-deletion**	Epenthesis (CC > CvC) ~ **C2 deletion**	Telugu

and Yang 2001).[16] Some examples from White Hmong include *cake* → [kʰê] and French *Adam* [adam] → [ʔàdà]. Notable is the fact that all languages listed above that choose deletion repair in coda position have a strong preference for monosyllabic morphemes. But note that even these languages do not systematically prefer deletion for onset clusters. For example, Burmese chooses epenthesis for word-initial clusters but debuccalization and deletion for word-final clusters, as in *Christ* → [kʰə.ɾɪʔ] (Chang 2009).

Yet another group of languages show both epenthesis and deletion depending on the segmental composition of the offending structure. These languages include Cantonese (Silverman 1992), Fijian (Schütz 2004; Shinohara 2006; Kenstowicz 2007), Hawaiian (Adler 2006), Inuktitut (Pollard 2008), Hawaiian Japanese (Higa 1970; Smith 2006), Mandarin (Miao 2006), Maori (Harlow 2004), Marshallese (Shinohara 2006), and Yoruba (Shinohara 2006). Examples from Fijian, *east* → [ˈisi], *wolf* → [ˌoˈliva], illustrate the variability of repair patterns.

The question then is what factors contribute to the choice between epenthesis and deletion in coda position. The P-map theory proposes that, all things being equal, consonants with more salient perceptual cues are more likely to be retained than those with less salient cues (Steriade 2009; see Shinohara (2006) and Fleischhacker (2005) in particular for the application of the P-map theory for variable deletion in onset and coda clusters, respectively. See also Miao (2006) for a detailed examination of onset and coda repair patterns for different segmental types in Mandarin).

[16] Golston and Yang (2001) state that "syllable-final consonants are categorically deleted in White Hmong." While the majority of relevant data available in Golston and Yang (2001) indeed show deletion of word-final coda consonants, there are some exceptional forms showing final epenthesis (French [lyk] *Luke* → [lɨ.kà]).

As for grammar-external factors, Smith (2006) demonstrates that English loanwords in Japanese are systematically adapted with epenthesis in the homeland variety, with deletion attested only sporadically, but deletion is much more prevalent in Hawaiian Japanese. She attributes the difference to the different channel of borrowing: the homeland variety is more influenced by the written input and is thus more likely to preserve the input material, while the Hawaiian variety is more influenced by the aural input and is more prone to perceptual adaptation. Miao (2006) proposes a similar explanation to account for why coda deletion is much more frequent in English loanwords than in German loanwords in Mandarin.[17]

The choice in the specific location and the quality of the epenthetic segment also often lacks a straightforward precedent in native phonology (see Fleischhacker 2005 for a P-map-based account of the location of the epenthetic vowel in word-initial clusters). Different strategies for determining the quality of the epenthetic segment include default insertion, vowel harmony, and consonantal assimilation (cf. de Lacy 2006; Uffmann 2006). The specific patterns of vowel epenthesis have been examined in detail in various languages, e.g. Farsi (Shademan 2003), Fijian (Kenstowicz 2007), Hindi (Singh 1985), Inuktitut (Pollard 2008), Kikuyu (Mwihaki 2001), Korean (Oh 1992; Heo 2006; K. Kim 2009), Mandarin (Miao 2006), Cook Island Maori (Kitto and de Lacy 1999), Samoan (Uffmann 2006), Sesotho (Rose and Demuth 2006), Selayarese (Broselow 1999), Shona (Uffmann 2006), Sranan (Uffmann 2006), and Yoruba (Pulleyblank 1988, Akinlabi 1993). Here again, explanations proposed for the choice of epenthetic vowel and its location are very diverse, from the spreading of (contrastively marked) features determined by native phonology (Akinlabi 1993; Pollard 2008; K. Kim 2009) to the emergence of the unmarked (Lombardi 2002; Uffmann 2006), perceptual similarity (Y. Kang 2003; Kenstowicz 2007; Steriade 2008), and epenthesis by misperception (Dupoux *et al.* 1999; Kabak and Idsardi 2007). See CHAPTER 67: VOWEL EPENTHESIS.[18]

4.2 *Co-occurrence restrictions*

Languages may also impose restrictions on possible combinations of vowels and consonants, and such restrictions can induce adaptation of the foreign input. Again, we are faced with the too-many-solutions problem. In the case of segmental adaptation discussed in §3, the features that compete for preservation are all consonantal features or all vocalic features, but in the case of sequential co-occurrence restrictions, vocalic and consonantal features are pitted against each other. Here I will discuss a few such examples.

In Inuktitut, vowels are allophonically retracted next to a uvular consonant, as in /tiriganniq/ [tiriganniq] 'fox' and /pirurtuq/ [pirurtoq] 'fruits' (Denis and Pollard 2008). Therefore, on the surface, a sequence of a retracted vowel and a

[17] Miao (2006) also notes that loans used in commercial brand names are more likely to show deletion than other loans.

[18] Consonantal epenthesis is also a common repair for a vowel-initial word or vowel hiatus in languages that require onsets. Examples of consonantal epenthesis in loanwords are found in Egyptian Arabic (Hafez 1996), Burmese (Chang, forthcoming), Jahai (Burenhult 2001), and White Hmong (Golston and Yang 2001), among others.

non-uvular consonant cannot occur. Thus, when the English input contains illicit sequences such as /ɪk/, there are at least two logically possible repairs to bring the foreign input in line with native phonotactics: change the vowel quality to [ik] or change the place of articulation of the dorsal consonant to [ɪq]. Actual data suggest that the consonantal place remains stable; the vowel quality changes, as in *six* → [siksi], *[sɪqsi] (based on Pollard 2008). This seems like a case of straightforward maintenance of contrastive features of the native language, since in Inuktitut, the distinction between tense and lax vowel is allophonic, but the contrast between velar and uvular places of articulation is phonemic.

French and Spanish loanwords in Moroccan Arabic, on the other hand, show the opposite pattern (Kenstowicz and Louriz 2009). In Moroccan Arabic, the vowels /i a u/ are allophonically lowered and/or retracted to [e ɑ o] when adjacent to emphatic (= pharyngealized) consonants, as in [sif] 'sword' *vs.* [sˤef] 'summer'. In other words, the contrastive distinction in Moroccan Arabic is the emphatic *vs.* non-emphatic distinction in consonants, whereas the vowel-quality difference is an allophonic property. In actual adaptations, mid vowels of French and Spanish are adapted as their corresponding mid vowels in Moroccan Arabic, but with an emphatic version of the adjacent consonants, as in French *taupe* [top] → Moroccan Arabic [tˤobˤbˤ-ɑ] 'rat (FEM)', retaining the allophonic distinction of the native language over a contrastive distinction. The adaptation of the English sequences [ɑn] and [æŋ] in Mandarin, as examined by Hsieh *et al.* (2009), presents a similar pattern. In Mandarin, the backness of a low vowel is predictable from the context; specifically, it is front before /n/, but is back before /ŋ/, a restriction referred to as "rhyme harmony" by Duanmu (2000). English, on the other hand, allows the free combination of low vowels and nasal codas, including [ɑn] and [æŋ], which violates the phonotactic constraints of Mandarin. Surprisingly, in loanword adaptation, the backness of the vowel is preferentially preserved over the place of articulation of the nasal consonants ([ɑn] → [ɑŋ], [æŋ] → [an]), despite the fact that the place feature of a nasal consonant is contrastive and the backness of a low vowel is not in Mandarin. Some representative examples are *monsoon* [mɑnsun] → *mang.xun* [maŋ.ɕyn] and *bank* [bæŋk] → *ban.ke* [pan.kʰɤ]. They attribute this pattern to the perceptual saliency of vowel place contrasts over that of consonantal nasal place contrasts, which are known to be perceptually not very salient (Jun 2004; Steriade 2009).

While the place feature of a consonant seems vulnerable in the Mandarin and the Moroccan Arabic examples, consonantal manner features such as [nasal] seem to be more stable than vowel features. Burmese (Chang 2009) provides a relevant example. In Burmese, only glottal stops are allowed in coda position and coda nasals are adapted as nasalization on the preceding vowel, as in *champagne* → [ʃȁpḗĩ]. However, not all Burmese vowels have a nasalized counterpart. When English words contain a vowel that lacks a nasal counterpart in Burmese, such as [ɛ ɔ i u], followed by a nasal coda, further adaptation is necessary. In such cases, the vowel quality is changed to allow nasalization, rather than failing to preserve nasality, as shown in *November* → [noʊ̀vȉbà], *[noʊ̀vɛ̀bà]. This indicates that preservation of the input nasality takes precedence over preservation of the input vowel quality. However, the limited number of cases examined in the literature does not allow us to conjecture on any cross-linguistic generalizations.

5 Conclusion

The study of loanwords has played an important role in the development of phonological theories in recent years, and loanword phonology presents a rich empirical ground for examining many topical questions in the field of phonology. The issues that the study of loanwords bears upon, directly or indirectly, include the role of output constraints *vs.* processes, the phonetics–phonology interface (more specifically, the role of perceptual factors in shaping phonological patterns), the role of native phonological contrasts in phonological processes, the productivity of stochastic generalizations, and the role of innate *vs.* acquired knowledge.

At the same time, loanword adaptation is conditioned by many extragrammatical factors, such as the role of orthography, the channel of borrowing, the degree of bilingualism, etc. The diverse nature of the contact situation poses an interesting challenge and yet also presents a natural locus of interface between theoretical phonology and sociolinguistics. It has been pointed out that some of the disagreements in the debate on the nature of loanword phonology stem from differing assumptions on what is classified as a loanword (cf. Rose and Demuth 2006); some linguists focus on online adaptations (by monolinguals or bilinguals), while others focus on established loanwords that are sanctioned by norms of the community. While these two endpoints likely exhibit slightly different patterns of adaptation and varying degrees of variability, we expect the two to be related in a systematic way – the output of the initial online adaptation serves as the input for a successive chain of speakers in the rest of the community, eventually leading to the establishment of norms. Some researchers are already addressing the question of how loanwords are transformed over time, in order to provide a more comprehensive and dynamic picture of loanword phonology (Poplack and Sankoff 1984; Crawford 2007, 2008; Davidson 2007; Y. Kang 2008b; Y. Kang *et al.* 2008; Friesner 2009a).

ACKNOWLEDGMENTS

I would like to thank Michael Kenstowicz, Sharon Peperkamp, and Andrea Hoa Pham, as well as the editors and two anonymous reviewers, for their helpful comments and suggestions.

REFERENCES

Aceto, Michael. 1996. Early Saramaccan syllable structure: An analysis of complex onsets from Schumann's 1978 manuscript. *Journal of Pidgin and Creole Languages* 11. 23–44.

Adler, Allison N. 2006. Faithfulness and perception in loanword adaptation: A case study from Hawaiian. *Lingua* 116. 1024–1045.

Ahn, Sang-Cheol & Gregory K. Iverson. 2004. Dimensions in Korean laryngeal phonology. *Journal of East Asian Linguistics* 13. 345–379.

Akinlabi, Akinbiyi. 1993. Underspecification and the phonology of Yoruba /r/. *Linguistic Inquiry* 24. 139–160.

Alber, Birgit & Ingo Plag. 2001. Epenthesis, deletion and the emergence of the optimal syllable in creole: The case of Sranan. *Lingua* 111. 811–840.

Arsenault, Paul. 2009. Coronal features and retroflexion in Dhivehi and other Indo-Aryan languages. *Toronto Working Papers in Linguistics* 30. 17–33.
Barker, Milton. 1969. The phonological adaptation of French loanwords in Vietnamese. *Mon-Khmer Studies Journal* 3. 138–147.
Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.
Best, Catherine T. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In Judith C. Goodman & Howard C. Nusbaum (eds.) *The development of speech perception: The transition from speech sounds to spoken words*, 167–224. Cambridge, MA: MIT Press.
Boersma, Paul & Silke Hamann. 2009. Loanword adaptation as first-language phonological perception. In Calabrese & Wetzels (2009), 11–58.
Brannen, Kathleen. 2002. The role of perception in differential substitution. *Canadian Journal of Linguistics* 47. 1–46.
Broselow, Ellen. 1992. The structure of fricative-stop onsets. Unpublished ms., State University of New York, Stony Brook.
Broselow, Ellen. 1999. Stress, epenthesis, and segment transformation in Selayarese loans. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 25. 311–325.
Broselow, Ellen. 2006. Loanword phonology. In Keith Brown (ed.) *Encyclopedia of language and linguistics*, 2nd edn., vol. 7, 286–290. Oxford: Elsevier.
Broselow, Ellen. 2009. Stress adaptation in loanword phonology: Perception and learnability. In Paul Boersma & Silke Hamann (eds.) *Phonology in perception*, 191–234. Berlin & New York: Mouton de Gruyter.
Brown, Cynthia A. 2000. The interrelation between speech perception and phonological acquisition from infant to adult. In John Archibald (ed.) *Second language acquisition and linguistic theory*, 4–65. Oxford: Blackwell.
Burenhult, Niclas. 2001. Loanword phonology in Jahai. *Lund University Department of Linguistics Working Papers* 48. 5–14.
Cain, Horst. 1986. *A lexicon of foreign loan-words in the Samoan language*. Cologne: Böhlau Verlag.
Calabrese, Andrea. 2009. Perception, production and acoustic inputs in loanword phonology. In Calabrese & Wetzels (2009), 59–114.
Calabrese, Andrea & W. Leo Wetzels (eds.) 2009. *Loan phonology*. Amsterdam & Philadelphia: John Benjamins.
Chang, Charles B. 2009. English loanword adaptation in Burmese. *Journal of the Southeast Asian Linguistics Society* 1. 77–94.
Chang, Charles B. Forthcoming. Phonetics vs. phonology in loanword adaptation: Revisiting the role of the bilingual. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 34.
Clements, G. N. 2001. Representational economy in constraint-based phonology. In T. A. Hall (ed.) *Distinctive feature theory*, 71–146. Berlin & New York: Mouton de Gruyter.
Crawford, Clifford. 2007. The role of loanword diffusion in changing adaptation patterns: A study of coronal stops in Japanese borrowings. *Working Papers of the Cornell Phonetics Laboratory* 16. 32–56
Crawford, Clifford. 2008. An evolutionary account of loanword-induced sound change in Japanese. *University of Pennsylvania Working Papers in Linguistics* 14. 56–80.
Davidson, Lisa. 2007. The relationship between the perception of non-native phonotactics and loanword adaptation. *Phonology* 24. 261–286.
Davidson, Lisa, Peter W. Jusczyk & Paul Smolensky. 2004. The initial and final states: Theoretical implications and experimental explorations of Richness of the Base. In Kager *et al.* (2004), 321–368.
Davidson, Lisa & Rolf Noyer. 1997. Loan phonology in Huave: Nativization and the ranking of faithfulness constraints. *Proceedings of the West Coast Conference on Formal Linguistics* 15. 65–79.

Davis, Stuart & Mee-Hui Cho. 2006. Phonetics versus phonology: English word final /s/ in Korean loanword phonology. *Lingua* 116. 1008–1023.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.

Denis, Derek & Mark Pollard. 2008. An acoustic analysis of the vowel space of Inuktitut: Evidence for allophony. Paper presented at Inuktitut Linguistics Workshop, University of Toronto.

Detey, Sylvain & Jean-Luc Nespoulous. 2008. Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua* 118. 66–81.

Dohlus, Katrin. 2005. Phonetics or phonology: Asymmetries in loanword adaptations – French and German mid front rounded vowels in Japanese. *ZAS Papers in Linguistics* 42. 117–135.

Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.

Duanmu, San. 2000. *The phonology of Standard Chinese*. Oxford: Oxford University Press.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25. 1568–1578.

Dupoux, Emmanuel, Christopher Pallier, Núria Sebastián-Gallés & Jacques Mehler. 1997. A destressing "deafness" in French? *Journal of Memory and Language* 36. 406–421.

Fleischhacker, Heidi. 2005. Similarity in phonology: Evidence from reduplication and loan adaptation. Ph.D. dissertation, University of California, Los Angeles.

Friesner, Michael L. 2009a. The social and linguistic predictors of the outcomes of borrowing in the speech community of Montreal. Ph.D. dissertation, University of Pennsylvania.

Friesner, Michael L. 2009b. The adaptation of Romanian loanwords from Turkish and French. In Calabrese & Wetzels (2009), 115–130.

Gandour, Jackson. 1979. Tonal rules for English loanwords in Thai. In T. L. Thongkum, P. Kullavanijaya, P. V. Panupong & K. Tingsabadh (eds.) *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugénie J. A. Henderson*, 94–105. Bangkok: Chulalongkorn University Press.

Gbéto, Flavien. 2000. *Les emprunts linguistiques d'origine européenne en Fon (Nouveau Kwa, Gbe: Bénin)*. Cologne: Rüdiger Köppe Verlag.

Geraghty, Paul & Jan Tent. 2004a. From lowlands to islands: Dutch loans in Polynesia. In Geraghty & Tent (2004b), 99–143.

Geraghty, Paul & Jan Tent (eds.) 2004b. *Borrowing: A Pacific perspective*. Canberra: Australian National University.

Golston, Chris & Phong Yang. 2001. White Hmong loanword phonology. In Caroline Féry, Antony Dubach Green & Ruben van de Vijver (eds.) *Proceedings of HILP* 5, 40–57. Potsdam: University of Potsdam.

Greene, Laurie A. 1999. *A grammar of Belizean Creole: Compilations from two existing United States dialects*. New York: Peter Lang.

Hafez, Ola. 1996. Phonological and morphological integration of loanwords into Egyptian Arabic. *Égypte/Monde Arabe* (première série) 27–28. 383–410.

Hancin-Bhatt, Barbara Jean. 1994. Segment transfer: A consequence of a dynamic system. *Second Language Research* 10. 241–269.

Harlow, Ray. 2004. Borrowing and its alternatives in Māori. In Geraghty & Tent (2004b), 145–169.

Haugen, Einar. 1950. The analysis of linguistic borrowing. *Language* 26. 210–231.

Heffernan, Kevin. 2007. The role of phonemic contrast in the formation of Sino-Japanese. *Journal of East Asian Linguistics* 16. 61–86.

Heo, Younghyon. 2006. Postpalatal epenthesis in Korean and Japanese: The role of contrast. In William O'Grady, Jong-Yurl Yoon, Jae-Young Han, Ponghyung Lee, Hae-Yeon Kim, Sung-Hoon Hong, Seunghye Chung & Fernando Torres (eds.) *Inquiries into Korean Linguistics II*, 157–165. Seoul: Taehaksa.

Herd, Jonathon. 2005. Loanword adaptation and the evaluation of similarity. *Toronto Working Papers in Linguistics* 24. 65–116.

Higa, Masanori. 1970. The sociolinguistic significance of borrowed words in the Japanese spoken in Hawaii. *University of Hawaii Working Papers in Linguistics* 2. 125–140.

Holden, Kyril. 1976. Assimilation rates of borrowings and phonological productivity. *Language* 52. 131–147.

Hsieh, Feng-fan. 2006. High infidelity: The non-mapping of Japanese accent onto Taiwanese tone. *MIT Working Papers in Linguistics* 52. 1–27.

Hsieh, Feng-fan & Michael Kenstowicz. 2008. Phonetic knowledge in tonal adaptation: Mandarin and English loanwords in Lhasa Tibetan. *Journal of East Asian Linguistics* 17. 279–297.

Hsieh, Feng-fan, Michael Kenstowicz & Xiaomin Mou. 2009. Mandarin adaptations of coda nasals in English loanwords. In Calabrese & Wetzels (2009), 131–154.

Hyman, Larry M. 1970. The role of borrowing in the justification of phonological grammars. *Studies in African Linguistics* 1. 1–48.

Ito, Chiyuki & Michael Kenstowicz. 2009. Mandarin loanwords in Yanbian Korean I: Laryngeal features. *Onin Kenkyuu* 12. 61–72.

Ito, Chiyuki, Yoonjung Kang & Michael Kenstowicz. 2006. The adaptation of Japanese loanwords into Korean. *MIT Working Papers in Linguistics* 52. 65–104.

Itô, Junko & Armin Mester. 1995. Japanese phonology. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 817–838. Cambridge, MA & Oxford: Blackwell.

Itô, Junko & Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.) *The handbook of Japanese linguistics*, 62–100. Malden, MA & Oxford: Blackwell.

Itô, Junko & Armin Mester. 2001. Covert generalizations in Optimality Theory: The role of stratal faithfulness constraints. *Studies in Phonetics, Phonology, and Morphology* 7. 273–299.

Iverson, Gregory K. & Ahrong Lee. 2006. Perception of contrast in Korean loanword adaptation. *Korean Linguistics* 13. 49–87.

Jacobs, Haike & Carlos Gussenhoven. 2000. Loan phonology: Perception, salience, the lexicon and OT. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: Phonology, syntax, and acquisition*, 193–210. Oxford: Oxford University Press.

Jun, Jongho. 2004. Place assimilation. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 58–86. Cambridge: Cambridge University Press.

Kabak, Baris & William J. Idsardi. 2007. Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech* 50. 23–52.

Kager, René, Joe Pater & Wim Zonneveld (eds.) 2004. *Constraints in phonological acquisition*. Cambridge: Cambridge University Press.

Kang, Hyunsook. 1996. English loanwords in Korean. *Studies in Phonetics, Phonology, and Morphology* 2. 21–47.

Kang, Hyunsook. 2006. An acoustic study of the perceptual significance of F2 transition of /w/ in English and Korean. *Speech Sciences* 13. 7–21.

Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20. 219–273.

Kang, Yoonjung. 2008a. Interlanguage segmental mapping as evidence for the nature of lexical representation. *Language and Linguistics Compass* 2. 85–100.

Kang, Yoonjung. 2008b. Tensification of voiced stops in English loanwords in Korean. *Harvard Studies in Korean Linguistics* 12. 179–192.

Kang, Yoonjung. 2008c. The adaptation of English /s/ in Korean. In John Whitman, Ponghyung Lee, Chang-Kook Suh, Sungdai Cho, Yeun-Jin Jung, Jae Young Chung &

Jeong-Im Han (eds.) *Inquiries into Korean Linguistics III*, 1–14. Seoul: Hankook Munhwasa.

Kang, Yoonjung. 2009. English /z/ in 1930s Korean. *Simon Fraser University Working Papers in Linguistics* 2.

Kang, Yoonjung. 2010. Tutorial overview: Suprasegmental adaptation in loanwords. *Lingua* 120. 2295–2310.

Kang, Yoonjung, Michael Kenstowicz & Chiyuki Ito. 2008. Hybrid loans: A study of English loanwords transmitted to Korean via Japanese. *Journal of East Asian Linguistics* 17. 299–316.

Karttunen, Frances. 1977. Finnish in America: A case study in monogenerational language change. In Ben Blount & Mary Sanches (eds.) *The social dimensions of language change*, 173–184. New York: Academic Press.

Karvonen, Daniel. 2005. Word prosody in Finnish. Ph.D. dissertation, University of California, Santa Cruz.

Katayama, Motoko. 1998. Optimality Theory and Japanese loanword phonology. Ph.D. dissertation, University of California, Santa Cruz.

Kato, Sachiko. 2006. Loanword adaptation in Japanese: Consonant gemination of English loanwords. *MIT Working Papers in Linguistics* 52. 105–137.

Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82. 536–574.

Kawahara, Shigeto. 2008. Phonetic naturalness and unnaturalness in Japanese loanword phonology. *Journal of East Asian Linguistics* 17. 317–330.

Kenstowicz, Michael. 2003. The role of perception in loanword phonology: A review of *Les emprunts linguistiques d'origine européenne en Fon*. *Studies in African Linguistics* 32. 95–112.

Kenstowicz, Michael. 2005. The phonetics and phonology of loanword adaptation. In S.-J. Rhee (ed.) *Proceedings of ECKL 1: Proceedings of 1st European Conference on Korean Linguistics*, 17–32. Seoul: Hankook Publishing Co.

Kenstowicz, Michael. 2007. Salience and similarity in loanword adaptation: A case study from Fijian. *Language Sciences* 29. 316–340.

Kenstowicz, Michael & Nabila Louriz. 2009. Reverse engineering: Emphatic consonants and the adaptation of vowels in French loanwords into Moroccan Arabic. *Brill's Annual of Afroasiatic Languages and Linguistics* 1. 41–74.

Kenstowicz, Michael & Hyang-Sook Sohn. 2001. Accentual adaptation in North Kyungsang Korean. In Michael Kenstowicz (ed.) *Ken Hale: A life in language*, 239–270. Cambridge, MA: MIT Press.

Kenstowicz, Michael & Atiwong Suchato. 2006. Issues in loanword adaptation: A case study from Thai. *Lingua* 116. 921–949.

Kertész, Zsuzsa. 2003. Vowel harmony and the stratified lexicon of Hungarian. *The Odd Yearbook* 7. 62–77.

Kim, Hyunsoon. 2008. Loanword adaptation between Japanese and Korean: Evidence for L1 feature-driven perception. *Journal of East Asian Linguistics* 17. 331–346.

Kim, Hyunsoon. 2009. Korean adaptation of English affricates and fricatives in a feature-driven model of loanword adaptation. In Calabrese & Wetzels (2009), 155–180.

Kim, Kyumin. 2009. Coronals in epenthesis in loanwords in comparison to velars. *Toronto Working Papers in Linguistics* 30. 53–67.

Kim, Soohee & Emily Curtis. 2002. Phonetic duration of English /s/ and its borrowing into Korean. In Noriko M. Akatsuka & Susan Strauss (eds.) *Japanese/Korean Linguistics* 10, 406–419. Palo Alto: CSLI.

Kitto, Catherine & Paul de Lacy. 1999. Correspondence and epenthetic quality. In Catherine Kitto & Carolyn Smallwood (eds.) *Proceedings of AFLA VI: The Sixth Meeting of the Austronesian Formal Linguistics Association*, 181–200. Toronto: Department of Linguistics, University of Toronto.

Kubozono, Haruo. 2006. Where does loanword prosody come from? A case study of Japanese loanword accent. *Lingua* 116. 1140–1170.
Kubozono, Haruo, Junko Ito & Armin Mester. 2008. Consonant gemination in Japanese loanword phonology. Paper presented at the 18th International Congress of Linguistics, Seoul.
LaCharité, Darlene & Carole Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry* 36. 223–258.
Lee, Dongmyung. 2009. The loanword tonology of South Kyungsang Korean. Ph.D. dissertation, Indiana University.
Lehiste, Ilse. 1988. *Lecture notes on language contact*. Cambridge, MA: MIT Press.
Lin, Yen-Hwei. 2008. Variable vowel adaptation in Standard Mandarin loanwords. *Journal of East Asian Linguistics* 17. 363–380.
Lombardi, Linda. 2002. Coronal epenthesis and markedness. *Phonology* 19. 219–251.
Luke, Kang-kwong & Chaakming Lau. 2008. On loanword truncation in Cantonese. *Journal of East Asian Linguistics* 17. 347–362.
Massaro, Dominic W. & Michael M. Cohen. 1983. Phonological context in speech perception. *Perception and Psychophysics* 34. 338–348.
Miao, Ruiqin. 2006. Loanword adaptation in Mandarin Chinese: Perceptual, phonological and sociolinguistic factors. Ph.D. dissertation, State University of New York, Stony Brook.
Mwihaki, Alice. 2001. Consonant–vowel harmony: Evidence from the phonotactics of loanword adaptation. *Poznan Studies in Contemporary Linguistics* 37. 139–145.
Oh, Mira. 1992. Palatal consonants, labial consonants, and vowel epenthesis in Korean loanword phonology. *Linguistic Journal of Korea* 17. 141–161.
Oh, Mira. 2004. English stop adaptation as output-to-output correspondence. *Onin Kenkyuu* 7. 165–172.
Paradis, Carole. 1988. Towards a theory of constraint violations. *McGill Working Papers in Linguistics* 5. 1–43.
Paradis, Carole. 2006. The unnatural /Cju/ (< foreign /Cy/) sequence in Russian loanwords: A problem for the perceptual view. *Lingua* 116. 976–995.
Paradis, Carole & Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33. 379–430.
Paradis, Carole & Darlene LaCharité. 2008. Apparent phonetic approximation: English loanwords in Old Quebec French. *Journal of Linguistics* 44. 87–128.
Paradis, Carole & Darlene LaCharité. 2009. English loanwords in Old Quebec French: Fewer bilinguals does not mean a great increase in naive phonetic approximation. *Langues et linguistique* 32. 82–117.
Paradis, Carole & Antoine Tremblay. 2009. Nondistinctive features in loanword adaptation: The unimportance of English aspiration in Mandarin Chinese phoneme categorization. In Calabrese & Wetzels (2009), 211–224.
Peperkamp, Sharon. 2005. A psycholinguistic theory of loanword adaptation. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 30. 341–352.
Peperkamp, Sharon & Emmanuel Dupoux. 2003. Reinterpreting loanword adaptations: The role of perception. In M. J. Solé, D. Recasens & J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*, 367–370. Barcelona: Causal Productions.
Peperkamp, Sharon, Inga Vendelin & Kimihiro Nakamura. 2008. On the perceptual origin of loanword adaptations: Experimental evidence from Japanese. *Phonology* 25. 129–164.
Pollard, Mark G. 2008. Epenthetic vowel quality in English loanwords in Inuktitut. M.A. thesis, University of Toronto.
Poplack, Shana & David Sankoff. 1984. Borrowing: The synchrony of integration. *Linguistics* 22. 99–135.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Pulleyblank, Douglas. 1988. Vocalic underspecification in Yoruba. *Linguistic Inquiry* 19. 233–270.
Repetti, Lori. 1993. The integration of foreign loans in the phonology of Italian. *Italica* 70. 182–196.
Repetti, Lori. 2006. The emergence of marked structures in the integration of loans in Italian. In Randall Gess & Deborah Arteaga (eds.) *Historical Romance linguistics: Retrospective and perspectives*, 209–235. Amsterdam & Philadelphia: John Benjamins.
Repetti, Lori. 2009. Gemination in English loans in American varieties of Italian. In Calabrese & Wetzels (2009), 225–240.
Rose, Yvan & Katherine Demuth. 2006. Vowel epenthesis in loanword adaptation: Representational and phonetic considerations. *Lingua* 116. 1112–1139.
Sabino, Robin. 1990. Towards a phonology of Negerhollands: An analysis of phonological variation. Ph.D. dissertation, University of Pennsylvania.
Schütz, Albert J. 1978. English loanwords in Fijian. In Albert J. Schütz (ed.) *Fijian language studies: Borrowing and pidginization*, 1–50. Suva: Fiji Museum.
Schütz, Albert J. 2004. English loanwords in Fijian. In Geraghty & Tent (2004b), 253–294.
Shademan, Shabnam. 2003. Epenthetic vowel harmony in Farsi. M.A. thesis, University of California, Los Angeles.
Shinohara, Shigeko. 1997. Analyse phonologique de l'adaptation japonaise de mots étrangers. Thèse de doctorat, Université Paris III.
Shinohara, Shigeko. 2000. Default accentuation and foot structure in Japanese: Evidence from Japanese adaptations of French words. *Journal of East Asian Linguistics* 9. 55–96.
Shinohara, Shigeko. 2004. Emergence of Universal Grammar in foreign word adaptations. In Kager *et al.* (2004), 292–320.
Shinohara, Shigeko. 2006. Perceptual effects in final cluster reduction patterns. *Lingua* 116. 1046–1078.
Silverman, Daniel. 1992. Multiple scansions in loanword phonology: Evidence from Cantonese. *Phonology* 9. 289–328.
Singh, Rajendra. 1985. Prosodic adaptation in interphonology. *Lingua* 67. 269–282.
Smith, Jennifer L. 2006. Loan phonology is not all perception: Evidence from Japanese loan doublets. In Timothy J. Vance & Kimberly A. Jones (eds.) *Japanese/Korean Linguistics* 14, 63–74. Palo Alto: CLSI.
Smolensky, Paul. 1996. The initial state and "Richness of the Base" in Optimality Theory. Unpublished ms., Johns Hopkins University, Baltimore (ROA-154).
Sridhar, Shikaripur N. 1990. *Kannada*. London: Routledge.
Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 219–250. San Diego: Academic Press.
Steriade, Donca. 2007. Perceptual repair and syllable structure: A reply to Kabak and Idsardi. Unpublished ms., MIT.
Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Studies in honor of Paul Kiparsky*, 151–180. Cambridge, MA: MIT Press.
Takahashi, Shoichi. 2006. Topics in Palauan loanword phonology. *MIT Working Papers in Linguistics* 52. 165–200.
Tinelli, Henri. 1981. *Creole phonology*. The Hague: Mouton.
Uffmann, Christian. 2006. Epenthetic vowel quality in loanwords: Empirical and formal issues. *Lingua* 116. 1079–1111.
Vendelin, Inga & Sharon Peperkamp. 2006. The influence of orthography on loanword adaptations. *Lingua* 116. 996–1007.
Walter, Mary Ann. 2006. Grammatical gender via lexical statistics: The case of Arabic-to-Spanish loanwords. *MIT Working Papers in Linguistics* 52. 201–225.

Werker, Janet F. & Richard C. Tees. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75. 1866–1878.

Wetzels, W. Leo. 2009. Nasal harmony and the representation of nasality in Maxacalí: Evidence from Portuguese loans. In Calabrese & Wetzels (2009), 241–270.

Wu, Hsiao-hung Iris. 2006. Stress to tone: A study of tone loans in Mandarin Chinese. *MIT Working Papers in Linguistics* 52. 227–253.

Yip, Moira. 1993. Cantonese loanword phonology and Optimality Theory. *Journal of East Asian Linguistics* 2. 261–291.

Yip, Moira. 2002. Necessary but not sufficient: Perceptual influences in loanword phonology. *Journal of the Phonetic Society of Japan* 6. 4–21.

Yip, Moira. 2006. The symbiosis between perception and grammar in loanword phonology. *Lingua* 116. 950–975.

Zuraw, Kie. 2000. Patterned exceptions in phonology. Ph.D. dissertation, University of California, Los Angeles.

6 Self-organization in Phonology

Andrew Wedel

Structure can arise in a system in many different ways. Self-organization is one general mechanism for structure formation which has relatively recently been explored as a possible contributor to patterns found in language. The aims of this chapter are: (i) to provide an overview of self-organization as a general mechanism for structure formation; (ii) to describe some of the ways that self-organizational processes can interact with other familiar mechanisms for structure formation; and (iii) to review selected work done to date which argues that particular phonological phenomena may arise through the contribution of self-organizational mechanisms.

Self-organization has been argued to play a role in a wide variety of phonological phenomena, from the development of a phonological grammar in acquisition through the systemic grammatical changes that occur in language over many generations. A central mechanism in self-organization is feedback, in which the properties of a current state of a system are dependent in some way on those of a preceding state. When such states are linked in a temporal chain, structure can develop in ways that cannot be described in terms of a single set of causal steps. Arguments that phonological patterns can emerge in this way are not new: Lindblom *et al.* (1984) for example argued that phonological inventories are shaped by interactions between phonological categories, mediated by constraints on effort and perceptibility rather than through, for example, innate constraints on features and their combinations (see CHAPTER 17: DISTINCTIVE FEATURES). In the last decade, however, the scope of self-organization-based accounts has rapidly widened to include phonological acquisition, evolution of complex grammatical patterns (CHAPTER 101: THE INTERPRETATION OF PHONOLOGICAL PATTERNS IN FIRST LANGUAGE ACQUISITION), evolution and maintenance of phonological contrast (CHAPTER 2: CONTRAST), grammatical "conspiracies" (CHAPTER 70: CONSPIRACIES), and many other phenomena. A representative set of these accounts will be discussed in detail below in §3, but it will be helpful to begin with a general introduction to key concepts in self-organization as a general structure formation mechanism. Supplementary simulations illustrating these concepts are included in the online version of this chapter.

1 Self-organization as a pathway to structure

"Self-organization" is not a concept with crisp edges. Instead, it is a big-tent term covering the many ways structure can form in *non-linear, dissipative* systems (Kauffman 1995; Ball 1999; Camazine *et al.* 2001; Heylighen 2001). Non-linear systems are those in which the properties of the system as a whole cannot be understood in terms of the properties of individual system elements, in other words, those in which new properties emerge through interaction. Cooked egg white is a familiar example of a non-linear system. Egg white consists primarily of the protein albumin, and prior to cooking the globular albumin molecules slide past one another easily, producing a translucent semi-liquid. When the temperature is raised beyond a certain point, the albumin protein chains unfold and stick to each other, creating a large, highly interlocked structure that is opaque and stiff. These properties of cooked egg white cannot be understood in terms of the summed properties of individual unfolded albumin molecules, but from their interaction as they stick together. Likewise, many features of language are dependent on the interaction of elements at some lower level of description for their existence, most obviously in that language change arises through language use and transmission across generations.

Dissipative systems are those in which a given state or structure is maintained through a constant flux of energy or matter. As a consequence, an account of a structure within a dissipative system includes time at some level. A ripple in a creek provides a familiar example of a higher order structure produced through a flux at a lower level of description. At one level of description, a ripple is an independent element of a creek that can interact with other elements at that level, such as a floating leaf or another ripple. At a lower level of description, it is a vast and constantly changing set of water molecules interacting with each other and the creek bed as they move. If the flow of water stops, the ripple disappears. Any given language is dissipative in the sense that it is instantiated through usage and transmission, just as a ripple is instantiated through flow. As we will see in the examples below in §3, self-organizational accounts of structure formation in language depend on cycles of use and/or acquisition.

Structure arises in non-linear, dissipative systems when many similar elements or events interact over time to produce persistent changes at some higher level of organization. Typically, structure formation in self-organizing systems is the result of positive and negative feedback loops engendered by the interaction of system elements with each other or with the environment. Positive feedback (also sometimes referred to as autocatalysis) arises when a given event makes a similar event more likely in the future. An example is the population growth that occurs when individuals have offspring at greater than the replacement rate. In this case, the birth of each additional individual makes a subsequent birth more likely. Positive feedback promotes change and can result in runaway processes. As an example of a potentially central role of positive feedback in language, I have argued that positive feedback in the form of similarity bias in production and perception may drive the development of coherent grammatical patterns despite storage of low-level phonotactic detail in the lexicon (CHAPTER 1: UNDERLYING REPRESENTATIONS), which should otherwise promote lexical idiosyncrasy (Wedel 2007).

Negative feedback arises when an event makes a similar event less likely in the future, as when a growing population outstrips its supply of resources. In this case, each additional birth *lowers* the probability of a subsequent birth through increased competition. Negative feedback promotes stability. Both positive and negative feedback represent types of non-linearity, because the description of patterns resulting from feedback must make reference to interactions between system elements. Self-organization often occurs in systems through positive feedback between system-internal elements that is prevented from snowballing beyond a certain point by negative feedback. Examples of this sort include population growth limited by finite resources, thunderstorm structure in which a growing updraft is constrained by a resulting downdraft, and economic bubbles, burst by the collapse of credit. In §3, we will see several examples of the ways negative feedback may inhibit loss of phonemic contrast over the course of language change (CHAPTER 93: SOUND CHANGE).

Self-organized systems frequently exhibit *emergence*. Emergence in this context refers to the generation of a higher order structure that interacts meaningfully with other structures of the system at this level of description. Our earlier example of a ripple in a stream serves as a familiar instance of emergence: the influence of a ripple on other system elements (a leaf, another ripple) is most usefully described in terms of our understanding of the behavior of ripples, rather than our understanding of water molecules. Likewise in language, for example, many phonological patterns can be described in terms of the interaction of (possibly conflicting) phonotactic generalizations or constraints. In models in which phonotactic generalizations are derived over the course of acquisition and usage from the lexical items that instantiate them, these generalizations are emergent (see e.g. Blevins 2004; Wedel 2006; Mielke 2008).

Finally, self-organized systems frequently exhibit phase transitions between semi-stable states defined by *attractors*. An attractor is a system state (or set of states) that nearby states tend to evolve toward. A simple visual metaphor for a system with multiple attractors is a surface with multiple basins. If a ball is placed somewhere on this surface, it will tend to roll to the bottom of whatever basin it happens to be in. Phase transitions correspond to the transition from one basin of attraction to another and are accompanied by a shift in the behavior of a system. In our visual analogy, if we begin to shake the surface, the ball will begin to bounce around within its basin and may eventually by chance roll up and over into a new basin, where it remains until it again rolls up and over into a new basin. Within the domain of morphology, local similarity bias in the form of analogical extension has been argued to create attractors that influence the course of morphological change over time (see e.g. Hock 2003; Garrett 2008). Likewise, pockets of formally similar irregulars ("gangs") have been shown to be more likely to recruit new members than formally isolated irregulars (Bybee and Moder 1983; Stemberger and MacWhinney 1988). Under this general model, coherent generalizations over forms act as emergent attractors. These patterns of similarity-based extension and, plausibly, resistance to extension are consistent with a model in which local similarity effects play a significant role in the formation of larger-scale morphological regularities. For a detailed implementation of this type of model simulating the evolution of past tense forms in Old English, see Hare and Elman (1995). In a similar fashion, I have argued that similarity biases at the level of sound categories may underlie the development of regular patterns in phonology (Wedel 2007), as well as the outcome of conflicts between phonological and morphological regularities (Wedel 2009); (see also CHAPTER 87: NEIGHBORHOOD EFFECTS).

"The Game of Life" (J. H. Conway, reported in Gardner 1970) provides a simple example of a deterministic, self-organized system that exhibits all these properties (examples of The Game of Life can be found on the web; see also the online version of this chapter). The Game of Life is a simple cellular automaton that occupies an infinite, two-dimensional orthogonal grid, the cells of which can be either "alive" or "dead." There are three simple rules governing cell birth and death, each of which makes reference to a cell's eight immediate neighbors: (i) if a living cell has fewer than two living neighbors, it dies; (ii) if a living cell has more than three living neighbors, it dies; (iii) if a dead cell has exactly three living neighbors, it becomes alive.

The grid is initialized with some seed pattern of living cells, and then left to evolve according to these three rules. Some seed patterns result in uninteresting outcomes: if the distribution of living cells is too sparse, all cells quickly die; conversely, some seed patterns are stable and do not change even though the rules continue to be applied. (Four cells arranged in a square is one example of such a stationary pattern.) Other seed patterns produce oscillating structures or structures that move in a consistent direction across the grid. "Gosper's Glider Gun" is a particularly beautiful example of the complexity that can arise through the interaction of these simple rules over time. (A movie of Gosper's Glider Gun can be found on the web or in the online version of this chapter.)

The Game of Life exhibits many of the typical features of self-organizing systems. Structure formation depends on the interaction between elements (it is non-linear) and the application of cell birth and death rules over time (it is dissipative). It also requires the interaction of context-dependent positive and negative feedback processes: depending on the local context, the birth of a cell can cause the birth of additional cells in the next round or it can cause death. As in many self-organizing systems, structure arises in The Game of Life through positive feedback that is held in check by negative feedback. Finally, this system exhibits emergence, in which distinct groupings of living cells function as units with predictable behavior. In Gosper's Glider Gun, for example, two large groupings of cells bounce off of stationary square groupings at the edges of the system, and then bounce off of each other. In the process of bouncing off of each other, they create small self-contained "gliders" that embark on an infinite journey away from the center.

2 Self-organization in interaction with other influences on structure

Self-organization does not operate in a vacuum. It contributes structure in a context supplied by the system and its environment, and the properties of the system that support and direct the emergence of new structure can have any source, including innate pre-specification. For example, features of the environment can supply negative feedback or serve as templates that give initial direction to self-organized structure. In The Game of Life, self-organized structure formation is dependent on the properties of the environment (the orthogonal grid), and on the predetermined character of interactions between cells. The structures that develop are also critically dependent on the initial seed pattern of living and dead cells which serves as an organizing template.

From a design point of view, self-organization is a powerful tool: if the details of a complex structure can be constructed through emergence instead of direct specification by some other means, it can be encoded considerably more compactly than otherwise possible (Gell-Mann 1992). For example, the specification of Gosper's Glider Gun requires only a description of the environment, the rules for cell birth and death, and the initial seed pattern. Further, different complex structures can be created by minimal changes to this description, such as the properties of the environment, the seed pattern, or the rules for cell birth and death.

Similarly, many biological structures are thought to emerge from self-organizational pathways which are given shape and direction by innately specified contexts. For example, the spots and stripes that are found throughout the animal kingdom have been proposed to emerge through a single basic system with slight variations involving the competition between diffusing activator and inhibitor molecules in an animal's skin. (This model was originally proposed by Alan Turing in 1950; for discussion see Ball 1999: ch. 4.) Different shapes and patterns of coat markings can be produced simply by changing the relative diffusion speeds of the activator and inhibitor, or by changing the shapes of the underlying pigment-producing cells. This is a much more informationally compact way to produce a complex coat pattern than, for example, specifying the state of each and every individual pigment-producing cell. Furthermore, patterns that arise through self-organizational pathways are often very robust to perturbation, because of the role that attractors play in the evolution of the system. Chain shifts are possible linguistic examples of the ability of self-organized patterns to survive perturbation, i.e. to persist despite change at some other level in the system (discussed in §3.6 below; see also Gordon 2002; Wedel 2006; Ettlinger 2007). For an overview of some of the many biological patterns that are thought to arise through self-organization, including examples of the interaction between self-organizational pathways and other mechanisms, see Camazine *et al.* (2001).

There are many resources available on the web to learn more about self-organization and related concepts. Excellent published resources include Kauffman (1995), Ball (1999), Camazine *et al.* (2001), and Heylighen (2001).

3 Self-organization as a pathway for structure formation in language

Much recent work approaches language as a complex adaptive system in which grammatical patterns are emergent properties resulting from the repeated interaction of the many different elements that make up a larger language system: innate and acquired biases; forms at multiple levels of representation; interacting spheres of use; sociolinguistic networks; and chains of acquisition and transmission over longer time-scales. For representative examples, see Haspelmath (1999); Nettle (1999); Plaut and Kello (1999); de Boer (2000); Browman and Goldstein (2000); Croft (2000); Lindblom (2000); Bybee (2001); Kirby (1999); Oudeyer (2002); Bod *et al.* (2003); Blevins (2004); Wedel (2007); Boersma and Hamann (2008); Kirby *et al.* (2008); Mielke (2008); Blevins and Wedel (2009).

A priori, there are at least two plausible reasons to think that self-organizational pathways may contribute to some of the patterns we find in language. The first is simply that at many levels and time-scales, language provides the necessary conditions to support spontaneous emergence of patterns through self-organizational pathways (see e.g. Lindblom *et al.* 1984; Ohala 1989; Lindblom 1992; Keller 1994; Labov 1994; Cziko 1995; Dennet 1995; Elman 1995; Deacon 1997; Cooper 1999; Hurford 1999; Steels 2000; Bybee 2001; Blevins 2004; MacWhinney 2006; Pisoni and Levi 2007; Beckner *et al.* 2009; and many others). Language involves the repeated interaction of many similar elements, in similar ways, at many levels of description and time-scales. Variation and bias in language acquisition and use provide many potential feedback pathways. Basic language-external constraints, from articulatory and perceptual factors through general categorization mechanisms to cross-culturally common salience relationships, all provide structures and templates that could give common shape to self-organized patterns. Because structure tends to emerge spontaneously under these conditions, it would be surprising if self-organizational pathways do not contribute to the formation of some of the many observed patterns in language, whether language-particular or crosslinguistically frequent. Put another way, if we found that self-organizational mechanisms played no role in the emergence of any observed language patterns, our burden would be to explain why not.

The second, perhaps less compelling reason derives from design principles. As briefly reviewed above, the specification of a complex pattern can be much more compact and the resulting structure more robust to perturbation when created through self-organizational pathways. To the extent that the language faculty has evolved under functional constraints for use and that grammars continue to do so diachronically, self-organization represents a powerful mechanism for structure in the "blind watchmaker's toolbox" (Dawkins 1986).

3.1 *Self-organization in phonology*

Explorations of self-organizational accounts for linguistic patterns can be understood in terms of the general scientific goal of explaining more with less. Just as a good Optimality Theory (Prince and Smolensky 1993) account attempts to explain new patterns through rankings of existing constraints, an account of linguistic patterns making use of a self-organizational pathway attempts to explain a complex pattern through the interaction of simpler independent mechanisms. Often, authors of these accounts argue that a linguistic pattern emerges through the interaction of domain-general factors rather than through innate grammatical mechanisms, but it is worth emphasizing that this is not a necessary feature of a self-organizational account. Just as a self-organized biological pattern may arise from innately specified processes, a self-organized linguistic pattern could arise from simpler language-specific structures. Some models have been framed in these terms under the rubric of "biolinguistics"; see e.g. Medeiros (2008).

Many self-organizational accounts make use of computational simulation, either as an existence proof that a given structure can arise through interactions between some defined set of system properties, and/or as a supporting illustration for verbal or analytic arguments. Simulation is particularly useful in this context, because self-organization proceeds through chains of circular causation progressively building structure over time. As a consequence, verbal descriptions of proposed self-organizational processes are often hard to assess critically. More importantly, interacting feedback loops are notorious for producing counter-intuitive results, so a computational implementation of a model provides both an important research tool for a theorist and a demonstration for a reader that the model operates as expected (Peck 2004). The following is a brief survey of several phenomena in phonology that have been proposed to arise in part through self-organizational pathways. This is of course just a sample of the many insightful self-organization-based models of phonological phenomena, and interested readers are encouraged to explore the literature further.

3.2 *Early phonological acquisition*

In early phonological acquisition (CHAPTER 101: THE INTERPRETATION OF PHONOLOGICAL PATTERNS IN FIRST LANGUAGE ACQUISITION), initial relatively accurate word imitation is followed by a period of less accurate, but more systematic productions (Ferguson and Farwell 1975, reviewed in Vihman *et al.* 2009). This is reminiscent of the U-shaped learning curve of irregular morphological forms in which irregulars are initially reproduced faithfully, followed by a period of over-regularization, followed in turn by increasingly accurate production. Further, while these production patterns are consistent for a given child, they differ across children, suggesting that the pathway to phonological competence is not prespecified at this level (Beckman and Edwards 2000; Vihman and Croft 2007; Vihman *et al.* 2009). Vihman *et al.* (2009) propose that this phenomenon can be explained in a model based on the ability of infants to acquire individual word gestalts, combined with an ability to generalize over those gestalts through feedback from their own production (see also Pierrehumbert 2003 for related discussion). Under this model, an infant's set of practiced babbles provide the seed patterns for initial generalizations over learned word gestalts. Feedback between these initial generalizations and word productions allows the infant to develop practiced sub-lexical "templates" (see also Fikkert 2007; Fikkert and Levelt 2008). The substitution of these generalized phonological "templates" in place of gestalts accounts for the period of poorer production accuracy in matching the adult pronunciation of given words, yet greater precision between individual utterances. Accuracy subsequently improves as the myriad interactions with caregivers further shape the trajectory of learning. Vihman *et al.* argue that a self-organizing, feedback-driven model of this kind is particularly well suited to explain both the highly individual initial production templates observed in children, and their subsequent convergence on a community standard of pronunciation.

3.3 *Conspiracies in historical phonology*

Conspiracies, in which a seemingly disparate set of processes all result in a common pattern, are widespread in phonology (Kisseberth 1970; CHAPTER 70: CONSPIRACIES). Blust identifies many different types of diachronic change in Austronesian languages that conspire to create a disyllabic word, in many cases restoring a historical disyllable that had lost or gained a syllable through other changes (Blust 2007). In the history of Javanese, for example, reduplication, epenthesis, deletion, and loss of a morpheme boundary have been favored if the product of change is a disyllabic word. Blust suggests that conspiracies arise when a pattern in a language becomes particularly salient, leading it to function as a "linguistic attractor" in language change (Cooper 1999). Within variationist approaches to language change (e.g. Ohala 1989; Labov 1994; Bybee 2001; Blevins 2004; and many others; see also CHAPTER 92: VARIABILITY), a salient pattern can influence the course of language change by biasing categorization (cf. the notion of "Change" in Evolutionary Phonology; Blevins 2004), and/or by biasing the range of variation in production. (For a simulation of pattern feedback in production biasing language change, see Wedel 2007.) In this context, Blust notes that disyllables have been reported to make up 94 percent of the set of content words in proto-Austronesian, and that in many modern Austronesian languages the disyllable remains the dominant word type.

A wide range of evidence is consistent with the hypothesis that such linguistic attractors bias individual behavior and thereby influence the course of language change. Many studies have shown that variation in linguistic behavior is biased toward previous experience: both grammaticality judgments (e.g. Albright 2002; Krott *et al.* 2002; Pierrehumbert 2006a) and production variation/errors (e.g. Bybee and Moder 1983; Dell *et al.* 2000; Vitevitch and Sommers 2003; Gonnerman *et al.* 2007; and many more) are biased by similarity and by pattern type-frequency at a wide range of representational levels (for reviews on various of these topics, see Bybee 2001; Ernestus and Baayen 2003; Bybee and McClelland 2005; Pierrehumbert 2006b; Baayen 2007; Pisoni and Levi 2007). Patterns of simulated language change in simple model systems (Wedel 2007, 2009) are also consistent with this notion: when production and perception errors by simulated agents are biased toward previous perception and production experience, change is strongly influenced by pre-existing patterns in the system. Further evidence concerning the influence of attractors in language change could be sought using iterated artificial language learning and transmission paradigms of the sort pioneered by Kirby *et al.* (2008). Finally, it is worth noting that within models such as Evolutionary Phonology, synchronic alternation patterns are created through diachronic change rather than through mechanisms localized within a single individual's language faculty (see Blevins 2004: ch. 3 for a review of earlier theories of this type). Under this model, this account of diachronic conspiracies provides the basis for an account of synchronic conspiracies as well.

3.4 *Actuation vs. propagation of change*

A significant question in historical linguistics is how an initially isolated change can survive and propagate throughout a community, given that language learners tend to converge on a common community standard. To the extent that this is the case, isolated variants should never be able to gain a foothold in a speech community, because every learner is exposed to many speakers (see Keller 1994: 99 and Nettle 1999 for discussion). In a foundational paper, Nettle (1999) uses a well-articulated simulation to explore factors that are required to allow randomly occurring variants to become established, assuming the existence of a stratified social structure. He finds that given reasonable assumptions (i.e. that *ceteris paribus*, learners tend to adopt the local majority pattern), random variation in acquisition is not sufficient to induce a population-wide transition from one pattern to another without being so pervasive as to obliterate any coherent pattern at all. He then shows that when significant prestige inequities are introduced in which a small number of individuals serve disproportionately as acquisition models, a novel variant can survive if it spreads in the population sufficiently to support itself through positive feedback. When a small number of individuals exert strongly disproportionate influence, the *effective population size* is small, allowing random events a greater chance of influencing the trajectory of change (see the literature on genetic drift in biological populations, e.g. the introductory article by Kliman *et al.* 2008). However, it is clear that functional articulatory and perceptual factors influence the course of change as well; otherwise, we should observe as many diachronic changes that are phonetically unnatural as those that are natural. Nettle explores the influence of functional biases in his model, and concludes that in order to enforce change alone, functional biases have to be sufficiently strong so that anti-functional patterns should never occur. Since this is not the case, Nettle argues that social factors are a critical engine of change, but that the rate of actuation and the efficiency of propagation must also be biased by functional factors that influence ease of production, perception, and acquisition.

3.5 Emergence of phonemes and inventory structure

Vowel inventories appear to be constructed to optimize perceptual contrast between neighboring vowels, given extant articulatory constraints (Liljencrants and Lindblom 1972). How does this apparent optimization come about? In an early self-organizational approach to this problem using a perception/production feedback loop, de Boer (2000) proposed that structure in vowel inventories emerges through interaction of language users under perceptual and production constraints, assuming a tendency for language users to imitate each other (see Browman and Goldstein 2000 for an abstractly similar self-organizational account couched within Articulatory Phonology; also CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATIONS). To test and illustrate this idea, de Boer constructed a simulation in which a group of agents can produce, perceive, and remember vowel pronunciations in the form of prototypes. (Agents are entities within a simulation that can change independently, here representing individual language users.) Within the simulation, agents speak and imitate each other, modifying their vowel categories in response to how successful their imitations are. In each round, a random pair of agents is chosen from the larger set of agents to act as speaker and hearer. The speaker articulates a randomly chosen vowel from memory with some random error; if it has no vowels in memory, it produces a random vowel within the available articulatory space. The hearer compares the formant values of the vowel to prototypes it has in memory and chooses the closest one. If it has no vowels in memory, it creates a similar vowel and calculates its associated articulatory parameters. The listener then repeats the matched prototype vowel for the speaker, who checks to see how close it was to the originally produced vowel. If the vowel is judged to be the same, the speaker agent gives the listener feedback that its imitation was successful. In that case, the listener shifts the parameters of its matching prototype closer to the vowel that it heard from the speaker. If the imitation was not successful, the listener checks its memory to find out how often that prototype has given rise to successful imitations. If it has been mostly unsuccessful, it moves that prototype closer to the sound it heard, just as in a successful imitation. If it has been mostly successful before, it may be that the speaker has an additional prototype vowel in that region of vowel space, and so the listener creates a new prototype to match approximately what it heard. Several additional processes come into play in this simulation: (i) if a vowel prototype is infrequently matched to a perceived vowel, it is discarded; (ii) if two vowel prototypes are too close together, they are merged; and (iii) new vowels are introduced by speakers at a low frequency (CHAPTER 90: FREQUENCY EFFECTS).

This simulation employs a number of features that may not correspond directly to actual features of language use (e.g. direct feedback on imitative success; the operative mechanisms of category loss and merger), but that is not its primary point. These mechanisms simply allow vowel inventories in individual agents to change over time in response to constraints on the differentiation of vowels that are perceived in individual usage events. De Boer shows that, given these constraints, populations of agents starting with empty vowel inventories develop jointly held phonetically natural vowel inventories. He concludes from this that the typological generalizations over vowel inventories found in natural language may arise through articulatory and perceptual constraints in usage rather than some more direct, innate specification. Coherent structure is primarily driven by positive feedback in this system, which comes in two forms: modification of prototypes toward perceived vowels, and merger of prototypes that get too close. These encourage the development of coherent vowel categories shared across the set of agents. Because vowels that are too perceptually confusable tend to be merged, the set of surviving vowels tends toward a perceptually "optimal" arrangement.

Oudeyer (2002, 2006) has used an abstractly similar, more physiologically grounded model of a perception/production feedback loop to argue that positive feedback inherent in processing can create categorial distinctions in the absence of any functional pressure. Research in response biases of cortical fields of neurons shows that their output is well predicted by the aggregate response of the entire field, rather than by the output of the most highly activated neuron. From the set of activities of all neurons, it has been found that one can predict the perceived stimulus or motor output by computing the population vector over the field, namely, the sum of all preferred outputs of the set of neurons multiplied by their activities (Georgeopoulos *et al.* 1986; for an account of the perceptual magnet effect (Kuhl 1991) based in this phenomenon, see Guenther and Gjaja 1996). The important feature of the population vector for our purposes is that it is shifted toward the center of the local distribution of outputs relative to the most highly activated neuron. Given a close mapping between perception and production (Oudeyer 2002; Fowler and Balantucci 2005), this property of cortical fields should produce positive feedback promoting the coalescence of perceptual-motor categories into well-defined distributions over many cycles of use.

In Oudeyer's (2002) model, linked motor and perceptual cortical fields are initially populated with randomly tuned neurons, such that there are no distinct coherent sound categories. Over the course of the simulation, randomly chosen production stimuli are produced by the motor field and processed by the perceptual field. In processing, each neuron in the perceptual field is activated by the production stimulus under the control of a Gaussian tuning factor responsive to the degree of match between the stimulus and a neuron's preferred vector. The preferred vectors of all neurons that have been activated to some degree by the stimulus are then shifted toward that of the maximally active neuron, producing a reversion to the local mean. This update function acts to incrementally consolidate the vectors exhibited by the neural map, influenced by random peaks in the distribution of stimuli produced early in the simulation. The perceptual and motor fields are linked by an update function that shifts vectors in the motor field in parallel to those in the perceptual field, closing the perception/production feedback loop. The resulting positive feedback between perception and production allows a rapid collapse of the originally random distribution of vectors in the sensory map into a small number of coherent sound-motor categories. Oudeyer interprets this feature of his model to suggest that native features of our neurological production and perception apparatus may be designed to develop categories of a particular granularity, and that this feature may play a role in the development of phoneme inventories.

3.6 *Merger* vs. *contrast maintenance*

When one sound category becomes more similar to another over the course of sound change, one possible outcome is category merger, as has occurred between /ɑ/ and /ɔ/ in western dialects of North American English (CHAPTER 80: MERGERS AND NEUTRALIZATION). Often however, the set of categories translates through phonetic space in a "chain shift," such that the system of contrasts is maintained even though the specific phonetic correlates of each category change (see Gordon 2002 for a review). A number of feedback-based models have been recently proposed that provide accounts for the mechanisms of both merger and/or shift and make predictions about the conditions under which either may occur. To understand these models, it will be helpful to review briefly the role of experienced phonetic variation in production and perception. A wide variety of experimental evidence indicates that individual percepts can leave detailed, long-lived traces in memory, and that these memory "exemplars" influence future perception and production behavior (for reviews, see Tenpenny 1995; Johnson 1997; also Pierrehumbert 2006b and the papers in Gahl and Yu 2006). The influence of perception on production (Goldinger 2000; Nielsen 2007) creates the possibility of a perception/production feedback loop in which the effect of biases anywhere in the cycle can potentially build up over time to shift behavior within a single generation. Pierrehumbert used an exemplar-based model of this loop to explore the consequences of feedback for merger between perceptually adjacent phonological categories (2001, 2002). In this model, categories consist of an abstract label and a set of stored perceptual exemplars that have been mapped to that category, where each exemplar is associated with an activation level that decays exponentially over time. No proposed mechanism in this particular model requires transmission between distinct agents, so as a simplification the simulation architecture uses a single category system in conversation with itself. Production proceeds by probabilistically choosing an exemplar in relation to activation level, averaging all the exemplar values within a set window around the chosen exemplar in proportion to their activations and then adding a small amount of normally distributed noise to that average. Averaging within a window around a single exemplar creates a reversion to the mean of the local distribution, just as the use of the population vector does in Oudeyer's simulation described above. Adding noise to production outputs keeps the distribution from collapsing to a single point through the effect of averaging and allows the system to evolve over time. To decide what label the new output should be categorized under, the summed activation level of exemplars within a set window around the output value is calculated for each label. The percept is then stored as a new exemplar under the category label with the highest score.

Pierrehumbert showed that, given this architecture, if two categories drift close enough such that they begin to compete for percepts along their mutual boundary, the category with greater overall exemplar activation tends to eventually absorb the less active category. This occurs through positive feedback between current activation and the ability to compete for percepts. All else being equal, an ambiguous percept is more likely to be mapped to a more active category than a less active category, which only results in the active category becoming yet more active with respect to the less active category. This snowballing feedback results in more and more percepts being mapped to the more active category, until the activation of the other category eventually falls low enough that it effectively no longer exists and is absorbed into the more frequent category. An example that can be modeled this way is the above mentioned merger of /ɔ/ with the more frequent /ɑ/ in western dialects of American English.

In all of the work reviewed above, local, similarity-based positive feedback drives coalescence of system elements into categorial groupings. In none of these accounts, however, is there any repulsive force that would prevent the steady merger of categories over time as they eventually drift into one another. As a consequence, the maintenance of multiple categories over the course of language change in these models would require either regular generation of new distinctions as in the de Boer (2000) model above or some mechanism to favor preservation of at least some existing contrasts. Boersma and Hamann (2008) approach the problem of contrast maintenance between existing sound categories through a constraint-based model that makes use of categorization accuracy on the part of a language learner. As a demonstration of their model, Boersma and Hamann simulate the evolution of category label/contents mappings within a unidimensional space. To concretize the model, they use the spectral frequency range of sibilants in human languages as the perceptual space. (In the following brief discussion of their model, I use an /s/ + /ʃ/ two-sibilant system as a running example, although more or fewer categories are possible.) The architecture of the model is vertical, in which a naive agent learns to associate part of the spectral frequency range with /s/ and another part with /ʃ/ by hearing examples from a teaching agent with feedback. After this learning phase, the agent becomes a teacher and produces examples of /s/ and /ʃ/ to a new learner agent, and so on. For the purposes of the argument, Boersma and Hamann assume that learning agents have acquired sound category labels from word patterns prior to the beginning of the simulation. As a result, learners know at the start how many sibilant categories their language has, but are not yet sure where their distributions lie within the frequency continuum. In addition, learners have the ability to learn an association between a given spectral frequency and /s/ or to /ʃ/ by constructing a ranking among optimality-theoretic constraints banning the mapping of a particular frequency to a particular category label (Boersma 1997). The architecture of the perception grammar incorporates both frequency of presentation and categorization accuracy, with the result that the grammar is maximally "certain" about mappings for sibilant frequencies that are further apart relative to those frequencies that were most often heard. Because subsequent production is based on sampling from the learned perception grammar rather than from the distribution of actually learned examples, an agent's production favors a distribution of sibilants that is slightly better separated than that which she heard herself. This creates a positive feedback loop that promotes increasing contrast between categories over many teacher/learner cycles. However, agents' productions are also influenced by ranked articulatory constraints that have the effect of biasing productions toward the center of the frequency continuum. Boersma and Hamann show that under the balancing influence of positive feedback via the perception grammar and negative feedback from articulatory constraints, categories evolve as well-spaced distributions with a joint center of gravity at the midpoint of the continuum.

The Boersma and Hamann model relies on error feedback from the lexicon within the learner to drive contrast maintenance. Another model for contrast maintenance has been proposed that operates in situations in which there is an *absence* of error feedback (Wedel 2004; Blevins and Wedel 2009; see also Ettlinger 2007). The relevant absence of error feedback in this model occurs when an ambiguous pronunciation is not rescued by an external context to determine its intended category mapping. As an example of external disambiguation, words like "too" and "two" in English are rarely confused by listeners, because they are used in distinct sentential and semantic contexts. In contrast, because of their semantic similarity, words within morphological paradigms are often distinguished in context primarily by their phonetic differences. For example, utterances like *I cook chicken well* and *I cooked chicken well* could be used in very similar contexts, in which case the tense of the verb *cook* is conveyed almost entirely by the audible presence or absence of the past tense [-t].

A hypothesis introduced in Wedel (2004) and explored more deeply in Blevins and Wedel (2009) is that this effectively greater "functional load" of word-internal phonetic information within paradigms may account for anti-homophony effects in paradigms. "Anti-homophony" refers to the failure of otherwise regular sound changes to occur in words when that change would render them homophonous with another word (see Blevins and Wedel 2009 for a review and examples). As in the Pierrehumbert (2001, 2002) model of category merger reviewed above, this model rests on evidence that category behavior is updated by experience: if a pronunciation is ambiguous in context, a hearer may map it to a category that the speaker did not intend, resulting in the effective trading of a variant between the two categories at their boundary. It is this "variant trading" between perceptually adjacent categories that drives the behavior of the model by preserving a crisp boundary between adjacent categories (see Blevins and Wedel 2009 for more discussion).

4 Looking forward

Our ability to build theories is limited by the knowledge that we already have. Whether or not any of the current self-organization-based accounts in phonology are "right," they are valuable in expanding our understanding of pattern formation mechanisms. Self-organization is ubiquitous in physical, biological, and cultural systems, and given that language provides the conditions for self-organization many times over, linguists should anticipate finding it as a contributing mechanism in this domain as well. Just as with any other type of account, however, showing that a particular structure could arise through self-organization does not mean that it does so. A hypothesis stands or falls on the empirical success of its predictions. Arguments from first principles about, for example, whether phonological patterns are more likely to derive from a language-specific cognitive faculty or a more general set of factors are less valuable than well-constructed tests of model predictions. Fortunately, a wide variety of techniques and approaches are now available for testing hypotheses, from the ever-growing array of psycholinguistic techniques through corpus studies and artificial language-learning paradigms.

Self-organizational models for structure formation make use of previously identified cognitive, articulatory, perceptual, or social factors as contributing building blocks. In turn, these models make further predictions about those factors, or may predict some yet undescribed phenomenon. For example, self-organizing models of phonological change through usage require a production/perception feedback loop that can drive small, but persistent and generalizable changes in post-acquisition phonological categories (e.g. Bybee 2001; Pierrehumbert 2001, 2002; Wedel 2007; Mielke 2008). Although more work needs to be done to establish their generality, results from a variety of psycholinguistic studies are consistent with this prediction (e.g. Goldinger 2000; Kraljic and Samuel 2006; Nielsen 2007). Likewise, each of the other models reviewed above makes new predictions that can be tested empirically. As phonologists bootstrap back and forth between model building and simulation on the one hand and empirical methods on the other, the field should gain a steadily better sense of whether and how self-organizational mechanisms contribute to the wide variety of phenomena that we study.

REFERENCES

Albright, Adam. 2002. Islands of reliability for regular morphology: Evidence from Italian. *Language* 78. 684–709.

Baayen, R. Harald. 2007. Storage and computation in the mental lexicon. In Gonia Jarema & Gary Libben (eds.) *The mental lexicon: Core perspectives*, 81–104. Amsterdam: Elsevier.

Ball, Phillip. 1999. *The self-made tapestry: Pattern formation in nature*. Oxford: Oxford University Press.

Beckman, Mary E. & Jan Edwards. 2000. The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development* 71. 240–249.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. 2009. Language is a complex adaptive system. *Language Learning* 59. Supplement 1. 1–26.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette & Andrew Wedel. 2009. Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26. 143–183.

Blust, Robert. 2007. Disyllabic attractors and anti-antigemination in Austronesian sound change. *Phonology* 24. 1–36.

Bod, Rens, Jennifer Hay & Stefanie Jannedy (eds.) 2003. *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Boer, Bart de. 2000. Self-organization in vowel systems. *Journal of Phonetics* 28. 441–465.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21. 43–58 (ROA-221).

Boersma, Paul & Silke Hamann. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25. 217–270.

Browman, Catherine P. & Louis Goldstein. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5. 25–34.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan & James L. McClelland. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* 22. 381–410.

Bybee, Joan & Carol Moder. 1983. Morphological classes as natural categories. *Language* 59. 251–270.

Camazine, Scott, Jean-Louis Deneubourg, Nigel R. Franks, James Sneyd, Guy Theraulaz & Eric Bonabeau. 2001. *Self-organization in biological systems*. Princeton: Princeton University Press.

Cooper, David L. 1999. *Linguistic attractors: The cognitive dynamics of language acquisition and change*. Amsterdam & Philadelphia: John Benjamins.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. Harlow: Longman.

Cziko, Gary. 1995. *Without miracles: Universal selection theory and the second Darwinian revolution*. Cambridge, MA: MIT Press.

Dawkins, Richard. 1986. *The blind watchmaker*. New York: W. W. Norton & Co.

Deacon, Terence W. 1997. *The symbolic species: The co-evolution of language and the brain*. New York: W. W. Norton & Co.

Dell, Gary S., Kristopher D. Reed, David R. Adams & Antje S. Meyer. 2000. Speech errors, phonotactic constraints and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26. 1355–1367.

Dennet, Daniel. 1995. *Darwin's dangerous idea: Evolution and the meanings of life*. New York: Simon & Schuster.

Elman, Jeffrey L. 1995. Language as a dynamical system. In Robert F. Port & Timothy van Gelder (eds.) *Mind as motion: Dynamical perspectives on behavior and cognition*. Cambridge, MA: MIT Press. 195–225.

Ernestus, Mirjam & R. Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79. 5–38.

Ettlinger, Marc. 2007. Shifting categories: An exemplar-based computational model of chain shifts. *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, 239–244. Mahwah, NJ: Lawrence Erlbaum.

Ferguson, Charles A. & Carol B. Farwell. 1975. Words and sounds in early language acquisition. *Language* 51. 419–439. Reprinted with appendix in William S.-Y. Wang (ed.) 1977. *The lexicon in phonological change*, 7–68. The Hague: Mouton.

Fikkert, Paula. 2007. Acquiring phonology. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 537–554. Cambridge: Cambridge University Press.

Fikkert, Paula & Clara C. Levelt. 2008. How does Place fall into place? The lexicon and emergent constraints in children's developing grammars. In Peter Avery, B. Elan Dresher & Keren Rice (eds.) *Contrast in phonology: Theory, perception, acquisition*, 231–268. Berlin & New York: Mouton de Gruyter.

Fowler, Carol A. & Bruno Balantucci. 2005. The relation of speech perception and speech production. In David B. Pisoni & Robert E. Remez (eds.) *The handbook of speech perception*, 633–652. Malden, MA & Oxford: Blackwell.

Gahl, Suzanne & Alan C. L. Yu (eds.) 2006. Special issue on exemplar-based models in linguistics. *The Linguistic Review* 23. 213–379.

Gardner, Martin. 1970. Mathematical games. *Scientific American* 223. 120–123.

Garrett, Andrew. 2008. Paradigmatic uniformity and markedness. In Jeff Good (ed.) *Linguistic universals and language change*, 125–143. Oxford: Oxford University Press.

Gell-Mann, Murray. 1992. Complexity and complex adaptive systems. In John A. Hawkins & Murray Gell-Mann (eds.) *The evolution of human languages*, 3–18. New York: Addison-Wesley.

Georgeopoulos, Apostolos, Andrew B. Schwartz & Ronald E. Ketter. 1986. Neuronal population coding of movement direction. *Science* 233. 1416–1419.

Goldinger, Stephen D. 2000. The role of perceptual episodes in lexical processing. In Anne Cutler, James M. McQueen & Rian Zondervan (eds.) *Proceedings of SWAP (Spoken Word Access Processes)*, 155–159. Nijmegen: Max Planck Institute for Psycholinguistics.

Gonnerman, Laura M., Mark S. Seidenberg & Elaine S. Andersen. 2007. Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology* 136. 323–345.

Gordon, Matthew J. 2002. Investigating chain shifts and mergers. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.) *The handbook of language variation and change*, 244–266. Oxford: Blackwell.

Guenther, Frank H. & Marin N. Gjaja. 1996. The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America* 100. 1111–1121.

Hare, Mary & Jeffrey L. Elman. 1995. Learning and morphological change. *Cognition* 56. 61–98.

Haspelmath, Martin. 1999. Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18. 180–205.

Heylighen, Francis. 2001. The science of self-organization and adaptivity. In L. Douglas Kiel (ed.) *Knowledge management, organizational intelligence and learning, and complexity*. Available (October 2010) at www.eolss.net.

Hock, Hans Henrich. 2003. Analogical change. In Brian D. Joseph & Richard Janda (eds.) *The handbook of historical linguistics*, 441-480. Malden, MA: Blackwell.

Hurford, James. 1999. The evolution of language and languages. In Robin Dunbar, Chris Knight & Camilla Power (eds.) *The evolution of culture*, 173–193. Edinburgh: Edinburgh University Press.

Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson & John W. Mullenix (eds.) *Talker variability in speech processing*, 145–165. San Diego: Academic Press.

Kauffman, Stuart. 1995. *At home in the universe: The search for the laws of self-organization and complexity*. Oxford: Oxford University Press.

Keller, Rudi. 1994 *On language change: The invisible hand in language*. London: Routledge.

Kirby, Simon. 1999. *Function, selection and innateness: The emergence of language universals*. Oxford: Oxford University Press.

Kirby, Simon, Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105. 10681–10686.

Kisseberth, Charles W. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1. 291–306.

Kliman, Richard, Bob Sheehy & Joanna Schultz. 2008. Genetic drift and effective population size. *Nature Education* 13. Available (July 2010) at www.nature.com/scitable/topicpage/Genetic-Drift-and-Effective-Population-Size-772523.

Kraljic, Tanya & Arthur Samuel. 2006. Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review* 13. 262–268.

Krott, Andrea, Robert Schreuder & R. Harald Baayen. 2002. Analogical hierarchy: Exemplar-based modeling of linkers in Dutch noun-noun compounds. In Royal Skousen, Deryle Lonsdale & Dilworth B. Parkinson (eds.) *Analogical modeling: An exemplar-based approach to language*, 181–206. Amsterdam & Philadelphia: John Benjamins.

Kuhl, Patricia K. 1991. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50. 93–107.

Labov, William. 1994. *Principles of linguistic change*, vol. 1: *Internal factors*. Oxford: Blackwell.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48. 839–862.

Lindblom, Björn. 1992. Phonological units as adaptive emergents of lexical development. In Charles A. Ferguson, Lise Menn & Carol Stoel-Gammon (eds.) *Phonological development: Models, research, implications*, 131–163. Timonium, MD: York Press.

Lindblom, Björn. 2000. Developmental origins of adult phonology: The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. *Phonetica* 57. 297–314.

Lindblom, Björn, Peter MacNeilage & Michael Studdert-Kennedy. 1984. Self-organizing processes and the explanation of language universals. In Brian Butterworth, Bernard Comrie & Östen Dahl (eds.) *Explanations for language universals*, 181–203. Berlin: Mouton.

MacWhinney, Brian. 2006. The emergence of linguistic form in time. *Connection Science* 17. 191–211.

Medeiros, David. 2008. Optimal growth in phrase structure. *Biolinguistics* 2. 156–195.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Nettle, Daniel. 1999. Using social impact theory to simulate language change. *Lingua* 108. 95–117.

Nielsen, Kuniko. 2007. The interaction between spontaneous imitation and linguistic knowledge. *UCLA Working Papers in Phonetics* 105. 125–137.

Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In Leiv Egil Breivik & Ernst Håkon Jahr (eds.) *Language change: Contributions to the study of its causes*, 173–198. Berlin: Mouton de Gruyter.

Oudeyer, Pierre-Yves. 2002. Phonemic coding might be a result of sensory-motor coupling dynamics. In Bridget Hallam, Dario Floreano, John Hallam, Gillian Hayes & Jean-Arcady Meyer (eds.) *Proceedings of the 7th International Conference on the Simulation of Adaptive Behavior*, 406–416. Cambridge, MA: MIT Press.

Oudeyer, Pierre-Yves. 2006. *Self-organization in the evolution of speech*. Oxford: Oxford University Press.

Peck, Steven L. 2004. Simulation as experiment: A philosophical reassessment for biological modeling. *Trends in Ecology and Evolution* 19. 530–534.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam & Philadelphia: John Benjamins.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology* 7, 101–139. Berlin & New York: Mouton de Gruyter.

Pierrehumbert, Janet B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46. 115–154.

Pierrehumbert, Janet B. 2006a. The statistical basis of an unnatural alternation. In Louis Goldstein, Douglas Whalen & Catherine T. Best (eds.) *Laboratory phonology* 8. 81–107. Berlin & New York: Mouton de Gruyter.

Pierrehumbert, Janet B. 2006b. The new toolkit. *Journal of Phonetics* 34. 516–530.

Pisoni, David B. & Susannah V. Levi. 2007. Some observations on representations and representational specificity in speech perception and spoken word recognition. In Gareth Gaskell (ed.) *The Oxford handbook of psycholinguistics*, 3–18. Oxford: Oxford University Press.

Plaut, David C. & Christopher T. Kello. 1999. The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In Brian MacWhinney (ed.) *Emergence of language*, 381–415. Mahwah, NJ: Lawrence Erlbaum.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Steels, Luc. 2000. Language as a complex adaptive system. In Marc Shoenauer (ed.) *Proceedings of PPSN VI, Lecture Notes in Computer Science*, 17–26. Berlin: Springer-Verlag.

Stemberger, Joseph & Brian MacWhinney. 1988. Are inflected forms stored in the lexicon? In Michael Hammond & Michael Noonan (eds.) *Theoretical morphology: Approaches in modern linguistics*, 101–116. New York: Academic Press.

Tenpenny, Patricia L. 1995. Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin and Review* 2. 339–363.

Vihman, Marilyn M. & William Croft. 2007. Phonological development: Toward a "radical" templatic phonology. *Linguistics* 45. 683–725.

Vihman, Marilyn M., Rory A. DePaolis & Tamar Keren-Portnoy. 2009. A dynamic systems approach to babbling and words. In Edith L. Bavin (ed.) *The Cambridge handbook of child language*, 163–182. Cambridge: Cambridge University Press.

Vitevitch, Michael S. & Mitchell S. Sommers. 2003. The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory and Cognition* 31. 491–504.

Wedel, Andrew. 2004. Category competition drives contrast maintenance within an exemplar-based production/perception loop. In John A. Goldsmith & Richard Wicentowski (eds.) *Proceedings of the 7th Meeting of the ACL Special Interest Group in Computational Phonology*, 1–10. Barcelona: Association for Computational Linguistics.

Wedel, Andrew. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23. 247–274.

Wedel, Andrew. 2007. Feedback and regularity in the lexicon. *Phonology* 24. 147–185.

Wedel, Andrew. 2009. Resolving pattern conflict: Variation and selection in phonology and morphology. In James P. Blevins & Juliette Blevins (eds.) *Analogy in grammar: Form and acquisition*, 83–100. Oxford: Oxford University Press.

96 Experimental Approaches in Theoretical Phonology

SHIGETO KAWAHARA

1 Introduction

This chapter provides an overview of how experimental work has informed phonological theories, and vice versa. This chapter starts with a historical overview; when phonology was being established as its own area of research, there was a sharp division between phonetics and phonology. This division was called into question, and issues concerning the phonetics–phonology interface are currently being extensively pursued by an approach that is now known as "laboratory phonology." After this historical overview, I discuss in some detail how phonetic experiments and phonological theories have informed each other.

1.1 *The tension between phonetics and phonology*

When phonology was being established as its own area of research, it was often assumed that phonology and phonetics were independent of one another. For example, Trubetzkoy (1939: 11) stated:

> The speech sounds . . . possess a large number of acoustic and articulatory properties. All of these are important for the phonetician since it is possible to answer correctly the question of how a specific sound is produced only if all of these properties are taken into consideration. Yet most of these properties are quite unimportant for the phonologist.

We still sometimes witness a sharp divide between phonetics and phonology in the current literature: some claim that phonology is an abstract, substance-free computational system, which should be separated out from phonetics: "patterns of phonetic substance are not relevant to phonological theory strictly defined" (Hale and Reiss 2000: 158; see Blaho 2008 for a recent review of this position). There has been an uneasiness about integrating phonetics into phonological studies, because of a belief that the phonetic module belongs to performance and the phonological module belongs to competence; i.e. phonetics does not belong to grammar per se (see e.g. Blaho 2008: 2). An assumption behind this claim is that phonetics involves automatic, universal mechanisms.

However, contrary to the view that phonetics consists of universal implementation rules, experiments have shown that phonetics is neither automatic nor universal; i.e. speakers control their phonetic behaviors, and cross-linguistic variation exists in the realm of phonetics (Keating 1985, 1988a; Kingston and Diehl 1994; see also §2.3.1 for more discussion). In this sense, any adequate model of grammar must integrate phonetics as a part of its model.

Phoneticians also responded to the thesis that we can and should study phonology without considering the phonetic mechanisms behind phonology (see Lindblom 1962, Ohala 1990b, and Diehl 1991 for general discussions).[1] Perhaps the best-known advocate of the objection is Ohala (1990b, among many other references). His general point is that many phonological patterns can be explained in terms of articulatory and perceptual factors, and therefore purely phonological explanations without considering phonetic substances can be arbitrary, circular, and post hoc. For example, many languages lack voiced stops in their inventory, a situation for which we could propose a redundancy rule [−son, −cont] → [−voice], as in *SPE* (Chomsky and Halle 1968), or a constraint *[−son, −cont, +voice], but these approaches miss the aerodynamic reason behind the dispreference for voiced stops. In order for speakers to maintain voicing, intraoral air pressure must be lower than subglottal air pressure, but the airflow required for voicing increases intraoral air pressure when the airway is significantly occluded. The increase in intraoral air pressure in turn makes it difficult to satisfy the aerodynamic condition. For this reason, it takes additional articulatory effort – e.g. larynx lowering, tongue advancement, etc. – to keep the intraoral air pressure sufficiently low to maintain voicing during stop closure (Jaeger 1978; Westbury 1979; Ohala 1983).

In addition to this kind of articulatory difficulty, perceptual factors demonstrably affect phonological patterns as well. For example, non-low back vowels are usually rounded, so that we could postulate a redundancy rule [−low, +back] → [+round] or a constraint *[−low, +back, −round], but again these explanations miss the generalization that rounding, by enlarging a resonance cavity, enhances the F2 difference between back and front vowels (Stevens *et al.* 1986; Diehl and Kluender 1989; Diehl 1991).

Finally, psycholinguistic factors also seem to play an important role in shaping phonological patterns. For instance, word-initial segments provide important cues for word recognition (Nooteboom 1981; Hawkins and Cutler 1988). Speakers thus seem to disfavor making phonological changes in word-initial positions, because such changes would result in difficulty in word recognition (Beckman 1997; Kawahara and Shinohara 2010).

In summary, there seem to be phonetic and/or psycholinguistic reasons behind many, if not all, phonological patterns. Therefore, according to Ohala, purging phonetic and psycholinguistic factors from phonological theory is misguided.

1.2 Current situation

Thus, on the one hand, there has been some reluctance to incorporate phonetic (and psycholinguistic) factors into phonological explanations. On the other hand, phonetics does seem to offer some insights into phonological patterns. For this

[1] The tension between phonetics and phonology existed before the inception of generative phonology (Chomsky and Halle 1968). See Ohala (1990b, 1999) for reviews of the history.

(and perhaps other) reason(s), there has been some unfortunate intellectual tension between phonetics and phonology, which Ohala describes as a "turf war" (1990b: 168), where people from each discipline felt that they had to delineate and defend their own territory.[2]

The situation, however, has been changing, as we witness the rise of a general approach which has come to be known as "laboratory phonology" (see Cohn 2010 for sociological aspects of the development of laboratory phonology in the field of general linguistics). The following quote from Beckman and Kingston (1990: 5) succinctly summarizes the spirit of this approach:

> We believe that the time has come to undo the assumed division of labor between phonologists and other speech scientists; we believe this division of labor creates a harmful illusion that we can compartmentalize phonological facts from phonetic facts. At the very least, we maintain that the endeavor of modeling the grammar and the physics of speech can only benefit from explicit argument on this point.

As the following discussion shows in more detail, many experimental studies have contributed to theoretical debates. The rest of the discussion proceeds as follows. In §2, I discuss how experimental approaches have informed phonological theories. In §3, I reverse direction and discuss cases in which theories have informed experiments. Although I try to be comprehensive in my review, there is necessarily a limit. For further examples and discussion, readers are referred to contributions in the *Laboratory Phonology* series (Kingston and Beckman 1990 *et seq.*), as well as in other volumes and papers devoted to this and other related issues (Ohala 1986b; Ohala and Jaeger 1986; Diehl 1991; Hayes *et al.* 2004; Kingston 2007; Solé *et al.* 2007; Coetzee *et al.* 2009).

2 How experiments have informed theory

2.1 *Beyond introspection-based data*

In generative linguistics, native speakers' intuition – or introspection – is the primary source of data, because "the set of grammatical sentences cannot be identified with any particular corpus of utterances obtained by the linguist in his field work" (Chomsky 1957: 15). Since generative phonology aims to study competence, i.e. what speakers know about their language, rather than performance, i.e. how speakers use the language, the only way to assess competence, it was believed, was introspection (though see Schütze 1996 for a critical discussion). Contrary to this research tradition, phonetic and psycholinguistic experiments have offered important insights into knowledge of grammar.

2.1.1 *Wug tests*

The first good example of experiments that have complemented the introspection-based approach is a wug test. In wug tests, named after an experiment by Berko (1958), native speakers are asked to pronounce novel words. Berko tested

[2] An anonymous reviewer points out that there may also be "punting," when people say that some other subfield is responsible for a phenomenon that they cannot account for.

whether English-speaking children acquire a rule of voicing assimilation in the English plural and other suffixes, and found that, given a nonce word like *wug*, most children pluralize it as *wug*[z], not as **wug*[s], showing that English-speaking children know that the plural suffix and the stem-final consonant must agree in voicing. In this way, the wug test has been used as a litmus test for the productivity of a phonological generalization. Recent years have witnessed a renewed interest in wug tests, which has provided some important insights into phonological knowledge, as summarized in (1).

(1) a. A standard assumption in generative phonology is that speakers assign a simple dichotomous grammatical/ungrammatical judgment to linguistic structures. In other words, speakers should treat all attested structures as equally good, and all ungrammatical structures as equally bad. However, several wug tests revealed that speakers can distinguish the relative grammaticality of two (un)grammatical structures (Shinohara 2004; Zuraw 2007).

b. More generally, the results of wug tests often show stochastic, rather than dichotomous, patterns (Albright and Hayes 2003; Hayes and Londe 2006).

c. Some experiments have shown that the probability of speakers applying a certain phonological process in a wug test reflects the frequency of the items that undergo that phonological process in their language (Bybee 1999; Zuraw 2000; Albright and Hayes 2003; Ernestus and Baayen 2003; Hayes and Londe 2006; Hayes *et al.* 2009; CHAPTER 90: FREQUENCY EFFECTS).

d. In some experiments, speakers either fail to replicate some statistical patterns in the lexicon (Becker *et al.* 2008) or at least show bias against reproducing some arbitrary, though statistically significant, patterns in the lexicon (Hayes *et al.* 2009).

e. Some phonological patterns are not productive (at least under a wug test), which leads to the suspicion that they are not a part of the speakers' grammar. Patterns whose productivity wug tests have failed to reveal include English velar softening (Ohala 1974; though see Pierrehumbert 2006), Japanese verb conjugations (Vance 1987: ch. 12) and Polish raising (Sanders 2001). See also Zimmer (1969) for a test of morpheme structure conditions in Turkish.

2.1.2 *Well-formedness judgment studies*

Another type of experiment which complements generative phonology's introspection-based approach is well-formedness judgment experiments. In these experiments, native speakers are asked to judge the naturalness of particular words or phonological processes (they can also take the form of word-likeliness judgments). These experiments, like wug tests, reveal, for example, that speakers can distinguish the relative grammaticality of two (un)grammatical structures (Pertz and Bever 1975; Coetzee 2008) and show that grammatical patterns exhibit a stochastic, rather than a simple dichotomous grammatical/ungrammatical, distinction (Hayes 2000; Albright and Hayes 2003; Fanselow *et al.* 2006). Well-formedness judgments are also known to reflect the frequency of the target items (e.g. Frisch *et al.* 2000; see also CHAPTER 90: FREQUENCY EFFECTS).

2.2 *Addressing the quality of phonological data*

2.2.1 *Re-evaluating phonological data*

Experiments have also re-evaluated what is phonological and what is not. Phonetic experiments have shown that many textbook examples of "phonological patterns" do not involve categorical changes but instead involve gradient changes, suggesting that they might be phonetic processes (CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY). For example, English was thought to have a vowel nasalization rule before a nasal consonant, as in [bĩn] *bean* and [dĩn] *dean* (Fromkin and Rodman 1998: 280–281). Cohn (1993), however, based on an instrumental study measuring patterns of nasal airflow, showed that English nasalization differs from contrastive nasalization in French in that, the closer to the nasal consonant, the more nasal airflow was detected within a nasalized vowel. English nasalization is therefore gradient rather than categorical, in the sense that it does not alter the whole segment but instead the degree of nasalization changes within a segment. For this reason, Cohn concluded that English nasalization belongs to phonetics. Many other examples of phonological patterns have been argued to show similar gradient properties, which I list in (2):[3]

(2)
- a. Arabic tongue backing (emphasis) spreading (Keating 1990 and references cited therein; see also CHAPTER 25: PHARYNGEALS)
- b. English /l/-velarization in the coda (Sproat and Fujimura 1993)
- c. English flapping (Fox and Terbeek 1977; de Jong 1998; CHAPTER 113: FLAPPING IN AMERICAN ENGLISH)
- d. English phrasal nasal assimilation (Nolan 1992; Gow 2002; but cf. Ellis and Hardcastle 2002)
- e. English phrasal palatalization (Zsiga 1995)
- f. English and French schwa deletion (Fougeron and Steriade 1997; Davidson 2006b; CHAPTER 68: DELETION)
- g. Japanese tonal spreading in unaccented words (Pierrehumbert and Beckman 1988)
- h. Russian vowel reduction in second pre-tonic syllable (Barnes 2002; but cf. Padgett and Tabain 2005; CHAPTER 79: REDUCTION)

The abundance of such examples led Hayes to state "I occasionally wondered, 'Where is the normal phonology that I was trained to study?'" (1995: 68).

The list in (2) shows that many patterns that were believed to be phonological have turned out to be phonetic. A more complex example comes from the domain of intonation. In Japanese and many other languages, the height of tones generally declines toward the end of an utterance. The question arose whether this pattern of declination is due to phonetics or phonology. One could posit that

[3] Davidson (2006a) demonstrates that a "schwa" inserted in English speakers' production of non-native clusters differs from a lexical schwa. She argues that this "schwa" results from gestural mis-coordination, and hence differs from phonological epenthesis (see also Hall 2006 for related cross-linguistic phenomena). However, within the framework of Articulatory Phonology (Browman and Goldstein 1986), she also proposes that gestural mis-coordination arises in the phonological component, rather than in the phonetic component.

this declination is phonetic (Fujisaki and Sudo 1971; see Poser 1984: 200 for more references); for example, subglottal air pressure decreases towards the end of an utterance, and the height of tones naturally drops. On the other hand, McCawley (1968) proposes a phonological rule in Japanese that changes a high tone to a mid tone after another high tone within a phrase. It turns out that it would be most fruitful to approach intonation from both perspectives. Poser (1984) argues that Japanese has both local lowering of H after another H(L), which seems phonological, and gradient, steady declination throughout the utterance, which is phonetic.[4] Beckman and Pierrehumbert argue that a similar hybrid approach accounts for the complex pattern of intonation in both English and Japanese (Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988).

In addition to helping us to decide whether patterns under discussion are phonological or phonetic, some studies have called into question the existence of some phonological patterns per se. Based on the traditional description of Tswana, Hyman (2001) discusses a case of post-nasal devoicing, but Gouskova *et al.* (2006) argue, on the basis of a production experiment, that Tswana may not have a process of post-nasal devoicing after all. A later study, however, suggests that some, though not all, speakers do show evidence of post-nasal devoicing (Coetzee *et al.* 2007). A general lesson we can draw from this series of studies is that careful instrumental experiments help us to establish whether phonological patterns under discussion really exist.

2.2.2 *Incomplete neutralization*

While it is standardly assumed that phonological processes involve categorical changes (see §2.2.1 and §2.3.2), some experiments have called this assumption into question. Port and O'Dell (1985) report a production experiment on German where they found some acoustic differences between underlying voiceless stops and "voiceless" stops that are underlyingly voiced but devoiced by coda devoicing (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). They found appreciable differences between these two categories in terms of preceding vowel duration, closure voicing duration, closure duration, and aspiration duration. Further, they demonstrated that listeners can detect the differences between the two categories at more than chance frequency. They argue that coda devoicing in German is therefore incomplete.

Subsequent studies have found other cases of incomplete neutralization in many languages, including Cantonese (Yu 2007), Catalan (Dinnsen and Charles-Luce 1984), Dutch (Ernestus and Baayen 2006, 2007; Warner *et al.* 2006), English (Fourakis and Port 1986; Ohala 1986a), Japanese (Mori 2002), Lebanese Arabic (Gouskova and Hall 2009), Polish (Slowiaczek and Dinnsen 1985; Slowiaczek and Szymanska 1989) and Russian (Chen 1970; Dmitrieva 2005; Padgett and Tabain 2005); see also CHAPTER 80: MERGERS AND NEUTRALIZATION. Some studies have argued, however, that these experimental results are largely or entirely due to extra-grammatical factors such as speakers' familiarity with English, orthographic influences, and hyperarticulation in a laboratory setting (Fourakis and Iverson 1984; Jassem and Richter 1989; Warner *et al.* 2006). The status of incomplete neutralization is much debated in the literature (see Warner *et al.* 2004; Port and Leary 2005 for reviews), but these

[4] Downstep may apply iteratively (Poser 1984; Kubozono 1988), resulting in quasi-gradient behavior. See §2.2.1 and §2.3.2 for the categorical nature of phonological patterns.

experiments at least show that we need to be careful when talking about the categoricality of phonological changes.[5] See §2.3.2 for more on the discussion on the categorical nature of phonological alternations.

2.3 *Bearing on the architecture of the grammar*

Not only have experiments served to evaluate the quality of phonological data, some phonetic studies have provided important insights into the general architecture of the grammar.[6]

2.3.1 Against universal phonetics

In *SPE*, the output of phonology was considered to be "the phonetic transcription" (*SPE*: 293), which lacked "properties of the signal that are supplied by universal rules" (*SPE*: 235). Keating (1985, 1988a) characterizes this view as phonetics involving universal, automatic rules (see also Kingston and Diehl 1994: §1.2). Phonetic studies soon showed that this view is too simplistic. For example, Chen (1970) compared durations of vowels before voiced consonants and those before voiceless consonants in seven languages (English, French, Russian, Korean, German, Spanish, and Norwegian), and showed that different languages show different degrees of lengthening before voiced consonants. Keating (1979) (reported in Keating 1985) followed up on this result, and showed that neither Czech nor Polish shows a reliable effect of voicing on preceding vowel duration. It therefore seems that the degree of lengthening before voiced consonants is language-specific. Similarly, an acoustic experiment by Port *et al.* (1980) showed that in Japanese vowel durations are heavily affected by the duration of adjacent consonants, but in Arabic such patterns are not evident, concluding that rhythmic compensation is not universal. These examples show that phonetic implementation is neither automatic nor universal. See Port and Leary (2005) for recent summaries of language-specific phonetic patterns.

2.3.2 The phonetics–phonology divide

As briefly discussed in §2.2.1, many experiments have identified a crucial difference between phonetics and phonology: phonological patterns involve complete categorical changes, whereas phonetics yields gradient outcomes (Keating 1990; Cohn 1993, 2006; Zsiga 1995; Tsuchida 1997; Barnes 2002). Experimental results played an essential role in establishing this difference. For instance, an electropalatographic study by Zsiga (1995) showed that English possesses two kinds of palatalization: complete palatalization, which we find in a morphophonological process, as in *press* [prɛs] *vs. pressure* [prɛʃər], and gradient palatalization, which we find across a word boundary, as in *miss you* [mɪʃju]. Zsiga found that the

[5] Some phonologists admit that some phonological changes are incomplete and propose a model of phonology that handles incomplete neutralization (van Oostendorp 2008; Gouskova and Hall 2009). Others consider the results of neutralization as lacking phonological/phonetic specifications (Steriade 1995, 1997), following the theory of phonetic underspecification (Keating 1988b; also Hsu 1996, cited in Steriade 1995, 1997). Yet others consider these incomplete neutralization patterns to be implemented in the phonetic component (Fourakis and Port 1986).

[6] Another topic that would fit in this subsection is the search for the phonetic basis of distinctive features. Due to space limitations, I cannot provide comprehensive discussion. See Kingston (2007) for a summary.

former [ʃ] is [ʃ] throughout its constriction, whereas the latter [ʃ] starts like an [s] and ends like an [ʃ]. An explanation we can give is that the former process involves a categorical phonological change, whereas the latter process is a gradient phonetic gestural overlap.

Pycha (2009) demonstrated another difference: comparing phonological lengthening (i.e. gemination) and phonetic lengthening at phrase edges in Hungarian, she found that phonological lengthening always targets the closure phase of affricates, whereas phonetic phrase edge lengthening affects portions that are adjacent to the boundaries. In this way, experiments have identified characteristics of phonetics that distinguish it from phonology. See Keating (1996: 263) for constellations of other properties that distinguish phonetics and phonology. See also Anderson (1981) for a general discussion of the phonetics–phonology divide.

2.3.3 *The phonetics–morphology interface?*

As exemplified by the two palatalization processes in English, morphophonological processes tend to involve categorical changes, whereas phonetic processes yield gradient outputs. A general assumption in generative studies is thus that phonology can be sensitive to morphology, but phonetics is not. The inaccessibility of morphological structures to phonetics was assumed in Chomsky and Halle (1968), where morphological boundaries are erased at the end of each transformational cycle (*SPE*: 15). The Bracket Erasure Convention in Lexical Phonology (Kiparsky 1982) also removes morphological boundaries after each level of derivation (CHAPTER 85: CYCLICITY). As a result, word-internal structures are inaccessible to later post-lexical rules or phonetics.

As an illustration, take the case of minimal word requirements. Many languages require (lexical) words to be of a certain minimal length, and this requirement is expressed in terms of abstract prosodic units (McCarthy and Prince 1986), but not in terms of raw phonetic duration (Cohn 1998). For example, even though English tense [i] is shorter than lax [æ] in raw duration ([i] = 100 milliseconds; [æ] = 123 milliseconds, according to Strange *et al.* 2004), [pi] is well formed, but [pæ] is not. Therefore, the minimal word requirement operates on abstract phonological units rather than on raw phonetic duration. This sort of requirement can be sensitive to morphological information. For example, in Yoruba, only nouns are required to be maximally disyllabic (Pulleyblank 1988: 250, fn. 24). On the other hand, no known languages seem to vary raw phonetic durations depending on morphological categories. Phonological requirements, therefore, may refer to morphological information whereas phonetic implementation cannot.[7] This thesis has been taken for granted and rarely questioned or addressed in the phonological literature.

However, Cho (2001) directly addressed this issue using EMA (electromagnetic articulography), and found that, in Korean, gestural timing is more variable across a morpheme boundary than within a morpheme, and also more variable across a non-lexicalized compound boundary than across a lexicalized compound boundary. Also, Sproat and Fujimura (1993) used X-ray microbeam technology and compared the amount of dorsal retraction of English coda [l] at various boundaries,

7 In turn, morphological processes can be sensitive to phonological information (i.e. phonologically conditioned allomorphy; see McCarthy 2002: 183 for references), but do not seem to be controlled by phonetic information (though see Bybee 1999 for a case of a morphological pattern that manipulates a non-contrastive feature).

including Level I and Level II boundaries, and found a difference between these two contexts. These experimental findings suggest that morphological boundaries may be visible to phonetic implementation rules.[8]

Another debate was initiated by Steriade (2000), who challenges the immunity of phonetics to morphological information. She argues that there are cases of phonetic analogy, a requirement that paradigmatically related words be phonetically similar. For example, derived words are required to be identical in raw phonetic duration to their corresponding bases. This phonetic analogy is proposed to explain why flapping applies in words like *capi*[ɾ]*alistic* (cf. *capi*[ɾ]*al*), whereas it fails to apply in words like *militaristic* (cf. *mili*[t]*ary*) with a similar stress pattern: the applicability of flapping in derived words depends on whether flapping is possible in the base words. However, Riehl (2003) found in a production experiment that the transfer of flapping from a base form to related words was not robustly observed. On the basis of an acoustic study, Riehl also challenged the assumption that the distinction between [t] and [ɾ] is made solely in terms of constriction duration (see Fox and Terbeek 1977 and de Jong 1998 for the phonetics of flapping; also CHAPTER 113: FLAPPING IN AMERICAN ENGLISH).

In relation to incomplete neutralization discussed in §2.2.2, Ernestus and Baayen (2006, 2007) argue for another case of morphological influence on phonetics. They found that there is slight voicing left in "devoiced" final consonants, and argue that this voicing is due to the activation of morphologically related words with a voiced consonant (CHAPTER 83: PARADIGMS). In summary, whether phonetics has access to morphological information or not is still under debate; experiments will be able to contribute much to this debate (see Bybee 1999, Barnes and Kavitskaya 2003, Davis 2005, Cohn 2006, and Yu 2007 for further discussion on the phonetics–morphology interaction and phonetic analogy).

2.4 *Arguments for and against the psychological reality of grammar*

Not only have experiments addressed what grammar should look like, they have also examined whether grammar is psychologically real in speakers' minds. Many experiments have addressed the question of whether the rules, constraints, and structures that linguists posit are psycholinguistically real or are merely theoretical devices that help us explain the linguistic patterns (Zimmer 1969; Ohala 1974, 1986a; Cena 1978; McCawley 1986). A general concern behind this work is that the psychological reality of a grammatical postulate has sometimes been confused with the analytical success of that postulate. As McCawley (1986: 28) puts it, "Chomsky's [(1986)] policy that the subject matter of linguistics is psychological in nature does not provide any reason for assuming that the purported facts that linguists have hitherto adduced as evidence for or against particular analyses are psychological in nature, nor even that they are strictly speaking facts." Psychological reality of phonological data should not be taken for granted, and must be explicitly tested. Some wug tests in fact revealed that some phonological patterns are not reflected in speakers' behaviors (see (1e)).

[8] A question that remains with respect to these results is whether the differences could be attributed to differences in the presence of prosodic boundaries like foot boundaries or prosodic words, which the phonetics is presumably able to see.

Recent development of experimental techniques has allowed us to address the question of psychological reality from a different perspective. In particular, a number of perception experiments showed that phonological constraints affect speech perception – given ambiguous acoustic signals, speakers are biased against categorizing the stimuli as those not allowed by their phonological grammar (Massaro and Cohen 1983; Pitt 1998; Dupoux *et al.* 1999; Moreton 2002; Berent *et al.* 2007; Coetzee 2008). A classic work by Massaro and Cohen took advantage of word-initial phonotactic restrictions in English, where only [ɹ] is allowed after [t], only [l] is allowed after [s], and both are allowed after [p]. They created a continuum from [ɹ] to [l] by varying F3 and presented the continuum in these contexts, and found that speakers judge tokens as [ɹ] most frequently after [t], less frequently after [p], and least frequently after [s]. These results showed that phonotactic restrictions in speakers' grammars affect how they categorize the speech signals.

Extending this work, some studies showed that some particular phonological hypotheses are psychologically real. For example, in Japanese, only foreign words, but not native or Sino-Japanese words, allow word-final long [aa] and singleton [p] (Itô and Mester 1995). Moreton and Amano (1999) showed that once listeners hear [p] in the stimuli, cueing foreignness of the stimuli, they are more likely to judge the word-final [a] as long [aa]. Gelbart and Kawahara (2007) extended this result and showed that, as long as real foreign words are presented, a similar bias towards allowing word-final long [aa] perceptually is observed, even in the absence of phonological cues to the lexical affiliation. See Gelbart (2005) for similar results from other languages.

On a slightly different line of research, acoustic studies have provided evidence for particular prosodic structures (Maddieson 1993; Broselow *et al.* 1997; Frazier 2006) or tonal representations (Morén and Zsiga 2006). Broselow *et al.* (1997) show that language-particular prosodic structures, each motivated in terms of stress placement, are manifested in different phonetic implementation patterns. Yet another line of research argues for the psychological reality of underspecification (Archangeli 1988). For example, a priming study by Lahiri and Reetz (2002) shows that labial and dorsal signals can activate coronal input. They argue that, assuming coronals are underspecified in the mental lexicon, all labial, coronal, and dorsal consonants can be matched up with underlying coronals (see also Lahiri and Marslen-Wilson 1991; CHAPTER 12: CORONALS). These studies aim to show that theoretical devices that have been proposed, such as lexical stratification, prosodic structure, or underspecification, may not merely be abstract theoretical constructs, but may be psychologically real, influencing our speech behaviors (see Goldrick, forthcoming, for discussion).

2.5 *Sources of phonological patterns*

Finally, many experiments have addressed the issue of sources of phonological patterns. This tradition has been most rigorously pursued by Ohala (e.g. 1983, 1990b), but has been taken up by many other researchers. For example, in many languages [kʲ] (or [k] before front vowels) changes diachronically into [ʧ]. The ubiquity of this sound change (and its synchronic correspondence) may be attributed to the acoustic affinity between [kʲ] and [ʧ] (Ohala 1989; Guion 1998; Chang *et al.* 2001; Wilson 2006). Raising of F2 via palatalization makes [k] sound similar to palatal consonants, and a long period of aspiration of dorsal [k] makes it sound similar to an affricate. A perception experiment by Guion (1998) demonstrated in fact that listeners often misperceive [kʲ] as [tʃ], showing that the sound change [kʲ] → [tʃ] may be due to acoustic similarity. Furthermore, Chang *et al.* (2001) point out that the directionality of the [kʲ] → [tʃ] change is rarely if ever reversed, and demonstrate experimentally that listeners may perceive [kʲ] as [tʃ], but not vice versa.

Another example is provided by the fact that in many languages a vowel must be long after a glide. Traditionally, this restriction has been analyzed as a case of compensatory lengthening: the first vowel in vowel sequences obtains a mora by a universal convention, loses its mora when it becomes a glide, with the floating mora being reassociated with the following vowel, resulting in a long vowel (Hayes 1989; see also CHAPTER 64: COMPENSATORY LENGTHENING). Myers and Hansen (2005) offer an alternative explanation: given a sequence of two vocoids, the boundary between them is blurry, and listeners may misattribute the gradient transition to the second vowel. The misattribution would result in a percept of long second vowels. Their perception experiments supported their hypothesis: the longer the transition duration, the more likely listeners judge the second vowel to be long (CHAPTER 20: THE REPRESENTATION OF VOWEL LENGTH).

The list of other experiments which have searched for the basis of phonological patterns would include, but is not limited to, the following: Kohler 1990; Ohala 1990a; Hura *et al.* 1992; Kawasaki-Fukumori 1992; Huang 2001; Hume and Johnson 2001; Barnes 2002; Mielke 2003; Kawahara 2006; Kochetov 2006; Myers and Hansen 2007) as well as those discussed in §3.1. See also Blevins (2004) and Ohala's other work (e.g. Ohala and Lorentz 1977; Ohala 1981, 1983) for further cases of phonetic origins of phonological patterns.

3 Experiments informed by phonology

So far I have been focusing on how experiments have informed phonological theories. However, the communication is by no means one-way. So we now turn our attention to how phonological observations and theories helped us design phonetic experiments and led to important discoveries.

3.1 Experiments motivated by phonological observations

As discussed in §2.5, many experiments have attempted to make sense of why certain phonological patterns occur. Put in a different perspective, this tradition has allowed us to reveal aspects of our phonetic systems by addressing why phonology works in the way that it does. To illustrate this point with another example, an influential tradition of this line of research is that of Adaptive Dispersion Theory, initiated by Lindblom and his colleagues (Liljencrants and Lindblom 1972; Lindblom 1986; Diehl *et al.* 2004) and pursued in a number of studies (e.g. Flemming 1995; Boersma 1998; Padgett 2002). This theory sets out to address why languages have the sets of vowels that they have. For example, languages that have three contrastive vowels usually have [a i u] rather than, say, [ə i ʌ], and languages that have five contrastive vowels have [a i u ɛ ɔ] rather than [ə ɨ ʌ ɤ ɪ]. The general idea is that speakers keep contrasting elements maximally (Liljencrants and Lindblom 1972) – or sufficiently (Lindblom 1986) – distinct from one another; this thesis has received support from experimental work (Engstrand and Krull 1994; Padgett 2002), as well as from corpus-based cross-linguistic analyses (Kingston 2007; see also CHAPTER 2: CONTRAST). This research tradition shows that an attempt to explain phonological patterns provides important insights into our speech behaviors. In this sense, taken together with the discussion in §2.5, phonological observations and phonetic experiments inform one another.

3.2 *Experiments motivated by phonological theories*

Not only can phonological observations lead to interesting phonetic hypotheses and experiments, specific phonological hypotheses can sometimes provide a guideline for where to look in experimental work. For example, in traditional analyses of Japanese intonation (Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988; Venditti 2005), Japanese was not thought to have an Intonational Phrase. Selkirk (2005), however, on the basis of cross-linguistic patterning, proposes a general theory of syntax–phonology mapping where clause edges should generally correspond to Intonational Phrase edges. Guided by this theory, Kawahara and Shinya (2008) investigate the intonational properties of clause edges in Japanese and find evidence for Intonational Phrase edges. In particular, they find that the left edges of clauses show larger initial rises and stronger pitch reset compared to VP edges, and clauses are also characterized by final lowering of tones, final creakiness and pause at their right edges. This work shows that theories can provide a guideline as to what to look for in phonetic studies.

Another example comes from articulatory studies on transparent segments in harmony contexts (Gick *et al.* 2006; Beňuš and Gafos 2007; Walker *et al.* 2008; CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS). Several authors have proposed that autosegmental spreading is strictly local and can never skip a segment (Ní Chiosáin and Padgett 1997; Gafos 1998; Walker 1998). Transparent segments in harmony patterns pose a problem for this theory because it looks as though these segments are "skipped." Recent articulatory studies have shown, however, that "transparent" segments also undergo harmony (e.g. tongue body backing in back vowel harmony in Hungarian and the tip blade gesture in Kinyarwanda consonant harmony), without causing much perceptual effect. This outcome is as predicted by strict locality, because transparent segments, too, undergo harmony phonologically. Again, the theory of strict locality has led to experiments that reveal a non-trivial aspect of transparent segments in harmony contexts. See Hayes (1999) for related discussion on theory-driven experiments.

3.3 *Testing specific phonological hypotheses and beyond*

Specific phonological hypotheses can motivate specific hypothesis testing, which has often resulted in further insights into the intricacy of the phonetics–phonology interface. To take one example, Steriade (1997, 2009) proposes that the less perceptible a phonological contrast is (in a particular context), the more likely it is to be neutralized. Some work has shown that at least some contrasts that are likely to be neutralized are indeed less perceptible than non-neutralizing contrasts (Kawahara 2006; Kochetov 2006; CHAPTER 80: MERGERS AND NEUTRALIZATION). Kochetov (2006) showed further, however, that not all differences in phonetic perceptibility are reflected in phonological patterns. Once again, Steriade's specific hypothesis about the interaction between phonetic perceptibility and phonological patterns motivated experimental testing, which revealed the complex interaction between phonetics and phonology.

To take another example, many languages require that lexical words be minimally bimoraic or bisyllabic (§2.3.3). Japanese, however, allows monomoraic lexical words, and Mori (2002) tested whether Japanese does indeed violate the minimality requirement. She found that, when monomoraic words are pronounced without a case particle, they undergo lengthening, while longer words do not show such lengthening. In this sense, Japanese does satisfy the minimal word requirement. However, she further found that lengthened monomoraic roots are not as long as bimoraic roots, instantiating a case of incomplete neutralization (§2.2.2).

To summarize, these experiments show that specific phonological hypotheses can inform experiments, which often in turn provide insight into the complex interaction between phonetics and phonology. The list of theories that have motivated specific experimentation includes the sonority sequencing principle (Broselow and Finer 1991), Optimality Theory's (Prince and Smolensky 1993) transitivity of constraints (Guest *et al.* 2000), the Emergence of the Unmarked (Broselow *et al.* 1998), and positional faithfulness theory (Kawahara and Shinohara 2010).

4 Summary

Phonetic and psycholinguistic experiments have contributed much to the development of phonological theories, and they will continue to do so. In (3) and (4) I summarize how experiments have informed phonological theories and vice versa.

(3) *How experiments inform theory*

a. Provide data beyond those available through introspection.
b. Re-examine the quality of phonological data.
c. Address questions about the architecture of the grammar.
d. Show and examine the psychological reality of the grammar.
e. Find the sources of phonological patterns.

(4) *How phonology informs experiments*

a. Helps to find restrictions on – and the nature of – speech through phonological patterns.
b. Provides a guide as to where and what to look for in phonetic experiments.
c. Motivates specific hypothesis testing.

ACKNOWLEDGMENTS

Thanks to the audiences at Kyoto University, members at Kyushu University, especially Toshio Matsuura, and my students in my graduate seminar in Spring 2009 and the graduate phonetics class in Spring 2010 at Rutgers University, especially Aaron Braver, Will Bennett, and Jimmy Bruno. Particular thanks are due to Osamu Fujimura, Beth Hume, Michael Kenstowicz, Marc van Oostendorp, Jaye Padgett, and two anonymous reviewers for their comments on earlier versions of this chapter, although they are not responsible for remaining errors. The preparation of this chapter was partially supported by a Research Council Grant from Rutgers University.

REFERENCES

Albright, Adam & Bruce Hayes. 2003. Rules *vs.* analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119–161.

Anderson, Stephen R. 1981. Why phonology isn't "natural." *Linguistic Inquiry* 12. 493–539.

Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5. 183–207.

Barnes, Jonathan. 2002. Positional neutralization: A phonologization approach to typological patterns. Ph.D. dissertation, University of California, Berkeley.

Barnes, Jonathan & Darya Kavitskaya. 2003. Phonetic analogy and schwa deletion in French. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 28. 39–50.

Becker, Michael, Nihan Ketrez & Andrew Nevins. 2008. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish devoicing neutralization. Unpublished ms., Reed College, Yale University & Harvard University (ROA-1001).

Beckman, Jill N. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14. 1–46.

Beckman, Mary E. & John Kingston. 1990. Introduction. In Kingston & Beckman (1990), 1–16.

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.

Beňuš, Štefan & Adamantios I. Gafos. 2007. Articulatory characteristics of Hungarian "transparent" vowels. *Journal of Phonetics* 35. 271–300.

Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104. 591–630.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.

Blaho, Sylvia. 2008. The syntax of phonology: A radically substance-free approach. Ph.D. dissertation, University of Tromsø.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns.* Cambridge: Cambridge University Press.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives.* The Hague: Holland Academic Graphics.

Broselow, Ellen & Daniel Finer. 1991. Parameter setting in second language phonology and syntax. *Second Language Research* 7. 35–59.

Broselow, Ellen, Su-I Chen & Marie Huffman. 1997. Syllable weight: Convergence of phonology and phonetics. *Phonology* 14. 47–82.

Broselow, Ellen, Su-I Chen & Chilin Wang. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20. 261–280.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Bybee, Joan. 1999. Usage-based phonology. In Darnell *et al.* (1999), 211–242.

Cena, R. M. 1978. When is a phonological generalization psychologically real? Bloomington: Indiana University Linguistics Club.

Chang, Steve S., Madelaine C. Plauché & John J. Ohala. 2001. Markedness and consonant confusion asymmetries. In Hume & Johnson (2001), 79–101.

Chen, Matthew Y. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22. 129–159.

Cho, Taehong. 2001. Effects of morpheme boundaries on gestural timing: Evidence from Korean. *Phonetica* 58. 129–162.

Chomsky, Noam. 1957. *Syntactic structures.* The Hague: Mouton.

Chomsky, Noam. 1986. Rules and representations. *Behavioral and Brain Sciences* 3. 1–15.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.
Coetzee, Andries W. 2008. Grammaticality and ungrammaticality in phonology. *Language* 84. 218–257.
Coetzee, Andries W., René Kager & Joe Pater (eds.) 2009. *Phonological models and experimental data.* Thematic issue, *Phonology* 26. 1–226.
Coetzee, Andries W., Susan Lin & Rigardt Pretorius. 2007. Post-nasal devoicing in Tswana. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences,* 861–867. Saarbrücken: Saarland University.
Cohn, Abigail C. 1993. Nasalisation in English: Phonology or phonetics. *Phonology* 10. 43–81.
Cohn, Abigail C. 1998. The phonetics–phonology interface revisited: Where's phonetics? *Texas Linguistic Forum* 41. 25–40.
Cohn, Abigail C. 2006. *Is there gradient phonology?* In Fanselow *et al.* (2006), 25–44.
Cohn, Abigail C. 2010. Laboratory Phonology: Past successes and current questions, challenges, and goals. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.) *Laboratory phonology 10,* 3–30. Berlin & New York: Mouton de Gruyter.
Darnell, Michael, Edith Moravscik, Michael Noonan, Frederick Newmeyer & Kathleen Wheatly (eds.) 1999. *Functionalism and formalism in linguistics.* 2 vols. Amsterdam & Philadelphia: John Benjamin.
Davidson, Lisa. 2006a. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34. 104–137.
Davidson, Lisa. 2006b. Schwa elision in fast speech: Segmental deletion or gestural overlap. *Phonetica* 63. 79–112.
Davis, Stuart. 2005. Capitalistic v. militaristic: The paradigm uniformity effect reconsidered. In Laura J. Downing, T. A. Hall & Renate Raffelsiefen (eds.) *Paradigms in phonological theory,* 107–121. Oxford: Oxford University Press.
de Jong, Kenneth J. 1998. Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics* 26. 283–310.
Diehl, Randy L. 1991. The role of phonetics within the study of speech. *Phonetica* 48. 120–134.
Diehl, Randy L. & Keith Kluender. 1989. On the objects of speech perception. *Ecological Psychology* 1. 121–144.
Diehl, Randy L., Björn Lindblom & Carl Creeger. 2004. Increasing realism of auditory representations yields further insights into vowel phonetics. In M. J. Solé, D. Recasens & J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences,* 1381–1384. Barcelona: Causal Productions.
Dinnsen, Daniel A. & Jan Charles-Luce. 1984. Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12. 49–60.
Dmitrieva, Olga. 2005. Incomplete neutralization in Russian final devoicing: Acoustic evidence from native speakers and second language learners. Poster presented at the 149th Meeting of the Acoustical Society of America.
Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25. 1568–1578.
Ellis, Lucy & W. J. Hardcastle. 2002. Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics* 30. 373–396.
Engstrand, Olle & Diana Krull. 1994. Durational correlates of quantity in Swedish, Finnish and Estonian: Cross-language evidence for a theory of adaptive dispersion. *Phonetica* 51. 80–91.
Ernestus, Mirjam & R. Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79. 5–38.

Ernestus, Mirjam & R. Harald Baayen. 2006. The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In Goldstein *et al.* (2006), 27–49.
Ernestus, Mirjam & R. Harald Baayen. 2007. Intraparadigmatic effects on the perception of voice. In Jeroen van der Weijer & Erik Jan van der Torre (eds.) *Voicing in Dutch*, 153–172. Amsterdam & Philadelphia: John Benjamins.
Fanselow, Gisbert, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.) 2006. *Gradience in grammar: Generative perspectives.* Oxford: Oxford University Press.
Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles.
Fougeron, Cécile & Donca Steriade. 1997. Does deletion of French schwa lead to neutralization of lexical distinctions? *Proceedings of the 5th European Conference on Speech Communication and Technology, University of Patras* 7. 943–946.
Fourakis, Marios & Gregory K. Iverson. 1984. On the "incomplete neutralization" of German final obstruents. *Phonetica* 41. 140–149.
Fourakis, Marios & Robert F. Port. 1986. Stop epenthesis in English. *Journal of Phonetics* 14. 197–221.
Fox, Robert A. & Dale Terbeek. 1977. Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics* 5. 27–34.
Frazier, Melissa. 2006. Output–output faithfulness to moraic stucture: Evidence from American English. *Papers from the Annual Meeting of the North East Linguistic Society* 36. 1–14.
Frisch, Stefan A., Nathan R. Large & David B. Pisoni. 2000. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42. 481–496.
Fromkin, Victoria A. & Robert Rodman. 1998. *An introduction to language*, 6th edn. Fort Worth: Harcourt Brace College.
Fujisaki, Hiroya & Hiroshi Sudo. 1971. A generative model for the prosody of connected speech in Japanese. *Annual Report, Engineering Research Institute, University of Tokyo* 30. 75–80.
Gafos, Adamantios I. 1998. Eliminating long-distance consonantal spreading. *Natural Language and Linguistic Theory* 16. 223–278.
Gelbart, Ben. 2005. Perception of foreignness. Ph.D. dissertation, University of Massachusetts, Amherst.
Gelbart, Ben & Shigeto Kawahara. 2007. Lexical cues to foreignness in Japanese. *MIT Working Papers in Linguistics* 55. 49–60.
Gick, Bryan, Douglas Pulleyblank, Fiona Campbell & Ngessimo Mutaka. 2006. Low vowels and transparency in Kinande vowel harmony. *Phonology* 23. 1–20.
Goldrick, Matthew. Forthcoming. Utilizing psychological realism to advance phonological theory. In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.) *The handbook of phonological theory*, 2nd edn. Malden, MA & Oxford: Wiley-Blackwell.
Goldstein, Louis, Douglas Whalen & Catherine T. Best (eds.) 2006. *Laboratory phonology 8.* Berlin & New York: Mouton de Gruyter.
Gouskova, Maria & Nancy Hall. 2009. Acoustics of epenthetic vowels in Lebanese Arabic. In Steve Parker (ed.) *Phonological argumentation: Essays on evidence and motivation*, 203–225. London: Equinox.
Gouskova, Maria, Elizabeth Zsiga & One Tlale. 2006. On the status of voiced obstruents in Tswana: Against *ND. *Papers from the Annual Meeting of the North East Linguistic Society* 36. 721–734.
Gow, David. 2002. Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* 28. 163–179.
Guest, Daniel J., Gary S. Dell & Jennifer S. Cole. 2000. Violable constraints in language production: Testing the transitivity assumption of Optimality Theory. *Journal of Memory and Language* 42. 272–299.

Guion, Susan. 1998. The role of perception in the sound change of velar palatalization. *Phonetica* 55. 18–52.

Hale, Mark & Charles Reiss. 2000. "Substance abuse" and "dysfunctionalism": Current trends in phonology. *Linguistic Inquiry* 31. 157–169.

Hall, Nancy. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23. 387–429.

Hawkins, John A. & Anne Cutler. 1988. Psycholinguistic factors in morphological asymmetry. In John A. Hawkins (ed.) *Explaining language universals*, 280–317. Oxford: Blackwell.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce. 1995. On what to teach the undergraduates: Some changing orthodoxies in phonological theory. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm 3*, 59–77. Seoul: Hanshin.

Hayes, Bruce. 1999. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In Darnell *et al.* (1999), vol. 1: *General papers*, 243–285.

Hayes, Bruce. 2000. Gradient well-formedness in Optimality Theory. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: Phonology, syntax, and acquisition*, 88–120. Oxford: Oxford University Press.

Hayes, Bruce & Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.

Hayes, Bruce, Kie Zuraw, Péter Siptár & Zsuzsa Cziráky Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85. 822–863.

Hsu, Chai-Shune. 1996. Laryngeal features in Taiwanese. Unpublished ms., University of California, Los Angeles.

Huang, Tsan. 2001. The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. *Ohio State University Working Papers in Linguistics* 55. 23–42.

Hume, Elizabeth & Keith Johnson (eds.) 2001. *The role of speech perception in phonology*. San Diego: Academic Press.

Hura, Susan L., Björn Lindblom & Randy L. Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35. 59–72.

Hyman, Larry M. 2001. The limits of phonetic determinism in phonology: *NC revisited. In Hume & Johnson (2007), 141–185.

Itô, Junko & Armin Mester. 1995. Japanese phonology. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 817–838. Cambridge, MA & Oxford: Blackwell.

Jaeger, Jeri J. 1978. Speech aerodynamics and phonological universals. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 4. 311–325.

Jassem, Wiktor & Lutosława Richter. 1989. Neutralization of voicing in Polish obstruents. *Journal of Phonetics* 17. 317–325.

Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82. 536–574.

Kawahara, Shigeto & Kazuko Shinohara. 2010. Phonetic and psycholinguistic prominences in pun formation: Experimental evidence for positional faithfulness. In William McClure & Marcel den Dikken (eds.) *Japanese/Korean linguistics 18*, 177–188. Stanford: CSLI.

Kawahara, Shigeto & Takahito Shinya. 2008. The intonation of gapping and coordination in Japanese: Evidence for Intonational Phrase and Utterance. *Phonetica* 65. 62–105.

Kawasaki-Fukumori, Haruko. 1992. An acoustical basis for universal phonotactic constraints. *Language and Speech* 35. 73–86.

Keating, Patricia. 1979. A phonetic study of a voicing contrast in Polish. Ph.D. dissertation, Brown University.

Keating, Patricia. 1985. Universal phonetics and the organization of grammars. In Victoria A. Fromkin (ed.) *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 115–132. Orlando: Academic Press.

Keating, Patricia. 1988a. The phonology–phonetics interface. In Frederick J. Newmeyer (ed.) *Linguistics: The Cambridge survey*, vol. 1, 281–302. Cambridge: Cambridge University Press.

Keating, Patricia. 1988b. Underspecification in phonetics. *Phonology* 5. 275–292.

Keating, Patricia. 1990. The window model of coarticulation: Articulatory evidence. In Kingston & Beckman (1990), 451–470.

Keating, Patricia. 1996. The phonology–phonetics interface. *Studia Grammatica* 41. 262–278.

Kingston, John. 2007. The phonetics–phonology interface. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 401–434. Cambridge: Cambridge University Press.

Kingston, John & Mary E. Beckman (eds.) 1990. *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press.

Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70. 419–454.

Kiparsky, Paul. 1982. From Cyclic Phonology to Lexical Phonology. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part I, 131–175. Dordrecht: Foris.

Kochetov, Alexei. 2006. Testing licensing by cue: A case of Russian palatalized coronals. *Phonetica* 63. 113–148.

Kohler, Klaus J. 1990. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. Hardcastle & A. Marchal (eds.) *Speech production and speech modelling*, 69–92. Dordrecht: Kluwer.

Kubozono, Haruo. 1988. The organization of Japanese prosody. Ph.D. dissertation, University of Edinburgh. Published 1993, Tokyo: Kurosio.

Lahiri, Aditi & William Marslen-Wilson. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38. 245–294.

Lahiri, Aditi & Henning Reetz. 2002. Underspecified recognition. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology* 7, 637–675. Berlin & New York: Mouton de Gruyter.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48. 839–862.

Lindblom, Björn. 1962. Phonetic aspects of linguistic explanation. *Studia Linguistica* 32. 137–153.

Lindblom, Björn. 1986. Phonetic universals in vowel systems. In Ohala & Jaeger (1986), 13–44.

Maddieson, Ian. 1993. Splitting the mora. *UCLA Working Papers in Phonetics* 83. 9–18.

Massaro, Dominic W. & Michael M. Cohen. 1983. Phonological context in speech perception. *Perception and Psychophysics* 34. 338–348.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague & Paris: Mouton.

McCawley, James D. 1986. Today the world, tomorrow phonology. *Phonology Yearbook* 3. 27–43.

Mielke, Jeff. 2003. The interplay of speech perception and phonology: Experimental evidence from Turkish. *Phonetica* 60. 208–229.

Morén, Bruce & Elizabeth C. Zsiga. 2006. The lexical and post-lexical phonology of Thai tones. *Natural Language and Linguistic Theory* 24. 113–178.

Moreton, Elliott. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84. 55–71.

Moreton, Elliott & Shigeaki Amano. 1999. Phonotactics in the perception of Japanese vowel length: Evidence for long-distance dependencies. In *Proceedings of the 6th European Conference on Speech Communication and Technology, Budapest*, 2679–2682.

Mori, Yoko. 2002. Lengthening of Japanese monomoraic nouns. *Journal of Phonetics* 30. 689–708.

Myers, Scott & Benjamin B. Hansen. 2005. The origin of vowel-length neutralisation in vocoid sequences: Evidence from Finnish speakers. *Phonology* 22. 317–344.

Myers, Scott & Benjamin B. Hansen. 2007. The origin of vowel-length neutralization in final position: Evidence from Finnish speakers. *Natural Language and Linguistic Theory* 25. 157–193.

Ní Chiosáin, Máire & Jaye Padgett. 1997. Markedness, segment realization, and locality in spreading. Technical Report LRC-97-01, Linguistics Research Center, University of California, Santa Cruz.

Nolan, Francis. 1992. The descriptive role of segments: Evidence from assimilation. In Gerard J. Docherty & D. Robert Ladd (eds.) *Papers in laboratory phonology II: Gesture, segment, prosody*, 261–280. Cambridge: Cambridge University Press.

Nooteboom, Sieb G. 1981. Lexical retrieval from fragments of spoken words: Beginnings *vs.* endings. *Journal of Phonetics* 9. 407–424.

Ohala, John J. 1974. Experimental historical phonology. In John M. Anderson & Charles Jones (eds.) *Historical linguistics*, vol. 2: *Theory and description in phonology*, 353–389. Amsterdam: North-Holland.

Ohala, John J. 1981. The listener as a source of sound change. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 17(2). 178–203.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.) *The production of speech*, 189–216. New York: Springer.

Ohala, John J. 1986a. Consumer's guide to evidence in phonology. *Phonology Yearbook* 3. 3–26.

Ohala, John J. (ed.) 1986b. The validation of phonological theories. Thematic issue. *Phonology Yearbook* 3. 1–252.

Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In Leiv Egil Breivik & Ernst Håkon Jahr (eds.) *Language change: Contributions to the study of its causes*, 173–198. Berlin: Mouton de Gruyter.

Ohala, John J. 1990a. The phonetics and phonology of aspects of assimilation. In Kingston & Beckman (1990), 258–275.

Ohala, John J. 1990b. There is no interface between phonology and phonetics: A personal view. *Journal of Phonetics* 18. 153–171.

Ohala, John J. 1999. The relation between phonetics and phonology. In W. J. Hardcastle & John Laver (eds.) *The handbook of phonetic sciences*, 674–694. Oxford & Cambridge, MA: Blackwell.

Ohala, John J. & Jeri J. Jaeger (eds.) 1986. *Experimental phonology*. Orlando: Academic Press.

Ohala, John J. & James Lorentz. 1977. The story of [w]: An exercise in the phonetic explanation for sound patterns. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 3. 577–599.

Oostendorp, Marc van. 2008. Incomplete devoicing in formal phonology. *Lingua* 118. 1362–1374.

Padgett, Jaye. 2002. Contrast and postvelar fronting in Russian. *Natural Language and Linguistic Theory* 21. 39–87.

Padgett, Jaye & Marija Tabain. 2005. Adaptive Dispersion Theory and phonological vowel reduction in Russian. *Phonetica* 62. 14–54.

Pertz, D. L. & T. G. Bever. 1975. Sensitivity to phonological universals in children and adolescents. *Language* 51. 149–162.

Pierrehumbert, Janet B. 2006. The statistical basis of an unnatural alternation. In Goldstein *et al.* (2006), 81–107.

Pierrehumbert, Janet B. & Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.

Pitt, Mark A. 1998. Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception and Psychophysics* 60. 941–951.
Port, Robert F. & Adam P. Leary. 2005. Against formal phonology. *Language* 81. 927–964.
Port, Robert F. & Michael O'Dell. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13. 455–471.
Port, Robert F., Salman Al-Ani & Shosaku Maeda. 1980. Temporal compensation and universal phonetics. *Phonetica* 37. 235–252.
Poser, William J. 1984. The phonetics and phonology of tone and intonation in Japanese. Ph.D. dissertation, MIT.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.
Pulleyblank, Douglas. 1988. Vocalic underspecification in Yoruba. *Linguistic Inquiry* 19. 233–270.
Pycha, Anne. 2009. Lengthened affricates as a test case for the phonetics–phonology interface. *Journal of the International Phonetic Association* 39. 1–31.
Riehl, Anastasia K. 2003. American English flapping: Perceptual and acoustic evidence against paradigm uniformity with phonetic features. *Working Papers of the Cornell Phonetics Laboratory* 15. 271–337.
Sanders, Nathan. 2001. Preserving synchronic parallelism: Diachrony and opacity in Polish. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 37. 501–516.
Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
Selkirk, Elisabeth. 2005. Comments on intonational phrasing in English. In Sónia Frota, Marina Vigário & Maria João Freitas (eds.) *Prosodies: With special reference to Iberian languages*, 11–58. Berlin & New York: Mouton de Gruyter.
Shinohara, Shigeko. 2004. Emergence of Universal Grammar in foreign word adaptations. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in phonological acquisition*, 292–320. Cambridge: Cambridge University Press.
Slowiaczek, Louisa M. & Daniel A. Dinnsen. 1985. On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13. 325–341.
Slowiaczek, Louisa M. & Helena J. Szymanska. 1989. Perception of word-final devoicing in Polish. *Journal of Phonetics* 17. 205–212.
Solé, Maria-Josep, Patrice Speeter Beddor & Manjari Ohala (eds.) 2007. *Experimental approaches to phonology*. Oxford: Oxford University Press.
Sproat, Richard & Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21. 291–311.
Steriade, Donca. 1995. Positional neutralization. Unpublished ms., University of California, Los Angeles.
Steriade, Donca. 1997. Phonetics in phonology: The case of laryngeal neutralization. Unpublished ms., University of California, Los Angeles.
Steriade, Donca. 2000. Paradigm uniformity and the phonetics–phonology boundary. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, 313–334. Cambridge: Cambridge University Press.
Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Studies in honor of Paul Kiparsky*, 151–179. Cambridge, MA: MIT Press.
Stevens, Kenneth N., Samuel J. Keyser & Haruko Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In Joseph S. Perkell & Dennis H. Klatt (eds.) *Invariance and variability in speech processes*, 426–449. Hillsdale, NJ: Lawrence Erlbaum.

Strange, Winifred, Ocke-Schwen Bohn, Sonja A. Trent & Kanae Nishi. 2004. Acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America* 115. 1791–1807.
Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.
Tsuchida, Ayako. 1997. Phonetics and phonology of Japanese vowel devoicing. Ph.D. dissertation, University of Cornell.
Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany: State University of New York Press.
Venditti, Jennifer. 2005. The ToBI model of Japanese intonation. In Sun-Ah Jun (ed.) *Prosodic typology: The phonology of intonation and phrasing*, 172–200. Oxford: Oxford University Press.
Walker, Rachel. 1998. Nasalization, neutral segments, and opacity effects. Ph.D. dissertation, University of California, Santa Cruz.
Walker, Rachel, Dani Byrd & Fidèle Mpiranya. 2008. An articulatory view of Kinyarwanda coronal harmony. *Phonology* 25. 499–535.
Warner, Natasha, Allard Jongman, Joan A. Sereno & Rachèl Kemps. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32. 251–276.
Warner, Natasha, Erin Good, Allard Jongman & Joan A. Sereno. 2006. Orthographic *vs.* morphological incomplete neutralization effects. *Journal of Phonetics* 34. 285–293.
Westbury, John R. 1979. Aspects of the temporal control of voicing in consonant clusters in English. Ph.D. dissertation, University of Texas, Austin.
Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30. 945–982.
Yu, Alan C. L. 2007. Understanding near mergers: The case of morphological tone in Cantonese. *Phonology* 24. 187–214.
Zimmer, Karl. 1969. Psychological correlates of some Turkish morpheme structure conditions. *Language* 45. 309–321.
Zsiga, Elizabeth C. 1995. An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. In Bruce Connell & Amalia Arvaniti (eds.) *Papers in laboratory phonology IV: Phonology and phonetic evidence*, 282–302. Cambridge: Cambridge University Press.
Zuraw, Kie. 2000. Patterned exceptions in phonology. Ph.D. dissertation, University of California, Los Angeles.
Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog infixation. *Language* 83. 277–316.

97 Tonogenesis

JOHN KINGSTON

1 Introduction

Tones have frequently appeared and multiplied under the influence of the laryngeal articulations of adjacent consonants (§2–§4), but much more rarely under the influence of the vowels that bear them (§6). The laryngeal articulations of following consonants readily introduce tones into previously toneless languages, but those of preceding consonants far more often split existing tones than introduce them where there were none before (§5). Opposite tones can arise from the same ostensible phonological source, apparently because the source can be pronounced so as to raise or lower F0 (§3.2.2 and §4). Finally, tone can arise from prosodic as well as segmental sources (§7). In all these sound changes, a predictable or redundant F0 difference becomes contrastive – i.e. is "phonologized" – once the contrast that introduced it neutralizes. These changes show how easily languages blur the distinction between phonetics and phonology and undermine their autonomy over time.

2 From consonants to tone

2.1 *Introduction*

The first two examples, Yabem and Korean (§2.2), show how tone can remain redundant on phonation contrasts in preceding consonants, and the next two, Kammu and Cham (§2.3), show how the contrast transfers from phonation contrasts to tone. This section closes with discussion of how redundant F0 differences are phonologized (§2.4), why voiced obstruents lower F0 (§2.5), how consonants' influence on F0 is represented phonologically (§2.6), and whether a consonant's phonological specification or its phonetic properties determine its tonogenetic effects (§2.7).

2.2 *Yabem and Korean: Contrastive phonation and redundant tone*

In Yabem, an Oceanic Austronesian language of the North Huon Gulf (Bradshaw 1979; Ross 1993), and in Korean (Jun 1996; Silva 2006; Kang and Guion 2008), tone remains redundant on phonation contrasts in preceding consonants (CHAPTER 2: CONTRAST).

2.2.1 *Yabem*

In Yabem (Ross 1993), tonogenesis is the product of tone–voicing harmony. The weak syllable of a two-syllable iambic foot acquired the tone and voicing of the strong syllable when the strong syllable began with a stop in Proto-Huon Gulf (PHG), and a strong syllable not beginning with a stop acquired the tone and voicing of the weak syllable when it began with a voiced stop in PHG,[1] as illustrated by the alternations in the three realis/irrealis verb paradigms in Table 97.1:

Table 97.1 Realis and irrealis paradigms of three Yabem verb stems. The irrealis morpheme is realized by prenasalizing the first voiced stop in the root, if it has one, but otherwise by selecting distinct allomorphs of the singular subject prefixes

	-dèŋ 'move towards'		-táŋ 'weep'		-luʔ 'vomit'	
	realis	*irrealis*	*realis*	*irrealis*	*realis*	*irrealis*
1 sg	gà-dèŋ	jà-ⁿdèŋ	ká-táŋ	já-táŋ	gà-lùʔ	já-lúʔ
2 sg	gò-dèŋ	ò-ⁿdèŋ	kó-táŋ	ó-táŋ	gò-lùʔ	ó-lúʔ
3 sg	gè-dèŋ	è-ⁿdèŋ	ké-táŋ	é-táŋ	gè-lùʔ	é-lúʔ
1 pl incl	dà-dèŋ	dà-ⁿdèŋ	tá-táŋ	tá-táŋ	tá-lúʔ	tá-lúʔ
1 pl excl, 2 pl	à-dèŋ	à-ⁿdèŋ	á-táŋ	á-táŋ	á-lúʔ	á-lúʔ
3 pl	sè-dèŋ	sè-ⁿdèŋ	sé-táŋ	sé-táŋ	sé-lúʔ	sé-lúʔ

The voiced and voiceless stems /-dèŋ/ and /-táŋ/ show the covariation of stop voicing and tone within the strong syllable of the foot. No tone is marked on /-luʔ/ 'vomit', because it varies between low when the prefix stop is voiced (after realis singular [gà-, gò-, gè-]), and high otherwise (after 1st plural inclusive [tá-] and 3rd plural [sé-] and in all irrealis forms). As this stem is high after prefixes that do not contain stops, high must be the default tone value for tone, with low occurring only when a voiced stop appears somewhere in the foot. Neither voicing nor its absence can be the default value, as alternations occur in both directions: 1st singular /gà-/ → [ká-] and 1st plural inclusive /tá-/ → [dà-]. These observations indicate that voicing is the source of low tone (as well as voicing any stop in a weak syllable's onset), and that high tone and the absence of voicing developed elsewhere.

In other examples, tone is contrastive, e.g. /áwé/ 'outside' *vs.* /àwè/ 'woman' and /óli/ 'body' *vs.* /òlì/ 'wages'. The source of the tones in these and similar

[1] Generally, Proto-Oceanic (PO) *b, *g > PHG *b, *g > Proto-North Huon Gulf (PNHG) *b, *g > Yabem /b g/ and low tone, and in some morphemes, PO *p, *k unpredictably voiced and lenited to PHG *v, *ɣ > PNHG zero reflexes in Yabem with low tone.

examples is apparently the PHG *v, *ɣ, which derive via voicing and lenition from PO **p, **k.

2.2.2 *Korean*

Following an aspirated [spread glottis] or tense [constricted glottis] stop (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION), the initial L of the accentual phrase's LHL melody found after a laryngeally unspecified lax stop is raised to H in Seoul and Chonnam Korean (Jun 1996) – elsewhere in the word, the F0 elevation is smaller and shorter lasting. Silva (2006) and Kang and Guion (2008) argue that tone is becoming contrastive after consonants that are not [constricted glottis], as the duration of aspiration in lax stops has lengthened in younger people's speech to match that of aspirated stops.

2.3 *Kammu and Cham: Contrastive tone and loss of phonation contrasts*

Kammu and Cham illustrate the transfer from phonation to tone contrasts.

2.3.1 *Kammu*

Table 97.2 shows that in two western dialects of the Mon-Khmer language, Kammu, low tones ("à") appear where syllables begin with voiced stops and sonorants in the eastern dialect and a higher tone (high "á" or rising-falling "â") where those

Table 97.2 Correspondences of tone and register in three western dialects of Kammu to the phonation contrast on initial stops (a) and sonorants (b) in the eastern dialect. Examples from Suwilai (2003). E. Kammu = Svantesson's (1989) "Southern Kammu"; the first tonal W. Kammu dialect = his "Northern Kammu"; and the register dialect of W. Kammu resembles his "Lamet"

	E. Kammu	*W. Kammu Tone 1*	*W. Kammu Tone 2*	*W. Kammu Register*	
a.	buːc	pùːc	pʰùːc	pṳːc	'rice wine'
	puːc	pûːc	púːc	pûːc	'to take off clothes'
	bok	pòk	pʰòk	po̤k	'to cut down a tree'
	pok	pók	pók	pók	'to take a bite'
	buːm	pùːm	pʰùːm	pṳːm	'to chew'
	puːm	pûːm	púːm	pûːm	'to fart'
	glaːŋ	klàːŋ	kʰlàːŋ	kla̤ːŋ	'stone'
	klaːŋ	klâːŋ	kláːŋ	klâːŋ	'eagle'
	ɟaŋ	càŋ	cʰàŋ	ca̤ŋ	'to weigh'
	caŋ	câŋ	cáŋ	câŋ	'astringent'
b.	ŋɔʔ	ŋɔ̀ʔ	ŋɔ̀ʔ	ŋɔ̤ʔ	'to fear'
	ʰŋɔʔ	ŋɔ̂ʔ	ŋɔ́ʔ	ŋɔ́ʔ	'paddy rice'
	raːŋ	ràːŋ	ràːŋ	ra̤ːŋ	'flower'
	ʰraːŋ	râːŋ	ráːŋ	ráːŋ	'tooth'
	waʔ	wàʔ	wàʔ	wa̤ʔ	'to chase'
	ʰwaʔ	wáʔ	wáʔ	wáʔ	'monkey'

syllables begin with voiceless stops and sonorants; the other western dialect shows a contrast in voice quality (also known as "register") instead (Svantesson 1989; Suwilai 2003). Whether the higher tone is high or rising–falling in the first tonal dialect depends on whether the vowel is short or long.

In both tonal dialects, voiceless sonorants have become voiced, and in the first, voiced stops have become voiceless unaspirated. These mergers *transfer* the contrast from phonation in the initial consonant to tone in the following vowel (cf. Hyman 2008 on "trans-phonologization"). In the second tonal dialect, the voiced stops of the eastern dialect have merged instead with the voiceless aspirated stops and thereby also transferred the contrast to tone. In the register dialect, a clear or tense voice quality and high tone ("ā" or "á") corresponds to voiceless initials in the non-tonal dialect, while a breathy voice quality and low pitch ("a̤") corresponds to voiced initials.[2] In this dialect, too, the contrast has been transferred completely from the initial consonant to the following vowel. Tone is rising–falling or high after [spread glottis] consonants (= voiceless aspirated stops, /s h/) and [constricted glottis] consonants (= /ʔj ʔw ɓ ɗ/ – the implosives correspond to glottalized nasals /ʔm ʔn/ in the non-tonal dialect). Surprisingly, tone is low following /ʔ/ itself.

2.3.2 *Cham*

Two Chamic languages, Eastern Cham (Phu *et al.* 1992) and Utsat (Maddieson and Pang 1993), have also developed tone from preceding consonants (see also Thurgood 1999; but cf. Brunelle 2005, for arguments that Eastern Cham is not tonal), while Western Cham has developed a register contrast instead. These developments are much like those observed in the Western Kammu dialects, except that Proto-Cham did not distinguish sonorants for phonation. Low tone emerged after voiced sonorants, unless the syllable ended in glottal stop, in which case high tone emerged.

2.4 *Exaggeration and then transfer*

In Yabem and Korean, the F0 differences remain largely predictable from the phonation differences in the consonants, yet they are exaggerated in both size and extent in both languages, so they cannot still be described as mere phonetic perturbations. The tonal Western Kammu dialects, Eastern Cham, and Utsat have taken the next step and transferred the contrast from these exaggerated but still redundant F0 differences to tone by no longer pronouncing at least some of the preceding consonants with distinct phonation. Jun (1996) shows that exaggerated F0 differences may co-exist with phonation differences, so exaggeration of one phonetic difference does not necessarily trade off at first with diminution of another,[3] but Silva (2006) and Kang and Guion (2008) show that such a trade-off may eventually occur and thereby transfer the contrast.

2 "Register" refers to breathy or lax *vs.* tense, clear or modal voice quality, low *vs.* high F0, and higher *vs.* lower vowel qualities, which may occur singly or in combination (Gregerson 1976; Huffman 1976; Thongkum 1987; Denning 1989). In many Mon-Khmer languages, the first value appears in words that began with voiced obstruents and sonorants in an earlier form of the language, while the second appears in words that began with voiceless consonants. The first values may all be concomitants of expanding the pharynx by advancing the tongue root and lowering the larynx (Lindau 1979; Edmondson and Esling 2006).

3 Phonation properties vary in their probability of occurrence, not in their value; for example, a stop may occur more or less often with aspiration but not with a greater or lesser degree.

2.5 The phonetics of low tone from voiced stops

In Yabem, Western Kammu, Eastern Cham, and Utsat, low tones developed on vowels following voiced stops and non-low tones elsewhere. These tones could have arisen from exaggerating the F0 lowering that is an automatic phonetic side-effect of an articulation whose purpose is to resolve the aerodynamic conflict between voicing and obstruency. On the one hand, producing the noise characteristic of obstruents requires that intraoral air pressure (P_o) rise, but, on the other hand, maintaining voicing requires that P_o remain enough below subglottal air pressure (P_s) that air continues to flow up through the glottis. An articulation that both maintains the pressure drop across the glottis and also lowers F0 is larynx lowering (Hombert 1978; Hombert *et al.* 1979), which slows the rise in P_o by enlarging the oral cavity (Westbury 1983) and slackens the vocal folds by tilting the cricoid cartilage forward relative to the thyroid cartilage (Honda *et al.* 1999). This slackening could be an automatic, unintended mechanical consequence of larynx lowering, because this articulation's purpose in voiced obstruents is to expand the oral cavity and keep the rise in P_o in check. Yet Honda *et al.* (1999), as well as Collier (1975), Ewan (1976), and Erickson *et al.* (1994), show that larynx lowering is also one of the maneuvers which speakers use to lower F0 deliberately. Their results suggest that larynx lowering in a voiced obstruent could also be intended to lower F0.

Why would a speaker wish to lower F0 when producing a voiced obstruent? Kingston and Diehl (1994, 1995) and Kingston *et al.* (2008) argue that a low F0 in vowels flanking a voiced stop integrates perceptually with voicing in the stop closure itself to enhance the percept of the presence of low-frequency energy in and near the stop (see also Stevens and Blumstein 1981). The percept of low-frequency energy (among other auditory qualities) rather than either voicing or low F0 individually may be what conveys that the stop is [voice]. Integration unifies the stop's phonation and the adjacent vowel's pitch phonetically, at a stage in perception between the signal's raw acoustics and the listener's recognition of the stop's value for an abstract distinctive feature [voice]. This line of reasoning suggests that the lower F0 next to voiced obstruents may be as deliberate a consequence of larynx lowering as keeping P_o from rising too fast.

This account extends naturally to the register differences that have developed from the [voice] contrast in Western Kammu, Lamet, Western Cham, and elsewhere (see also Henderson 1967, Matisoff 1973, Denning 1989, and Thurgood 2002 for recognition of this link). In breathy voice, the first harmonic is far more intense than higher harmonics compared to modal or tense voice, and in higher vowels, the first formant (F1) is lower than in lower vowels. While both differences can be interpreted as automatic concomitants of producing voicing in an adjacent stop, their exaggeration and phonologization in the development of register contrasts in these languages can equally easily be construed as an attempt to strengthen the percept of low-frequency energy next to the stop.[4]

[4] F0 is usually lower in breathy voice, as in tonal Western Kammu and Eastern Cham compared to registral Western Kammu and Western Cham. Obstruents that are sometimes described as breathy voiced also "depress" tone in many Southern Bantu languages (Rycroft 1980, 1983; Traill *et al.* 1987; Traill 1990; Cassimjee 1998; Cassimjee and Kisseberth 1998; Donnelly 2009). However, the depressors are often no longer breathy or even voiced (Schachter 1976; Traill 1990; Jessen and Roux 2002; Maddieson 2003; Strazny 2003; Downing and Gick 2005; Downing 2009), so depression has become autonomous from the original laryngeal articulation that first introduced it into the languages (Traill *et al.* 1987; Traill 1990; Maddieson 2003), like other transfers of contrast from phonation to tone.

2.6 *Stiff and slack*

Halle and Stevens (1971) propose a different unification, which substitutes features used to represent tone contrasts in vowels, [stiff] and [slack] vocal folds, for [voice] in obstruents (see also Löfqvist *et al.* 1989 for phonetic evidence, and Avery and Idsardi 2001 for a dramatic extension of Halle and Stevens's proposals). A stop specified for [slack] vocal folds would be pronounced with voicing and would lower F0 on the following vowel, while one specified for [stiff] would be pronounced without voicing and with higher F0 on the following vowel. Voicing in obstruents is thus a side-effect of the vocal folds being slack enough.

The reduction in vocal fold tension represented by [slack] both helps and hinders keeping P_o enough below P_s: slacker folds vibrate for a smaller pressure drop across the glottis than stiffer ones, but they also let a larger volume of air pass through the glottis per unit of time and thereby accelerate the rise in P_o. Speakers apparently rely instead on slowing the rise in P_o by expanding the oral cavity both actively (Bell-Berti and Hirose 1975; Westbury 1983) and passively (Ohala and Riordan 1979). Halle and Stevens's features do not therefore abstract away from the phonetic realization of the phonation and tone contrasts so much as subsume them under a single contrast, whose phonetic realization differs systematically as a function of the segment's value for other features (cf. Kingston and Diehl 1994; see also CHAPTER 17: DISTINCTIVE FEATURES).

The Korean facts show that obstruents with other phonation types than [voice] raise rather than lower F0. Jun (1996) attributes this difference to the [spread glottis] and [constricted glottis] stops also being [stiff], apparently as a byproduct of stronger thyroarytenoid (also known as vocalis) contraction compared to the lax stops – cricothyroid contraction does not differ between the three classes of stops (Hirose *et al.* 1974; Hirose *et al.* 1981). The differences in the amount and timing of thyroarytenoid contraction are intended to contribute to differences between these stops in the glottal opening's size and timing (see also Kagaya 1974) and in the [constricted glottis] stops to the firmness with which the vocal folds are pressed against one another, but they clearly affect F0 too.

2.7 *Phonetics or phonology?*

The discussion in §2.5 treats F0 lowering next to voiced obstruents as one of the intended correlates of an abstract [voice] contrast. F0 raising next to [spread glottis] and [constricted glottis] could be treated as a similarly deliberate correlate of these contrasts. Yet we have seen two clear instances of tonogenesis from phonetically predictable rather than contrastive properties, namely, predictably voiced sonorants induced a low tone in the tonal Western Kammu dialects and in Eastern Cham, and the predictably spread glottis voiceless fricative [s] induced a high tone in the tonal Western Kammu dialects. The low tone that emerged after voiced sonorants in the tonal Western Kammu dialects could be attributed to their contrasting with voiceless sonorants in their ancestor (and the Eastern Kammu dialect). However, sonorants did not contrast for [voice] in Proto-Cham, so the low tone that emerged in Eastern Cham after voiced sonorants cannot derive from a correlate of the [voice] contrast in that language. Similarly, voiceless fricatives do not contrast for [spread glottis] with voiced ones in Western Kammu. The participation of voiced sonorants and /s/ suggest that their tonogenetic effects are determined by their being pronounced with voicing or the glottis wide open (Löfqvist and Yoshioka 1980; Yoshioka *et al.* 1981) and not by their phonological specification (see also CHAPTER 8: SONORANTS).

The quandary here is this: if the F0 differences next to consonants contrasting the phonation features [voice], [spread glottis], and [constricted glottis] are the product of articulations that are controlled so as to produce those differences, why do the differences also appear next to consonants that do not contrast for any of these features? Their appearance next to this latter group of consonants suggests that they are instead automatic consequences of voicing, opening the glottis wide, etc., and that it is only the articulations which produce these properties that are controlled. Producing voicing in a sonorant, however, does not require any articulation such as larynx lowering to keep the rise in P_o in check, because no appreciable amount of air is trapped inside the oral cavity during their pronunciation. In the absence of such an articulation, F0 would also not be automatically lowered next to sonorants, however reliably voiced they would be. These facts suggest that voiced sonorants pattern with voiced obstruents because voicing makes them phonetically similar, and that F0 is deliberately lowered next to them to enhance this phonetic similarity.

The case of voiceless fricatives apparently differs: the glottis must be opened wide to provide a volume of airflow through the oral constriction downstream sufficient to produce turbulent, noisy airflow. Opening the glottis wide in voiceless aspirated stops is presumably also intended to produce turbulent, noisy airflow after the stop release, so both non-contrastive and contrastive glottal spreading have a common purpose. Even so, glottal spreading does not automatically raise F0 (Hombert *et al.* 1979; Kingston and Diehl 1994), and to do so would require deliberate stiffening of the folds next to both classes of consonants. In short, what looked like evidence that the F0 differences must be automatic turns out to be evidence that they must instead be controlled.

3 Tonogenesis from following consonants

The strikingly similar tonogenetic behavior of following consonants in Vietnamese and Chinese (§3.1) is presented first, followed by the discussion of Athabaskan (§3.2). The Athabaskan data are the first to show that opposite tones can evolve from the same ostensible phonological source (§3.2.2).

3.1 Vietnamese and Chinese

According to Haudricourt (1954b), the tone contrasts of present-day Vietnamese developed in two stages. First, a final stop induced a rising tone, a final voiceless fricative a falling tone, and a level tone developed in other syllables, which either were open or ended in a nasal (see the columns in Table 97.3). A considerable time later, each of these tones then split in two under the influence of a [voice] contrast in syllable-initial stops, with words beginning with voiced stops developing lower pitches than those beginning with voiceless ones (see the rows in Table 97.3). The resulting six-way tone contrast displayed in Table 97.3 can be observed today in the Vietnamese spoken in the northern part of Vietnam.

Thurgood (2002) argues that the tones attributed to final and initial stops in Vietnamese did not arise directly from the stops themselves but instead from differences in voice quality that perturbed F0. The distinction between a rising

Table 97.3 Tonogenesis in Vietnamese: "k" = any following stop. The italicized words are the Vietnamese names for each of the tones

		Following consonants		
		CV,CVN	CVk, CVʔ	CVs, CVh
Preceding consonants	voiceless	*pa > pa high level (*ngang*)	*pak > pak high rising (*sac*)	pas > pa high falling (*hoi*)
	voiced	*ba > pa low level (*huyen*)	*bak > pak low rising (*nang*)	bas > pa low falling (*nga*)

and a level tone initially arose from a contrast between creaky and modal voice in syllables not ending in stops or fricatives in Proto-Vietic (Diffloth 1989). Tones in syllables ending in stops subsequently merged with those arising in creaky voice syllables, because final stops were pronounced with a simultaneous glottal closure. Similarly, the subsequent tone split arose from an earlier difference between a breathy and a modal voice quality induced by preceding voiced and voiceless stops and not directly from the [voice] contrast itself. Only the falling tones that developed in syllables that ended in voiceless fricatives are attributed directly to the consonants themselves (Thurgood 2002: 336).[5]

Haudricourt (1954a), Pulleyblank (1962), Mei (1970), Baxter (1982), and Sagart (1999) propose that tones developed in much the same way in the evolution from Old Chinese (500 BCE) to Middle Chinese (500 CE). Between Old and Middle Chinese, the *ping* 'level' or A tone arose in syllables ending in vowels or sonorants, the *shang* 'rising' or B tone in syllables ending in a glottal stop, and the *qu* 'departing' or C tone in syllables originally ending in /s/, which had become /h/ by the time the tones developed. A fourth tone, *ru* 'entering' or D, arose in syllables ending in an oral stop. Between Middle Chinese and the present-day languages, these tones split into lower and higher reflexes when the voiced stops merged with the voiceless unaspirated stops. There is little disagreement about this second stage in the development of the present-day tones, but considerable controversy about the transition from Old to Middle Chinese; for alternative analyses, see Wang (1958), Benedict (1972), and Ballard (1988). Lack of space prevents me from doing more here than mention the fact of this controversy.

3.2 *Athabaskan*

The Athabaskan languages in northwestern Canada, southeastern Alaska, and the Apachean subgroup are tonal, while those in southwestern Alaska and along the Pacifiic Coast of Oregon and northern California are not (Leer 1979, 1999, 2001; Kingston 2005; Krauss 2005). Most of the latter retain a contrast between glottalic and non-glottalic stops at the ends of stems that has been replaced by tones in the tonal languages.

[5] The Tamang languages, a group of closely related Tibeto-Burman languages spoken in Nepal, synchronically illustrate the covariation of tone, voice quality, and voicing in preceding obstruents that was the hypothesized precursor to present-day Vietnamese (Mazaudon 1978, forthcoming; Mazaudon and Michaud 2008).

3.2.1 *Stem rime contents*

Three contrasts between stem-final consonants (1a–c) and two contrasts between stem nuclei (1d–e) interacted in tonogenesis:

(1) a. Glottalic *vs.* non-glottalic stops (K′ *vs.* K), sonorants (′R *vs.* R), and fricatives (′X *vs.* X)
 b. Stops *vs.* sonorants (K *vs.* R)
 c. Stops *vs.* fricatives (K *vs.* X)
 d. Full *vs.* reduced (or long *vs.* short) nuclei (VV and Vʔ *vs.* V)
 e. Full vowel nuclei ending with glottal constriction (Vʔ) *vs.* those not ending with glottal constriction (VV).

(The difference in the position of the glottalic diacritic (K′ *vs.* ′R and ′X) is explained below.)

If a stem rime lacked a coda consonant, the nucleus had to be full (VV or Vʔ), but reduced as well as full vowel nuclei occurred in rimes ending in a consonant (VC, VVC). A rime ending in a glottalic sonorant or fricative with a non-glottalic full vowel nucleus did not contrast with one with a glottalic full vowel nucleus, i.e. VV′R and VV′X = Vʔ′R and Vʔ′X. Otherwise, all possible combinations of rimes and nuclei occurred. Anticipating §3.2.2, the tones are referred to as the "marked" and "unmarked" (see also CHAPTER 4: MARKEDNESS).

When the vowel was reduced, marked tone developed in stems ending in a glottalic stop or sonorant (VK′, V′R), while unmarked tone developed in those ending in a non-glottalic stop or sonorant (VK, VR). When the vowel was full and non-glottalic, however, marked tone only developed in stems ending in glottalic sonorants (VV′R); otherwise, the unmarked tone developed (VVK′ as well as VVK, VVR). The glottalic articulation was thus lost without a trace in VVK′ stems in the tonal languages. Marked tone also developed in all stems with a glottalic full vowel (Vʔ, VʔK, VʔK′). Glottalic fricatives are derived in certain stem allomorphs by spirantizing glottalic stops (VK′ > V′X, VVK′ > VV′X), and they behave like glottalic sonorants in producing marked tone on preceding full as well as reduced vowel nuclei in the tonal languages.

Much as Thurgood (2002) did for Vietnamese (§3.1), Leer (1979, 1999) argued that marked tone did not develop directly from any glottalic articulation on a stem-final consonant; glottalic consonants instead introduced a voice quality he called "constriction" on the preceding vowel, which was subsequently replaced by the marked tone. As in the ancestor of Vietnamese, some stem nuclei already contrasted for constriction in Proto-Athabaskan, i.e. *Vʔ *vs.* *VV, and stems ending in glottalic consonants merged tonogenetically with those whose nuclei were constricted.

Kingston (1985, 1990, 2005) argues that sonorants and fricatives behaved differently from stops because they lacked a stop burst with which the laryngeal articulation could be coordinated (CHAPTER 8: SONORANTS; CHAPTER 28: THE REPRESENTATION OF FRICATIVES). The laryngeal articulation in sonorants and fricatives could therefore shift more easily to the beginning of the consonantal articulation than in a stop, where it remains bound to the release at the end of the closure. This timing difference would also explain why no contrast was possible in Proto-Athabaskan between VV and Vʔ nuclei in stems ending in glottalic sonorants or fricatives: the glottal constriction at the beginning of a sonorant or fricative consonant would coincide with that at the end of a glottalic full vowel and render VV'R and VV'X rimes indistinguishable from VʔR and VʔX rimes.

The coordination of the glottalic articulation with the release of a stop was not enough to prevent it from constricting all or most of a preceding reduced vowel, so marked tone developed in such stems, but in stems with full vowels, only a portion of the vowel would be constricted by co-articulation with the final consonant – too little for marked tone to develop on them.

3.2.2 *Opposite tones from the same source*

The marked tone differs in level between languages, for example, it is high in Chipewyan, but low in Gwich'in. Kingston (2005) argues that both high and low tones could have developed from stem-final glottalic consonants because their laryngeal articulation could have been pronounced so as to raise or lower F0. If the glottal constriction was achieved by contracting the thyroarytenoid muscle alone, then the voice quality of adjacent vowels would be creaky and F0 would be characteristically low; however, if the cricothyroid muscle was also contracted at the same time, the adjacent vowel's voice quality would instead be tense and its F0 would be characteristically high.

According to Leer (1999), the languages with the high-marked tone are found on the Canadian Cordillera's east side and those with the low-marked tone on the Cordillera's west side. Yet within each of these two geographical ranges are found closely related languages with the opposite tone, e.g. low-marked Dogrib is spoken among otherwise high-marked and closely related languages on the east side, while on the low-marked west side Northern Tutchone is high-marked next to its close relative, low-marked Southern Tutchone; and similarly Tanacross is high-marked next to its close relatives, low-marked Upper and Lower Tanana.

The glottalic *vs.* non-glottalic contrast must have been lost very early in stem-final stops in the tonal languages, as none of them retains any vestige of it in the consonants themselves. How then could these recent reversals come about? The answer probably lies in the retention of the contrast in stem-final sonorants as well as between glottalic and non-glottalic full vowels in these languages down to the present day. If the ancestor of the tonal languages originally split into a high-marked and a low-marked dialect when some of its speakers chose to pronounce all the glottalic consonants with cricothyroid as well as thyroarytenoid contraction, while others chose to do so with thyroarytenoid contraction alone, then there is no reason why speakers should have lost this freedom to choose how to pronounce glottalic consonants in the subsequent history of the family. Speakers of Dogrib, Northern Tutchone, and Tanacross could have exercised this choice quite recently in words ending in glottalic sonorants or whose nucleus was glottalic, where tone remained redundant on the laryngeal articulation of the stem-final consonant or the nucleus. The contrastive tones in what were once stop-final stems must have switched value at the same time as those in the stems where the tones were predictable. That is, all high and low tones were treated alike regardless of whether they were contrastive or redundant.

Athabaskan is by no means unique in developing the opposite tone values from ostensibly the same source. Glottal stop produced a rising tone in Vietnamese (Table 97.3) and Chinese, but a falling tone in Lhasa Tibetan (Mazaudon 1977). An [h] produced a falling tone in Vietnamese (Table 97.3), but a high tone in Utsat.

A high tone also emerged after voiceless aspirated stops, [s] and [h] in Korean, and on Punjabi vowels followed by [h] or breathy voiced stops in Middle Indic (Ohala 1973).

3.3 *Summary*

In this section, I have described how tone emerged from the laryngeal articulations of following consonants in the earliest ancestors of present-day Vietnamese, the Chinese languages and the tonal Athabaskan languages – [constricted glottis] in all three families and [spread glottis] in Vietnamese and Chinese. In more recent ancestors of present-day Vietnamese and Chinese, the [voice] contrast in preceding consonants split the tones that arose from this source. Tonogenesis in Athabaskan shows that a glottal constriction can be pronounced so as to either raise or lower F0, depending on whether the cricothyroid as well as the thyroarytenoid is contracted. Other examples show that F0 can be both raised and lowered while the glottis is constricted or spread. The next section shows that such cross-linguistic variability is pervasive. This variability is another, strong clue that the so-called "perturbations" of F0 caused by these various laryngeal articulations may instead be controlled.

4 Tone splitting in East and Southeast Asia

Examples presented in this section show that languages can differ in whether higher or lower tonal reflexes arise after [voice], [constricted glottis], or [spread glottis] consonants. This cross-linguistic variability is attributed to the freedom to raise or lower F0 next to sonorants of any phonation type.

4.1 *Two- and three-way splits and cross-linguistic variability*

Tone splitting under the influence of onset consonants is characteristic of the Sino-Tibetan, Hmong, Tai, and Kam-Sui families of languages in East and Southeast Asia. A two-way split induced by an earlier [voice] contrast in initial consonants has already been illustrated with Vietnamese; many other examples of two-way splits induced by [voice] can be found in this linguistic area (Haudricourt 1972; Matisoff 1972, 1973; Mazaudon 1977). Three-way splits (Table 97.4), in which [spread glottis] and [constricted glottis] induce distinct tones, are also not uncommon.

In Yung-chiang Kam (Table 97.4a), higher tones have evolved after [constricted glottis] than [spread glottis] consonants, while in Nakhorn Sithammarat Thai (Table 97.4b) it is the reflexes after [spread glottis] consonants that are higher. These outcomes are referred to henceforth as "Constricted High" and "Spread High" splits.

In all examples so far, reflIexes after [voice] consonants have been lower than those after other phonation types, but some languages exhibit higher reflIexes instead, when tones split both two ways (Table 97.5a) and three ways (Table 97.5b). These outcomes are "Voiced High" splits, as compared to the "Voiced Low" splits in Tables 97.3 and 97.4.

Table 97.4 Three-way tone splitting: "x > y" indicates that "x" changed into "y" in the course of the split. Labial symbols are used to represent the various phonation types. Commas separate distinct reconstructed sounds or their reflexes. Chao's 1–5 low–high pitch scale is used to represent tones here

	b m v ≠ p^h ^hm f (p) ≠ ^ʔb ^ʔm (p) voice ≠ spread ≠ constricted		A	B	C
a.	*Constricted High: Yung-chiang Kam*				
	High	p, ^ʔb > m, ^ʔm > m	55	53	323
	Mid	p^h, ^hm > m	35	453	13
	Low	b > p, m	212	33	31
b.	*Spread High: Nakhorn Sithammarat Thai*				
	High	p^h, ^hm > m	53	53	55
	Mid	^ʔb, p	24	24	33
	Low	b > p^h, m	31	35	11

Table 97.5 Voiced High two-way and three-way splits in Shan and Szu ta chai. Slashes indicate multiple reflexes of the same reconstructed sound

	b m v ≠ p^h ^hm f (p) ≠ ^ʔb ^ʔm (p) voice ≠ spread ≠ constricted		A	B	C	D
a.	*Two-way: Shan*					
	High	b > p, m	55	22	44	
	Low	p, ^hm > m, ^ʔb > m/w	334	11	22	
b.	*Three-way: Szu ta chai*					
	High	b > p, m	35	33	44	33
	Mid	(m)p, ^ʔm > m	33	44	13	31
	Low	p^h, mp^h > $mp^ɦ$, s > $z^ɦ$, ^hm > ^ɦm	13	34	11	11

In the face of such dramatic cross-linguistic variation, it's perhaps easy to ignore the three outcomes that are consistent across all these examples (and the many others that could be presented): (i) the reflexes that emerged after original [voice] consonants were distinct from those that emerged after consonants of other specified phonation types;[6] (ii) the reflexes that emerged after the one series of obstruents that is unspecified for phonation type, the voiceless unaspirated stops, were distinct from those that emerged after the [voice] obstruents; (iii) sonorants and obstruents of a given phonation type produce the same reflexes.

The cross-linguistic variation is limited to the differences in level between the tonal reflexes observed after consonants of each phonation type.

[6] None of the examples presented here shows comprehensively how fricatives behave in tone splits, although a hint is provided by Szu ta chai, where [s] patterns with the voiceless aspirated stops. For the purposes of tonogenesis, fricatives are analyzed as contrasting for [spread glottis] rather than [voice], with voicing differences being redundant, like sonorants. Both [f] and [^hm] are [spread glottis] and redundantly voiceless, while [v] and [m] are unspecified for [spread glottis] and redundantly voiced.

4.2 *An explanation*

The traditional explanation for the Voiced High developments has been that these languages were originally Voiced Low, in conformity with phonetic expectations, but after the initial [voice] contrast neutralized and the tone split became contrastive, the tones' values were no longer constrained phonetically, and the tones that had evolved after the original [voice] consonants exchanged values with those that evolved after consonants with other phonation types. A similar story might be devised for the Spread High *vs.* Constricted High difference: either [spread glottis] or [constricted glottis] consonants first induced higher tone reflexes in the three-way splits, and the higher reflexes later exchanged values with the lower ones.

Kingston and Solnit (1989) give three arguments against such post-merger exchange processes. First, exchange processes otherwise do not occur but would have to apply frequently in this linguistic area to this linguistic property (see also Alderete 2001 and Moreton 2004 for formal objections). Second, since it is the tone features that have developed on the vowels that apparently invert their values, inversion should apply across all proto-tones. But, once inversion occurred, the prior, uninverted stage would no longer be detectable, and there would be no way to know it had ever happened. If the mechanism that inverted tones could apply to individual proto-tones, inversion would be observable in that some proto-tones would exhibit post-inversion Voiced High reflexes, while others would retain the prior Voiced Low reflexes. However, such mixed languages do not occur (see Brown 1975 for further elaboration of this argument). Third, the difference between Spread High and Constricted High splits is observable prior to the merger of the laryngeal contrasts in the consonants, in a Spread High language in Table 97.4b, and for the Constricted High Karenic language, Pwo, in Table 97.6. If languages can differ while the tone reflexes still remain predictable from the consonants' laryngeal articulations, then there is no need to assume that one development is original and the other a later inversion. In the Voiced Low language, Shih men k'an (Hmong), and many others, consonants contrasting for [voice] remain unmerged, but I have not found any cases of a Voiced High split in which this contrast has not merged (CHAPTER 80: MERGERS AND NEUTRALIZATION).

Solnit and Kingston (1988) and Kingston and Solnit (1989) argue that the freedom to pronounce sonorants of any phonation type with a higher or lower F0 (but cf. Maddieson 1984) accounts for both Voiced High *vs.* Voiced Low and Spread High *vs.* Constricted High developments. The development of opposite tones from the same phonological sources in different Athabaskan languages was explained in §3.2.2 by appealing to the same freedom to pronounce sonorants differently

Table 97.6 Three-way Constricted High tone split in Pwo

	A	B
ʔb, p	33	41
pʰ, ʰm > m	11	41
b > pʰ, m	33	24

across languages. As in Athabaskan, once the sonorants' pronunciation and its effect on F0 had been determined in the East and Southeast Asian languages, the corresponding obstruents would be pronounced and affect F0 in the same way, even if that pronunciation would conflict with the obstruent's phonetic proclivity to perturb F0 in the opposite direction. A phonetic conflict may in fact only arise with [voice] obstruents, which lower F0, as both [spread glottis] and [constricted glottis] obstruents can be pronounced so as to raise or lower F0. Solnit and Kingston (1988) and Kingston and Solnit (1989) formalize this hypothesis by specifying first sonorants and then the corresponding obstruents for tones, using Halle and Stevens's (1971) features [stiff] and [slack], as well as for phonation type. The tone features are not substituted for [voice] in this analysis. Instead, a consonant's value for [stiff] and/or [slack] is predictable from/ redundant on its value for [voice], [spread glottis], or [constricted glottis] within a language, but not cross-linguistically.

This proposal does not account for entire language families or substantial sub-groups being either Constricted High or Spread High, while both Voiced Low and Voiced High developments are found within the same families, subgroups, and even between closely related languages. As only voiced obstruents are expected on phonetic grounds to have a consistent (lowering) effect on F0, we would expect them to constrain these changes, yet they are the most labile in tone splits.

5 Tones from preceding *vs.* following consonants: Splits *vs.* tonogenesis

Tones rarely arise *de novo* from phonation contrasts in preceding consonants, as they have done in Western Kammu dialects, Eastern Cham, Utsat, Yabem, and Korean; they are more likely to split existing tones, as they have frequently done in the Vietic, Sino-Tibetan, Hmong, Tai, and Kam-Sui families.

There is actually a paradox here. Hombert (1977) showed that the F0 perturbations induced by an initial [voice] contrast were much smaller and shorter lasting in Yoruba, where high, mid, and low tones contrast, than in English, where no contrasts between morphemes are carried by tone. The perturbations may be constrained in size and extent in Yoruba to prevent confusion of one tone with another. If the perturbations induced by initial consonants' laryngeal articulations are generally constrained in tone languages, then how could they ever bring about the splits just described? The answer must be that speakers of these languages have been sensitized to F0 differences between previously existing tones, and they would be more likely to attend to, rely on, and eventually transfer the contrast to the systematic F0 differences after preceding consonants, too, even if those differences were at first small and brief.

How, then, could tone have developed from an initial [voice] contrast in the Western Kammu dialects, Eastern Cham, or Utsat, where there was no tone before? Extensive contact with other languages whose tones have split under the influence of an initial [voice] contrast is the most likely explanation.[7] Contact

[7] Brunelle (2005) presents extensive sociolinguistic evidence that speakers of Eastern Cham did not have sufficient contact with Vietnamese, the most likely language to have provided a model for its tonogenesis.

with tone-splitting languages, however, cannot explain tonogenesis from preceding consonants in Yabem or Korean. Yabem is apparently not in contact with any other tonal languages, aside from its close relative Bukuwa. Close similarities between their tone systems indicate that tone must have developed in their immediate common ancestor, rather than Bukuwa being the source of Yabem's tones. Korean has, of course, been in prolonged contact with Chinese, but the phonologization of the F0 differences after [spread glottis] and [constricted glottis] *vs.* unspecified consonants is too recent to be explained by that contact, and too different from the tone split occasioned by the much earlier merger of the [voice] contrast in Chinese – many Chinese languages also maintain the contrast between voiceless unaspirated and aspirated stops, which is about to be transferred to a tone contrast in Korean. These languages and perhaps Eastern Cham (cf. Brunelle 2005) are at least incipient exceptions to the generalization that contrastive tone does not develop in non-tonal languages from laryngeal articulations in preceding consonants.

6 Tones from uncommon sources

While consonants have often induced or split tones on neighboring vowels, the vowels' own articulations have split tones in just two languages. The rarity of this development is surprising because F0 is higher in vowels in which the tongue is higher than in vowels in which the tongue is lower. This variation in the vowels' "intrinsic" F0 is so pervasive that it is thought to be a universal, mechanical consequence of raising the tongue, which pulls the larynx up via the hyoid bone (Ohala and Eukel 1987; Whalen and Levitt 1995; Whalen *et al.* 1999); but see Ladd and Silverman (1984), Steele (1986), and Kingston (1991, 1992) for contrary evidence. Despite these intrinsic F0 differences' ubiquity and apparently automatic character, it appears that vowel height has only split tones in the Angkuic language, U, a not-too-distant relative of the Kammu languages discussed in §2.3.1 above (Svantesson 1989), and in Lugbara, a Moru-Madi language of the Central Sudanic branch of Nilo-Saharan (Andersen 1986) – other putative cases have been argued not to be products of vowel height differences (Hombert 1978; Maddieson 1978; Schuh 1978).

6.1 *Tone from vowel height: U*

In U, the high level tone in the closely related language Hu has split: it remains high in syllables with high vowels, but became low in syllables with non-high vowels (Table 97.7a *vs.* b; forms from U's and Hu's close non-tonal relative, Lamet, are given when there is no Hu cognate showing the original vowel height).

6.2 *Tone from advanced tongue root: Lugbara*

Western Lugbara contrasts four tones: low [fì] 'it exploded', mid [fī] 'he entered', high [fí] 'intestines', and extra-high [fi̋] 'they entered'. Within a word, the vowels /ɛ ɔ/ are raised to [e o] when followed by /i/ or /u/, and /i u/ do not co-occur with /ɪ ʊ/. Andersen (1986) proposes that the language therefore distinguishes [+ATR] vowels, [i e o u], from [–ATR] vowels, [ɪ ɛ a ɔ ʊ].

Table 97.7 High tone in U where the vowel is high in Hu and/or Lamet (a) *vs.* low tone when the vowel is non-high in the related languages (b)

	U	*Hu*	*Lamet*	
a.	nchí	nsíʔ	si̯ʔ	'louse'
	si	(pa)síʔ	(pl)si̱ʔ	'rope'
	nthɨ́ɪ		ntu̯ʔ	'hole'
	thú		tu̯uʔ	'vegetable'
b.	sì	θéʔ	khe̯ʔ	'tree'
	ŋkhù	ŋkhóʔ		'rice'
	khù	sŋkhóʔ		'yesterday'

In the second syllable of disyllabic verb stems, high and extra-high tones are in complementary distribution: high tone occurs when the vowel is [−ATR], [ávɪ́] 'play' *vs.* extra-high tone when it is [+ATR] [átsi̋] 'walk'. In nouns, high tone only occurs on [−ATR] vowels, and the extra-high tone can occur on both [−ATR] and [+ATR] vowels, e.g. high [ȁrɪ́] 'blood' but *[ȁrí] and [ɔ̏tsɛ̋] 'dog' *vs.* [ɔ̏ɗɨ̋ɪ] 'thigh'. Andersen proposes that the high tone split in Western Lugbara and became high when the vowel bearing it was [−ATR] and extra-high when that vowel was [+ATR].

Higher F0 has been associated with [+ATR] in other languages (Denning 1989), so ATR differences may affect vowels' intrinsic F0 much like vowel height does, probably because [+ATR] vowels are often also somewhat higher than [−ATR] vowels.

6.3 *Why is tonogenesis from vowel height so rare?*

Tone splits from vowel height or ATR contrasts are decidedly rare. Why should a vowel's own intrinsic F0 be so unable to split its tone, while F0 perturbations brought about by neighboring consonants' laryngeal articulations can so readily do so? The most plausible answer to this question is that intrinsic F0 differences between vowels differing in height depend on the vowels being prominent (Reinholt Peterson 1978; Ladd and Silverman 1984; Steele 1986; Kingston 2007) or bearing a high(er) rather than low(er) tone (Hombert 1977; Zee 1980; Connell 2002), while the effects of consonants' laryngeal articulations on the F0 of adjacent vowels do not appear to depend on the vowels' prominence or tones. If the intrinsic F0 differences between vowels were less consistent across prosodic contexts than the F0 differences between consonants, learners would be less likely to reinterpret them as differences in tone.[8]

[8] The bias favoring the learning of modular constraints proposed by Moreton *et al.* (2008) and Moreton (2008a, 2008b) does not help here, because the constraints referring to tone and vowel height are no more a-modular than those referring to tone and consonants' laryngeal articulations.

7 From stress and intonation to tone: Tonogenesis in Germanic

Before concluding this chapter, I sketch the probable paths through which tone evolved in three North Germanic languages, Swedish, Norwegian, and Danish (Gårding 1977; Lahiri *et al.* 1999; Riad 1998, 2000, 2003), and in Central and Low Franconian (Hermans 1999; Gussenhoven 2000b; Boersma 2006; Kehrein 2007). Unlike all the other examples discussed in this chapter, the sources of tone in these languages are not the laryngeal articulations of segments, although the tones' distribution is partly regulated by or correlated with various segmental properties. The tones in these languages instead come from the F0 correlate of stress in the Scandinavian languages and from intonation in Central and Low Franconian.

7.1 North Germanic word accents

In the conservative variety of Swedish spoken in and around Stockholm, i.e. Central Standard Swedish, words contrast in whether they bear Accent 1 or Accent 2. Other dialects of Swedish, as well as Norwegian dialects, differ in whether the accents still contrast and in the timing of the F0 events that realize the accents relative to the segment strings of the words bearing them (Bruce and Gårding 1978; Bruce 1999; Gussenhoven and Bruce 1999). Despite these differences between dialects and languages, their various synchronic patterns all descend from the same reinterpretation of the F0 correlates of stress as word accents in late Proto-Nordic, between 800 and 1200 CE (Riad 1998, 2003).

On the surface, both word accents can be analyzed as a sequence of a H and L tone (Gussenhoven and Bruce 1999, among many others; see also CHAPTER 45: THE REPRESENTATION OF TONE; CHAPTER 32: THE REPRESENTATION OF INTONATION). In words bearing Accent 1, the L tone is aligned with the stressed syllable, while in words bearing Accent 2, it is the H tone that is aligned with that syllable. Thus the two word accents apparently do not contrast in the tones of which they are composed but in the alignment of those tones with respect to that word's primary stressed syllable (Bruce 1977). Phonetically, F0 is low from the beginning of the stressed vowel in words bearing Accent 1, while it starts high and falls to a low value across that syllable in words bearing Accent 2. For words bearing Accent 1, the H only appears if there is a preceding unstressed syllable. When they are the last word in a focused constituent, both Accent 1 and Accent 2 words bear a focal H tone that appears immediately after the word accent tones, and initial and final L% boundary tones appear at the beginning and end of each intonational phrase. All these tones are necessarily realized roughly a syllable later in phrases containing Accent 2 words than in those containing Accent 1 words.

Riad's (1998, 2003) phonological analysis differs from this superficial account. Rather than a contrast in the tune-to-text alignment of a HL contour that determines when the other tones of the intonational contour are realized, Accent 2 words contrast privatively in having a lexical H* aligned with their stressed syllable which is absent in Accent 1 words. Focus is still marked by an H tone, which follows the lexical H*, if any. The L that invariably precedes this H is not lexical but instead either part of focus marking, a default L, or one inserted to separate the lexical and focal Hs in Accent 2 words and satisfy the OCP. The melody of an Accent 1 word is therefore $[LH]_{Focus}$-L%, and that of an Accent 2 word H*-$[LH]_{Focus}$-L%.

The diachronic development of this phonological contrast began with a number of reducing sound changes in Proto-Nordic, particularly syncope of vowels in light (CV) syllables (sixth–ninth centuries CE), e.g. heavy–light–heavy /'doomi,jan/ > heavy–heavy */'døø,man/ 'to judge' and heavy–light /'gastiz/ > /'gæstr/ 'guest'. Stress was quantity-sensitive at this stage, and any heavy syllable would have been stressed. Syncope of medial vowels created stress clashes whenever the syllables that became adjacent as a result were both heavy. Stresses also clashed in many other words where heavy syllables abutted even prior to syncope, e.g. */'wur,ðoo/ 'words (NOM/ACC PL)'. Riad proposes that stress clash was resolved by destressing the second syllable through shortening and other changes that left the affected syllable light: */'døø,man/ > /'døø.ma/ and */'wur,ðoo/ > /'wor.ðu/.

Riad hypothesizes that high F0 was one of the phonetic correlates of stress at this time, and that this F0 peak survived destressing. Words that had undergone destressing would then have had two F0 peaks, one on the remaining primary stressed syllable and the other on the newly destressed syllable. After destressing, an F0 peak would no longer be a correlate of stress because only the first of these peaks would still be associated with a stressed syllable. Once destressing eliminated the prosodic equivalence of these two peaks, the first, the one still associated with the primary stressed syllable, was re-analyzed as a lexical H* tone. This is the path of development to present-day Accent 2 words. Words in which the primary stressed syllable was final would not have any stress clash to resolve, would not have undergone destressing, and would not ever have borne two prosodically non-equivalent F0 peaks. No re-analysis of the F0 peak on a final stressed syllable as a lexical H* would therefore take place, so what would become Accent 1 words would not acquire a specification for a lexical H*. The F0 peak on this syllable, as well as the one on the destressed syllable in Accent 2 words-to-be, would then be reinterpreted as intonational in origin, i.e. as the present-day focal H. Subsequent sound changes added syllables to words that were once monosyllabic and produced words with non-final stress and Accent 1 and thus possible contrasts with Accent 2, notably the appearance of svarabhakti vowels before syllabic sonorants, e.g. /akr/ > /aker/ 'field', /fogl/ > /fogel/ 'bird', and /sookn/ > /sokken/ 'parish', and enclisis of the definite article, e.g. **/and hinn/ 'duck the', */and-in/ > /anden/, cf. /anden/ 'the ghost', with Accent 2.

Lack of space prevents me from taking up the debates as to whether the present-day contrast between word accents is privative or equipollent, whether Accent 2 (Riad 1998, 2003) or Accent 1 (Lahiri *et al.* 2005) is synchronically marked if the contrast is privative, and whether the accents are lexically or morphologically determined in the present-day languages (see for discussion Riad 1998, 2003; Morén 2006).

7.2 Danish *stød*

Danish *stød* is phonetically a glottal stop or creaky voice that causes F0 to drop sharply at the end of the sonorant interval in a syllable bearing it. It is represented here with an apostrophe following the end of this interval. A syllable can bear *stød* if it is stressed and bimoraic (heavy), and its second mora is a vowel or a sonorant followed by another consonant. If the second mora is a sonorant consonant and word-final, then *stød* cannot occur. These conditions are referred to collectively as the "*stød* basis." Syllables bearing *stød* are typically either the sole syllable in monosyllabic words or final stressed syllables in longer words. Following Itô and Mester (1997), Riad (2000) analyses *stød* as the allophone of a general HL word contour that occurs when these conditions are met, i.e. when both tones are compressed into a single word-final stressed syllable. When the stressed syllable is non-final, the two tones can be distributed across two syllables and no *stød* occurs.

Given the constraints on its distribution, it should not be surprising that Danish *stød* corresponds to the similarly restricted Accent 1 in Swedish and Norwegian. The correspondence is not perfect, however, because the distribution of *stød* is constrained further by the *stød* basis, while Accent 1's distribution is only constrained by the original position of stress in the word. If *stød* is simply the allophone of a general HL contour that is realized under the segmental and prosodic conditions specified by the *stød* basis, then, as Riad (2000) emphasizes, these conditions only determine how that contour can be realized phonetically and do not restrict whether the underlying HL can occur, unlike the conditions on the distribution of Accent 1 in Swedish and Norwegian.

Riad (2000) proposes that *stød* developed from word accents in a North Germanic dialect quite like the Central Swedish dialect of Western Mälardalen, centered around the city of Eskilstuna, southwest of Stockholm (cf. Liberman 1982). The tones in this dialect are shifted earlier compared to Central Stockholm Swedish, such that the L% boundary tone is aligned with the stressed syllable, which crowds the tones of the focal contour earlier, too, displacing the focal L from the stressed syllable and aligning the focal H with it. A consequence of this crowding of the focal H and L% boundary tones onto the stressed syllable is that, in the appropriate pragmatic conditions, the L% boundary tone is realized as a steep drop in F0 that may end in creaky voice.

The Eskilstuna "curl" phonetically resembles Danish *stød*. However, *stød* occurs in a narrower range of contexts than curl (the *stød* basis), it is categorical in the contexts where it occurs rather than pragmatically conditioned (CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY), and word accents no longer contrast in Danish, unlike in Eskilstuna and in Western Mälardalen generally. The alignment of the L% boundary tone with the stressed syllable in the Eskilstuna-like ancestor of Danish motivated the segmental restrictions on the occurrence of *stød* that constitute its basis: the L tone needed sonorant material on which to realize itself. The shift of the focal H tone to the stressed syllable in this Eskilstuna-like ancestor would have aligned H tones with stressed syllables generally, which would lead to their re-analysis as stress correlates and the loss of any phonological distinction between the lexical H* of Accent 2 and the focal H. In words with Accent 1, nothing changes as a result of this re-analysis, but in words with Accent 2, the re-analysis of the focal H would have led speakers to hypothesize secondary stresses on the syllables bearing the focal H tone that were actually unstressed. At that point, none of the tones is lexical any longer. These sound changes effectively reverse the sequence of prosodic changes that originally created the word accents from the F0 peaks aligned with stressed syllables in late Proto-Nordic.

This analysis appears to predict, incorrectly, that *stød* should correspond to Accent 2 in addition to Accent 1. Riad (2000) points to two properties that would have blocked such correspondences. First, the restrictions imposed by the *stød* basis would block many words with Accent 2 from developing *stød*. Second, large numbers of vowel-final Danish words that correspond to Swedish or Norwegian words with Accent 2 end in schwa rather than in a full vowel, and schwa cannot bear *stød*. These characteristics jointly limit the number of possible Accent 2 words that could have met the conditions to develop *stød*.

7.3 *Scandinavian summary*

In Riad's (1998, 2003) account, word accents developed in the ancestor of Swedish and Norwegian once clash-resolving destressing of the second of two stressed syllables prosodically differentiated the H tone that remained aligned with the primary stressed syllable from that which had been aligned with the destressed syllable. The H that remained aligned was re-analyzed as a lexical H* tone, producing Accent 2. Final primary stressed syllables would not have destressed, the H tone aligned them would not have been re-analyzed as a lexical tone, and as a result Accent 1 words are not specified for tone lexically. Danish *stød* developed from word accents as a result of a shift of the L% boundary tone and the focal H tone to the stressed syllable, which led to a re-analysis of H tones as correlates of stress and the reintroduction of covert secondary stresses that reversed the sequence of sound changes that originally produced the word accents.

7.4 *Tonogenesis in Central and Low Franconian*

In Central and Low Franconian dialects, heavy, stressed syllables bear one of two tone patterns, referred to variously as "*Stosston*" or "*stoottoon*" *vs.* "*Schleifton*" or "*sleeptoon*," "*Schärfung*" *vs.* "*Trägheitakzent*," "acute" *vs.* "circumflex," and "Accent 1" *vs.* "Accent 2" (Gussenhoven 2000b; Boersma 2006; Kehrein 2007).[9] Their distribution depends on the height of the vowel in the stressed syllable of a word and the voicing of the following consonant, but neither dependency is a by-product of the effects that vowel height or voicing may have on F0. This sketch of their historical development is a synopsis of Boersma's (2006) account, broken into three stages: Old Low Franconian to Early Middle Limburgian (900–1100 CE), Early to Late Middle Limburgian (1300–1400 CE), and Late to Present-day Middle Limburgian. The first two stages are illustrated in Table 97.8:

Table 97.8 First two stages in the development of Low and Central Franconian tonogenesis. At each stage, the first row shows the underlying quantitative structure and the second row the alignment of the declarative HL melody. The interrogative LH melody would be aligned in the same fashion

	Accent 1		*Accent 2*				
stages	'sleep'	'bed'	'grape'	'walk'	'hold'	'brook'	'chest'
1	slɑ$_{\mu}$ɑ$_{\mu}$pən	bɛ$_{\mu}$d$_{\mu}$də	dru$_{\mu}$wvə	lɔ$_{\mu}$wpən	ɦɑ$_{\mu}$ldən	bɛ$_{\mu}$kə	ke$_{\mu}$stə
	slɑ́ɑ́pə̀n	bɛ́d̀də̀	drúwvə̀	lɔ́wpə̀n	ɦáldə̀n	bɛ́kə̀	késtə̀
2 OSL	slɑ$_{\mu}$ɑ$_{\mu}$pən	bɛ$_{\mu}$d$_{\mu}$də	dru$_{\mu}$u$_{\mu}$və	lɔ$_{\mu}$u$_{\mu}$pan	ɦa$_{\mu}$ldən	bɛ$_{\mu}$ɛ$_{\mu}$kə	ke$_{\mu}$stə
	slɑ́ɑ́pə̀n	bɛ́d̀də̀	drúúvə̀	lɔ́úpə̀n	ɦáldə̀n	bɛ́ɛ́kə̀	késtə̀

[9] Only the so-called "rule A" dialects are discussed. Rule B dialects reverse the assignment of accents in many, although not all, forms, but the diachronic path to reversal remains obscure.

In the first stage, a quantity contrast developed between bimoraic long non-high vowels ('sleep') and monomoraic long high vowels ('grape'), diphthongs ('walk'), and short vowel–sonorant sequences ('hold') (see also CHAPTER 20: THE REPRESENTATION OF VOWEL LENGTH; CHAPTER 57: QUANTITY-SENSITIVITY). Open and closed stressed syllables containing short vowels were also monomoraic ('brook' and 'chest', respectively), except when the following consonant was a geminate, whose first half also projected a mora ('bed'). Both tones of the declarative HL or interrogative LH focal accents could map onto the stressed syllable when its nucleus was a long non-high vowel or a short vowel followed by a geminate consonant, but only the first tone could map onto the single mora in all other stressed syllables because tones were mapped left-to-right, at most one tone to a mora, and at most one mora to a tone.

Quantity differed between high and non-high long-vowel vowels because high vowels are inherently shorter than non-high ones, prior lowering of short high vowels to a higher mid quality eliminated the quantity contrast in high vowels in Early Middle Limburgian, and the long high vowels consisted of a vocalic segment followed by a homorganic glide. They thus resembled diphthongs and short vowel–sonorant sequences structurally in consisting of a vocalic segment followed by a consonantal one. At this stage, the realization of the focal pitch accents remained entirely predictable from the quantity differences between the two kinds of stressed syllables.

Open syllable lengthening ('brook') eliminated this predictability and transferred the original quantity contrast to a pitch accent contrast in the second stage. Original long non-high vowels bore Accent 1, while the newly lengthened non-high vowels bore Accent 2. The long high vowels, diphthongs, and short vowel–sonorant sequences were necessarily reinterpreted as bimoraic after open syllable lengthening, but they, too, retained Accent 2. This accentual contrast preserved the original mapping and phonetic timing of the focal pitch accents relative to syllables of the first stage, at the expense of violating the constraint mapping just a single mora to each tone. Henceforth, F0 changed during the stressed syllable itself in Accent 1 words, while in Accent 2 words it changed between the stressed and following unstressed syllable, a difference reminiscent of the superficial timing difference between Accent 1 and Accent 2 words in Swedish and Norwegian.

In the third stage, short vowels in closed syllables that previously alternated with lengthened vowels in open syllables in nominal, verbal, and adjectival paradigms lengthened, e.g. /dak/ > /daak/ 'roof (FEM SG)', /dax/ > /daax/ 'day (MASC SG)', and /lam/ > /laam/ 'lame (SG)', by analogy with their plurals /daakə/, /daaɣə/ and /laamə/ (see also Gussenhoven 2000b). The singulars would have borne Accent 2, either because they all had short vowels previously or on analogy with their plurals, which as products of open syllable lengthening bore Accent 2 themselves.

Final schwa apocope introduced Accent 1 on words with long high vowels, diphthongs, and lengthened vowels followed by voiced consonants through a shift of the L that had been aligned with the schwa to the first syllable, e.g. /drúúvə̀/ > /drúúv̀/ > /drúùv/. Three properties made this tone's preservation and shift possible: the consonant preceding the schwa belonged to the same mora as the schwa; it therefore bore the tone mapped onto that mora; and that tone could be heard during a voiced consonant. That is, the pronunciation before apocope was [drúúʋ̀ə]. This shift produced new alternations between /lɛ́ɛ̀v/ 'live (1 SG)/ (IMP SG)' and /dáàɣ/ 'days' with Accent 1, on the one hand, and /lɛ́ɛ́vən/ 'to live' and /dááx/ 'day' with Accent 2, on the other. In the Central Franconian dialects described as undergoing Rule A, but not in those undergoing Rule A2, the extension of Accent 1 generalized to words in which the schwa did not drop, e.g. in /lɛ́ɛ̀vən/ 'to live'. Accent 1 was also retained in words with geminates when the geminate was voiced, /bɛ̀d̀də/ 'bed' > /bɛ́dd̀/ and /zóǹǹə/ 'sun' > /zóǹǹ/ (cf. /nákkə/ 'neck' > /nákk/). Later degemination produced /bɛ̂d and /zôn/, with Accent 1, *vs.* /nák/, with Accent 2. Accent 2 replaced Accent 1 when the geminate was voiceless, because the second tone of the pitch accent could not be heard during a voiceless consonant.

The development of the accentual contrast in Central and Low Franconian certainly depended on both vowel height and consonant voicing, each at a different stage in its history, but not in any way that reflected the direct influences these segmental contrasts may have on F0. Instead, it was the inherently shorter duration of high compared to non-high vowels, the absence of a quantity contrast in the high vowels, and the structural parallel between long high vowels, falling sonority diphthongs, and short vowel–sonorant sequences that determined the initial difference in the mapping of the two tones of the focal pitch accents. And it was the audibility of F0 values during voiced intervals, including those of voiced obstruents, that determined the later appearance of Accent 1 where Accent 2 might otherwise have been expected. The inception of the accentual contrast at stage 2, however, did not depend on either property, but instead on conserving the previously predictable mapping of tones to syllables after open syllable lengthening had largely eliminated the quantity contrast that previously determined that mapping.

Gussenhoven (2000b) and Kehrein (2007) present a number of additional empirical arguments for not treating the distribution of the accents as instances of tonogenesis from the intrinsic F0 differences between vowels differing in height or on vowels preceding rather than following obstruents differing voicing. First, vowel qualities changed subsequently in directions opposite to those expected from the tones they bear. Second, voiced consonants only induced Accent 1 in place of Accent 2 on high vowels in words of more than one syllable, across the board in Rule A and Rule A2 dialects following final schwa apocope, and when schwas were retained in Rule A dialects, while original monosyllables with high vowels consistently exhibit Accent 2. The presence of a following vowel should have been of no consequence if Accent 1 developed from the effect of the voicing contrast on the preceding vowel's F0. This difference is expected, however, if the voiced consonant instead preserved and made audible the tone on the second syllable. Perhaps the most compelling argument, however, is that Accent 1 and Accent 2 are only realized with a HL *vs.* sustained H melody on the stressed syllable in declaratives. In interrogatives, they are instead realized with LH *vs.* sustained L melodies. If both H and L tones can be regulated by both non-high and high vowels and by both voiced and voiceless obstruents within the same language, then the effects of vowel height and obstruent voicing differences on F0 cannot have determined the accents' distributions.

This last argument raises the question, should this accentual contrast's development be treated as an instance of tonogenesis at all? In the present-day languages, the tones themselves arise from the intonation, and there is no reason to think that they have not done so throughout the period when the accentual contrast developed. The contrast may therefore have always been in how intonational tones were aligned with the utterance's segmental material. Kehrein (2007) argues for just such a synchronic analysis in which there are no lexical tones, and a word's (original) segmental properties determine whether it projects a moraic trochee and Accent 1 alignment or a syllabic trochee and Accent 2 alignment. He also shows that phonetic evidence for a lexical tone can be hard to find in the F0 contours of present-day Central and Low Franconian words (Gussenhoven 2000a, 2000b; cf. Gussenhoven and Peters 2004).

7.5 Differences between Central and Low Franconian and Scandinavian

Boersma's analysis sharpens the synchronic and diachronic differences between the Central and Low Franconian accents and the Scandinavian accents. Even though tones are realized earlier in Accent 1 than in Accent 2 words, in both the tones remain properties of the intonation in Central and Low Franconian, while they now arise from the lexicon and/or morphology in the Scandinavian languages. Word accents developed in the Scandinavian languages as a result of clash-resolving destressing and the resulting reinterpretation of the F0 peaks that had previously been predictable correlates of stress. In Central and Low Franconian, inherent differences in duration between high and non-high vowels and differences in the audibility of F0 between voiced and voiceless targets determined instead how focal pitch accents were aligned with segments. The result in the Scandinavian languages was the development of a contrast between the presence and absence of a lexical tone, that is, true tonogenesis; but in Central and Low Franconian, the contrast was instead between prosodic structures that forced intonational tones to align differently.

8 Summary and conclusions

This chapter has reviewed the ways in which languages have acquired contrastive tone, particularly tonogenesis and tone splitting in East and Southeast Asian language families, including Sino-Tibetan, Hmong, Tai, Kam-Sui, the few tonal Mon-Khmer and Chamic languages, Athabaskan, Yabem, and Korean. The laryngeal articulations of consonants raised or lowered F0 in neighboring vowels, those effects on F0 were exaggerated, and contrast transferred from the consonants to tone once the consonants were otherwise no longer pronounced differently. Just two languages, U and Lugbara, have apparently split their tones under the influence of intrinsic F0 differences between vowels contrasting for height and ATR, respectively.

The chapter closed by discussing how the word accents developed in Swedish and Norwegian from clash-resolving destressing and the resulting re-analysis of F0 peaks as lexical and focal tones rather than as correlates of stress, how *stød* subsequently developed from these word accents in Danish when F0 peaks realigned with stressed syllables and were once again analyzable as stress correlates, and how contrasting patterns of tune–text alignment developed from original quantity contrasts and subsequent changes in quantity in Central and

Low Franconian. The Franconian case can be described as a transfer of contrast from segments to tone, but one where segmental properties affected the timing and audibility of intonational F0 targets rather than F0 values. In phonologizing a formerly predictable correlate of stress, the Scandinavian word accents also transfer a contrast: Accent 2 emerged in words in which secondary stresses were lost, to resolve stress clash *vs.* Accent 1 on words in which no stress clash occurred because the final syllable bore primary stress. Subsequent sound changes affecting Accent 1 words introduced following unstressed syllables, and words with non-final primary stress now contrast for word accent.

In Athabaskan and East and Southeast Asian languages, consonants with particular laryngeal articulations were the source of higher tones in one language and lower tones in another. This variability was attributed to the presence of contrasts for laryngeal articulations in sonorants as well as in obstruents. No phonetic constraints on laryngeal articulations in sonorants prevent them from being pronounced so as to either raise or lower F0. Individual languages, groups of languages, or whole families opt to pronounce the laryngeal articulations in sonorants one way or the other, and the obstruents with the same laryngeal articulations follow suit. Athabaskan differs from the East and Southeast Asian languages in preserving the laryngeal contrasts in sonorants and transferring them from obstruents, while it is the other way around for the East and Southeast Asian languages. As a result, particular groups of languages and individual languages in the Athabaskan family have repeatedly changed their pronunciation of the sonorants, while in East and Southeast Asian language groups are far more uniform, with the exception of the Voiced Low *vs.* Voiced High difference. These facts suggest that consonants should be specified for tone, using [stiff] and [slack] in addition to [voice], [spread glottis], and/or [constricted glottis].

This chapter does not account for tonogenesis in the many tonal Central American, African, and New Guinean languages (except Yabem). Tones have been reconstructed for the proto-languages from which the present-day Central American and African languages descend, so it has not been possible to work out how these languages became tonal. Catalogues are beginning to emerge of the types of tone languages found in New Guinea, particularly in the Trans-New Guinea stock (Donohue 1997, 2005; Ross 2005), but there the data vary too much to generalize about their historical development.

ACKNOWLEDGMENTS

I have been helped enormously in writing this chapter by Paul Boersma, Marc Brunelle, James Kirby, Wolfgang Kehrein, Tomas Riad, the editors Keren Rice and Marc van Oostendorp, two anonymous reviewers, and especially David Solnit. If its virtues are largely theirs, its errors remain mine.

REFERENCES

Alderete, John. 2001. *Morphologically governed accent in Optimality Theory.* New York & London: Routledge.

Andersen, Torben. 1986. Tone splitting and vowel quality: Evidence from Lugbara. *Studies in African Linguistics* 17. 55–68.

Avery, Peter & William J. Idsardi. 2001. Laryngeal dimensions, completion and enhancement. In T. A. Hall (ed.) *Distinctive feature theory*, 41–70. Berlin & New York: Mouton de Gruyter.

Ballard, W. L. 1988. *The history and development of tonal systems and tone alternations in South China*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Baxter, William H. 1982. *A handbook of Old Chinese phonology*. Berlin & New York: Mouton de Gruyter.

Bell-Berti, Fredericka & Hajime Hirose. 1975. Palatal activity in voicing distinctions: A simultaneous fiberoptic and electromyographic study. *Journal of Phonetics* 3. 69–74.

Benedict, Paul K. 1972. *Sino-Tibetan: A conspectus*. Cambridge: Cambridge University Press.

Boersma, Paul. 2006. The history of the Franconian tone contrast. Unpublished ms., University of Amsterdam.

Bradshaw, Joel. 1979. Obstruent harmony and tonogenesis in Jabêm. *Lingua* 49. 189–205.

Brown, J. Marvin. 1975. The great tone split: Did it work in two different ways? In Jimmy G. Harris & James R. Chamberlain (eds.) *Studies in Tai linguistics in honor of William J. Gedney*, 33–48. Bangkok: Office of State Universities.

Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.

Bruce, Gösta. 1999. Word tone in Scandinavian languages. In van der Hulst (1999), 605–633.

Bruce, Gösta & Eva Gårding. 1978. A prosodic typology for Swedish dialects. In Eva Gårding, Gösta Bruce & Robert Bannert (eds.) *Nordic prosody*, 219–228. Lund: Gleerup.

Brunelle, Marc. 2005. Register in Eastern Cham: Phonological, phonetic, and sociolinguistic approaches. Ph.D. dissertation, Cornell University.

Cassimjee, Farida. 1998. *IsiXhosa tonology: An Optimal Domains Theory analysis*. Munich: Lincom Europa.

Cassimjee, Farida & Charles W. Kisseberth. 1998. Depressor consonants in Mijikenda and Nguni. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 18. 26–40.

Collier, René. 1975. Physiological correlates of intonation patterns. *Journal of the Acoustical Society of America* 58. 249–255.

Connell, Bruce. 2002. Tone languages and the universality of intrinsic F0: Evidence from Africa. *Journal of Phonetics* 30. 101–129.

Denning, Keith. 1989. The diachronic development of phonological voice quality, with special reference to Dinka and other Nilotic languages. Ph.D. dissertation, Stanford University.

Diffloth, Gérard. 1989. Proto-Austroasiatic creaky voice. *Mon-Khmer Studies* 15. 139–154.

Donnelly, Simon. 2009. Tone and depression in Phuthi. *Language Sciences* 31. 161–178.

Donohue, Mark. 1997. Tone systems in New Guinea. *Linguistic Typology* 1. 347–386.

Donohue, Mark. 2005. Tone and the Trans New Guinea languages. In Kaji (2005), 33–54.

Downing, Laura J. 2009. On pitch lowering not linked to voicing: Nguni and Shona group depressors (with an appendix by Thilo C. Schadeberg). *Language Sciences* 31. 179–198.

Downing, Laura J. & Bryan Gick. 2005. Voiceless tone depressors in Nambya and Botswana Kalang'a. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 27. 65–80.

Edmondson, Jerold A. & John H. Esling. 2006. The valves of the throat and their functioning in tone, vocal register and stress: Laryngographic case studies. *Phonology* 23. 157–191.

Erickson, Donna, Kiyoshi Honda, Hiroyuki Hirai, Mary E. Beckman & Niimi Seiji. 1994. Global pitch range and the production of low tones in English intonation. *Proceedings of the 3rd International Conference on Spoken Language Processing*, 651–654. Yokohama: Acoustical Society of Japan.

Ewan, William G. 1976. Laryngeal behavior in speech. Ph.D. dissertation, University of California, Berkeley.

Fromkin, Victoria A. (ed.) 1978. *Tone: A linguistic survey*. New York: Academic Press.

Gårding, Eva. 1977. *The Scandinavian word accents*. Lund: Gleerup.

Gregerson, Kenneth J. 1976. Tongue-root and register in Mon-Khmer. In Jenner *et al.* (1976), 323–369.

Gussenhoven, Carlos. 2000a. The boundary tones are coming: On the non-peripheral realization of boundary tones. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, 132–151. Cambridge: Cambridge University Press.

Gussenhoven, Carlos. 2000b. On the origin and development of the Central Franconian tone contrast. In Lahiri (2000), 215–260.

Gussenhoven, Carlos & Gösta Bruce. 1999. Word prosody and intonation. In van der Hulst (1999), 233–271.

Gussenhoven, Carlos & Jörg Peters. 2004. A tonal analysis of Cologne Schärfung. *Phonology* 21. 251–285.

Halle, Morris & Kenneth N. Stevens. 1971. A note on laryngeal features. *MIT Research Laboratory of Electronics Quarterly Progress Report* 101. 198–213.

Hargus, Sharon & Keren Rice (eds.) 2005. *Athabaskan prosody*. Amsterdam & Philadelphia: John Benjamins.

Haudricourt, André-George. 1954a. Comment reconstruire le chinois archaique. *Word* 10. 351–364.

Haudricourt, André-George. 1954b. De l'origine des tons en vietnamien. *Journal Asiatique* 242. 69–82.

Haudricourt, André-George. 1972. Two-way and three-way splitting of tonal systems in some Far Eastern languages. In Jimmy G. Harris & Richard B. Noss (eds.) *Tai phonetics and phonology*, 58–86. Bangkok: Central Institute of English Language.

Henderson, Eugénie J. A. 1967. Grammar and tone in South-East Asian languages. *Wissenschaftliche Zeitschrift der Karl-Marx-Universität-Leipzig (Gesellschafts-und Sprachwissenschaftliche Reihe)* 1–2. 171–178.

Hermans, Ben. 1999. A description of tonal accent in a Limburgian dialect. In van der Hulst (1999), 633–658.

Hirose, Hajime, C.-Y. Lee & Tatsujiro Ushijima. 1974. Laryngeal control in Korean stop production. *Journal of Phonetics* 2. 145–152.

Hirose, Hajime, Hea Suk Park, Hirohide Yoshioka, Masayuki Sawashima & Hiroyuki Umeda. 1981. An electromyographic study of laryngeal adjustments for the Korean stops. *Annual Bulletin, Research Insitute of Logopedics and Phoniatrics, University of Tokyo* 15. 31–43.

Hombert, Jean-Marie. 1977. Consonant types, vowel height, and tone in Yoruba. *Studies in African Linguistics* 8. 173–190.

Hombert, Jean-Marie. 1978. Consonant types, vowel quality, and tone. In Fromkin (1978), 77–111.

Hombert, Jean-Marie, John J. Ohala & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55. 37–58.

Honda, Kiyoshi, Hiroyuki Hirai, Shinobu Masaki & Yasuhiro Shimada. 1999. Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech* 42. 401–411.

Huffman, Franklin. 1976. The register problem in fifteen Mon-Khmer languages. In Jenner *et al.* (1976), 575–589.

Hulst, Harry van der (ed.) 1999. *Word prosodic systems in the languages of Europe*. Berlin & New York: Mouton de Gruyter.

Hyman, Larry M. (ed.) 1973. *Consonant types and tone*. Los Angeles: University of Southern California.

Hyman, Larry M. 2008. Enlarging the scope of phonologization. *University of California Berkeley Phonology Lab Annual Report*, 382–409.

Itô, Junko & Armin Mester. 1997. Stø?det i dansk. Handout, Scandinavian Summer School in Generative Phonology, Hvalfjar Darströnd.

Jenner, Philip, Laurence Thompson & Stanley Starosta (eds.) 1976. *Austroasiatic Studies.* Honolulu: University Press of Hawaii.

Jessen, Michael & Justus Roux. 2002. Voice quality differences associated with stops and clicks in Xhosa. *Journal of Phonetics* 30. 1–52.

Jun, Sun-Ah. 1996. Influence of microprosody on macroprosody: A case of phrase-initial strengthening. *University of California Los Angeles Working Papers in Phonetics* 92. 97–116.

Kagaya, Ryohei. 1974. A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics* 2. 161–180.

Kaji, Shigeki (ed.) 1999. *Cross-linguistic studies of tonal phenomena: Tonogenesis, typology, and related topics.* Tokyo: ILCAA.

Kaji, Shigeki (ed.) 2005. *Cross-linguistic studies of tonal phenomena: Historical development, tone-syntax interface, and descriptive studies.* Tokyo: ILCAA.

Kang, Kyoung-Ho & Susan G. Guion. 2008. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America* 124. 3909–3917.

Kehrein, Wolfgang. 2007. There's no tone in Cologne: Against tone segment interactions in Franconian. Unpublished ms., University of Amsterdam.

Kingston, John. 1985. The phonetics and phonology of the timing of oral and glottal events. Ph.D. dissertation, University of California, Berkeley.

Kingston, John. 1990. Articulatory binding. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 406–434. Cambridge: Cambridge University Press.

Kingston, John. 1991. Integrating articulations in the perception of vowel height. *Phonetica* 48. 149–179.

Kingston, John. 1992. The phonetics and phonology of perceptually motivated articulatory coordination. *Language and Speech* 35. 99–113.

Kingston, John. 2005. The phonetics of Athabaskan tonogenesis. In Hargus & Rice (2005), 137–184.

Kingston, John. 2007. Segmental influences on F0: Automatic or controlled? In Carlos Gussenhoven & Tomas Riad (eds.) *Tones and tunes*, vol. 2: *Experimental studies in word and sentence prosody*, 171–210. Berlin & New York: Mouton de Gruyter.

Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70. 419–454.

Kingston, John & Randy L. Diehl. 1995. Intermediate properties in the perception of distinctive feature values. In Bruce Connell & Amalia Arvaniti (eds.) *Papers in laboratory phonology IV: Phonology and phonetic evidence*, 7–27. Cambridge: Cambridge University Press.

Kingston, John & David Solnit. 1989. The inadequacy of underspecification. *Papers from the Annual Meeting of the North East Linguistic Society* 19. 264–278.

Kingston, John, Randy L. Diehl, Cecilia J. Kirk & Wendy A. Castleman. 2008. On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics* 36. 28–54.

Krauss, Michael. 2005. Athabaskan tone. In Hargus & Rice (2005), 51–136.

Ladd, D. Robert & Kim E. A. Silverman. 1984. Vowel intrinsic pitch in connected speech. *Phonetica* 41. 31–40.

Lahiri, Aditi (ed.) *Analogy, leveling, markedness: Principles of change in phonology and morphology.* Berlin & New York: Mouton de Gruyter.

Lahiri, Aditi, Tomas Riad & Haike Jacobs. 1999. Diachronic prosody. In van der Hulst (1999), 335–421.

Lahiri, Aditi, Allison Wetterlin & Elisabet Jönsson-Steiner. 2005. Lexical specification of tone in North Germanic. *Nordic Journal of Linguistics* 28. 61–96.

Leer, Jeff. 1979. *Proto-Athabaskan verb stem variation. I: Phonology.* Fairbanks: Alaska Native Language Center.

Leer, Jeff. 1999. Tonogenesis in Athabaskan. In Kaji (1999), 37–66.

Leer, Jeff. 2001. Shift of tonal markedness in Northern Tlingit and Southern Athabaskan. In Shigeki Kaji (ed.) *Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese accentology, and other topics*, 61–89. Tokyo: ILCAA.

Liberman, Anatoly. 1982. *Germanic accentology: The Scandinavian languages*. Minneapolis: University of Minnesota Press.

Lindau, Mona. 1979. The feature expanded. *Journal of Phonetics* 7. 163–176.

Löfqvist, Anders & Hirohide Yoshioka. 1980. Laryngeal activity in Swedish obstruent clusters. *Journal of the Acoustical Society of America* 68. 792–801.

Löfqvist, Anders, Thomas Baer, Nancy McGarr & Robin Seider Story. 1989. The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America* 85. 1314–1321.

Maddieson, Ian. 1978. Universals of tone. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 2: *Phonology*, 335–365. Stanford: Stanford University Press.

Maddieson, Ian. 1984. The effects on F0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis. *Journal of Phonetics* 12. 9–15.

Maddieson, Ian. 2003. The sounds of Bantu languages. In Derek Nurse & Gérard Phillipson (eds.) *The Bantu languages*, 15–41. London: Routledge.

Maddieson, Ian & Keng-Fong Pang. 1993. Tone in Utsat. *Oceanic Linguistics Special Publications* 24. 75–89.

Matisoff, James A. 1972. *The Loloish tonal split revisited*. Berkeley: University of California.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In Hyman (1973), 71–95.

Mazaudon, Martine. 1977. Tibeto-Burman tonogenetics. *Linguistics of the Tibeto-Burman Area* 3. 1–123.

Mazaudon, Martine. 1978. Consonantal mutation and tonal split in the Tamang subfamily of Tibeto-Burman. *Kailash* 6. 157–179.

Mazaudon, Martine. Forthcoming. Paths to tone in the Tamang branch of Tibeto-Burman (Nepal). In Gunther de Vogelaer & Guido Seiler (ed.) *Dialects as a testing ground for theories of language change*. Amsterdam & Philadelphia: John Benjamins.

Mazaudon, Martine & Alexis Michaud. 2008. Tonal contrasts and initial consonants: A case study of Tamang, a "missing link" in tonogenesis. *Phonetica* 65. 231–256.

Mei, Tsu-lin. 1970. Tones and prosody in Middle Chinese and the origin of the Rising tone. *Harvard Journal of Asian Studies* 30. 86–100.

Morén, Bruce, 2006. Here/now versus there/then: Synchrony and diachrony in Scandinavian pitch accent. Paper presented at the 28th Generative Linguistics in the Old World Conference, Geneva.

Moreton, Elliott. 2004. Non-computable functions in Optimality Theory. In John J. McCarthy (ed.) *Optimality Theory in phonology: A reader*, 141–164. Oxford: Blackwell.

Moreton, Elliott. 2008a. Learning bias as a factor in phonological typology. *Proceedings of the West Coast Conference on Formal Linguistics* 26. 393–401.

Moreton, Elliott. 2008b. Modeling modularity bias in phonological pattern learning. *Proceedings of the West Coast Conference on Formal Linguistics* 27. 1–16.

Moreton, Elliott, Gary Feng & Jennifer L. Smith. 2008. Syllabification, sonority, and perception: New evidence from a language game. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 41. 341–355.

Ohala, John J. 1973. The physiology of tone. In Hyman (1973), 1–14.

Ohala, John J. & Brian Eukel. 1987. Explaining the intrinsic pitch of vowels. In Robert Channon & Linda Shockey (eds.) *In honour of Ilse Lehiste*, 207–215. Dordrecht: Foris.

Ohala, John J. & Carol J. Riordan. 1979. Passive vocal tract enlargement during voiced stops. In Jared J. Wolf & Dennis H. Klatt (eds.) *Speech communication papers*, 89–92. New York: Acoustical Society of America.

Phu, Van Han, Jerold A. Edmondson & Kenneth J. Gregerson. 1992. East Cham as a tone language. *Mon-Khmer Studies* 20. 31–43.

Pulleyblank, Edwin G. 1962. The consonantal system of Old Chinese. *Asia Major, Series 2*, 9. 58–144, 206–265.

Reinholt Peterson, Niels. 1978. Intrinsic fundamental frequency of Danish vowels. *Journal of Phonetics* 6. 177–189.

Riad, Tomas. 1998. The origin of Scandinavian tone accents. *Diachronica* 15. 63–98.

Riad, Tomas. 2000. The origin of the Danish *stød*. In Lahiri (2000), 261–300.

Riad, Tomas. 2003. Diachrony of the Scandinavian accent typology. In Paula Fikkert & Haike Jacobs (eds.) *Development in prosodic systems*, 91–144. Berlin & New York: Mouton de Gruyter.

Ross, Malcolm D. 1993. Tonogenesis in the North Huon Gulf Chain. *Oceanic Linguistics Special Publications* 24. 133–153.

Ross, Malcolm D. 2005. Towards a reconstruction of the history of tone in the Trans New Guinea family. In Kaji (2005), 3–31.

Rycroft, David K. 1980. *The depression feature in Nguni languages and its interaction with tone*, vol. 8. Grahamstown: Department of African Languages, Rhodes University.

Rycroft, David K. 1983. Tone patterns in Zimbabwean Ndebele. *Bulletin of the School of Oriental and African Studies, University of London* 46. 77–135.

Sagart, Laurent. 1999. The origin of Chinese tones. In Kaji (1999), 91–103.

Schachter, Paul. 1976. An unnatural class of consonants in Siswati. *Studies in African Linguistics*. Supplement 6. 211–220.

Schuh, Russell. 1978. Tone rules. In Fromkin (1978), 221–256.

Silva, David J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23. 287–308.

Solnit, David & John Kingston. 1988. Voiced high splits: Putting the blame on sonorants. Paper presented at the 63rd Annual Meeting of the Linguistic Society of America, New Orleans.

Steele, Shirley A. 1986. Interaction of vowel F0 and prosody. *Phonetica* 43. 92–105.

Stevens, Kenneth N. & Sheila Blumstein. 1981. The search for invariant acoustic correlates of phonetic features. In Peter D. Eimas & Joanne L. Miller (eds.) *Perspectives on the study of speech*, 1–38. Mahwah, NJ: Lawrence Erlbaum.

Strazny, Philipp. 2003. Depression in Zulu: Tonal effects of segmental features. In Jeroen van de Weijer, Vincent J. van Heuven & Harry van der Hulst (eds.) *The phonological spectrum: Segmental structure*, vol. 1, 223–239. Amsterdam & Philadelphia: John Benjamins.

Suwilai, Premsrirat. 2003. Khmu dialects: A case of register complex and tonogenesis. In Shigeki Kaji (ed.) *Cross-linguistic studies of tonal phenomena: Historical development, phonetics of tone, and descriptive studies*, 13–28. Tokyo: ILCAA.

Svantesson, Jan-Olof. 1989. Tonogenetic mechanisms in northern Mon-Khmer. *Phonetica* 46. 60–79.

Thongkum, Theraphan L. 1987. Phonation types in Mon-Khmer languages. *UCLA Working Papers in Phonetics* 76. 29–48.

Thurgood, Graham. 1999. *From ancient Cham to modern dialects: Two thousand years of language contact and change*. Honolulu: University of Hawai'i Press.

Thurgood, Graham. 2002. Vietnamese and tonogensis. *Diachronica* 19. 333–363.

Traill, Anthony. 1990. Depression without depressors. *South African Journal of African Languages* 10. 166–172.

Traill, Anthony, James S. M. Khumalo & Paul Fridjhon. 1987. Depressing facts about Zulu. *African Studies* 46. 255–274.

Wang, Li. 1958. *Hanyu shi gao* [*A sketch of the history of the Han (Chinese) language*]. Beijing: Kezue Chubanshe.

Westbury, John R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America* 73. 1322–1336.

Whalen, Douglas & Andrea G. Levitt. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23. 349–366.
Whalen, Douglas, Bryan Gick, Masanobu Kumada & Kiyoshi Honda. 1999. Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0. *Journal of Phonetics* 27. 125–142.
Yoshioka, Hirohide, Anders Löfqvist & Hajime Hirose. 1981. Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Journal of the Acoustical Society of America* 70. 1615–1623.
Zee, Eric. 1980. Tone and vowel quality. *Journal of Phonetics* 8. 247–258.

98 Speech Perception and Phonology

Andrew Martin
Sharon Peperkamp

1 Introduction

The task of speech perception involves converting a continuous, information-rich waveform into a more abstract representation. This mapping process is heavily language-dependent – every language divides up acoustic space differently, and the mapping is distorted by context-dependent phonological rules. This is not an easy job, but it is made easier in two complementary respects. During the first year of life the human perceptual apparatus is gradually optimized to better perceive the distinctions that are crucial in the ambient language's phonological system while ignoring irrelevant variation, and phonological systems themselves are optimized from the perspective of human perception.

This two-way interaction, in which perception adapts to phonology and phonology to perception, has long been of interest to phonologists, but only in recent decades have the tools necessary to explore the connection between the two become available. In this chapter we provide an overview of the debates that have arisen around each issue, as well as the research that bears on each debate (see also Hume and Johnson 2001). §2 discusses how phonology influences speech perception, both in the native language (§2.1) and in non-native ones (§2.2), and how second language perception relates to loanword adaptations (§2.3). §3 addresses the question of how perception influences phonology, beginning with an overview of the relevant typological data (§3.1), and concluding with a comparison of theoretical approaches to the data (§3.2). Finally, in §4 we briefly consider the ramifications of the bidirectional nature of the phonology–perception interaction.

2 Phonological influences in speech perception

2.1 *Native language perception*

How we perceive speech in our native language has long been a topic of interest. As early as the 1930s, phonologists were interested in the role of abstract phonemes in perception (CHAPTER 11: THE PHONEME). Sapir (1933) argued that phonemes are psychologically real, in that native speakers are typically unaware of the allophones in their language. The evidence he provides comes from native speakers transcribing phonemes regardless of their phonetic realization. For instance, native speakers of Nootka fail to transcribe the phoneme /i/ as [e] when it occurs after /h/. Likewise, Swadesh (1934) observed that phonemes are the basic percepts by which native speakers perceive their language.

More recently, the psychological reality of phonemes has been investigated experimentally. For instance, Otake *et al.* (1996) used a phoneme detection task to examine the processing of the moraic nasal consonant by native speakers of Japanese.[1] In Japanese, the phonetic realization of this consonant depends upon the place of articulation of the following consonant: in the words *kanpa* 'campaign', *bando* 'band', *kenri* 'right', and *tanka* 'type of poem', it is realized as [m], [n̪], [n], and [ŋ], respectively. Otake *et al.* found that Japanese listeners respond equally fast and accurately to the four realizations when asked to detect the sound that is spelled N in *rōmaji* (the Japanese writing system that uses the Latin alphabet). Especially revealing is the lack of a difference for the bilabial realization, since in onset position the presence of [m] is contrastive (cf. *mori* 'forest' and *nori* 'type of seaweed'). Hence, Japanese listeners perceive all phonetic realizations of the moraic nasal consonant as instances of a single underlying unit.

Several other studies have shown perception differences for allophonic as opposed to phonemic contrasts (CHAPTER 2: CONTRAST) using discrimination paradigms. In these paradigms, participants are asked, for instance, to indicate whether a given stimulus is identical to another one or not, or to which of two different stimuli a third one is identical. It has thus been shown that listeners have difficulty distinguishing among allophones of the same phoneme, which are typically perceived as more similar than phones in different phonemic categories, even when acoustic distance is equated. For instance, Whalen *et al.* (1997) studied the perception of the allophonic contrast [p–pʰ] (with [p] as in *happy*), compared to that of the phonemic contrasts [b–p] and [b–pʰ]. They found that perception of the allophonic contrast was worse than that of the phonemic ones. Peperkamp *et al.* (2003) showed that, likewise, French listeners have difficulty perceiving the allophonic contrast [ʁ–χ] (both segments being realizations of the French phoneme /r/), compared to the phonemic contrast [m–n].

Of course, what is an allophonic contrast in one language can be a phonemic contrast in another language. Hence, another way of examining the effect of phonological status of a given contrast is to present it to one group of listeners for whom it is phonemic and to another group for whom it is allophonic. Such a cross-linguistic study was carried out by Kazanina *et al.* (2006). They used magneto-encephalographic (MEG) recordings to examine the perception of the contrast between [t] and [d] by native speakers of English and Korean. In English, this contrast is phonemic, whereas in Korean it is allophonic: [d] only occurs as an allophone of /t/ between two sonorant segments (CHAPTER 111: LARYNGEAL CONTRAST

[1] There is at least one earlier study on the processing of abstract phonemes: Jaeger (1980) used a conditioning as well as a concept-formation paradigm to examine whether English listeners perceive different allophones of /k/ to be part of a single category. Unfortunately, the results of these experiments were inconclusive (mainly because of lack of data).

IN KOREAN). Kazanina *et al.* revealed a mismatch negativity response (a neural marker for automatic change detection) in English but not in Korean listeners whenever a series of tokens of [ta] was interrupted by a token of [da]. In other words, Korean listeners did not detect the change from [ta] to [da], perceiving the different consonants as two instances of the same category. Similarly, using both a rating task and a discrimination task, Boomershine *et al.* (2008) found that English and Spanish speakers differ with respect to the perception of the contrasts [d–ð] and [d–ɾ]; the former is phonemic in English but allophonic in Spanish, while the reverse holds for the latter. Both groups of participants rated the contrast that is phonemic in their language as more different than the allophonic one and were better at discriminating it.

The fact that listeners have difficulty perceiving allophonic contrasts does not mean that allophones are ignored at all levels of speech processing. On the contrary, listeners appear to use their knowledge about the allophonic rules in their language for the purposes of word recognition. This was shown, for instance, by Lahiri and Marslen-Wilson (1991), who studied the processing of nasal vowels by English and Bengali listeners. In English, nasal vowels only occur as allophones of oral vowels before nasal consonants, whereas in Bengali the distinction between oral and nasal vowels is phonemic. Using a gating task, in which participants have to guess which word they hear as they listen to successively longer stretches of it, Lahiri and Marslen-Wilson found that adult native speakers of English, but not of Bengali, interpret nasality on vowels as an indication that a nasal consonant follows. That is, when listening to a word up to its nasal vowel, English speakers replied with words in which the vowel is followed by a nasal consonant. Likewise, Otake *et al.* (1996), in their study on the Japanese moraic nasal mentioned before, found that Japanese but not Dutch listeners use the phonetic identity of the moraic nasal consonant to anticipate the place of articulation of the next consonant. In particular, they had more difficulty in detecting a given consonant when it occurred after an inappropriate realization of the moraic nasal, i.e. a realization that does not share the following consonant's place of articulation; for instance, they found it hard to detect /k/ when it immediately followed a non-velar realization of the moraic nasal.

Finally, it is important to note that most of the evidence for the phoneme in perception reviewed above concerns unconscious perception processes. These processes should not be confused with (conscious) awareness. Adults who are illiterate or use a non-alphabetic writing system, as well as preliterate children, typically do not have phonemic awareness, that is, conscious knowledge about the phonemes of their language. They are thus unable to carry out metalinguistic tasks, such as deleting the first sound of a word (Liberman *et al.* 1974; Morais *et al.* 1979; Read *et al.* 1986), although they can perform such tasks with larger units such as syllables (Liberman *et al.* 1974; Morais *et al.* 1986). Importantly, the influence of the native phoneme inventory on speech perception is independent of literacy. Indeed, this influence arises very early in life, long before the question of literacy comes into play. For instance, it has been shown that, like English-speaking adults, 10- to 12-month-old English-learning infants fail to discriminate the non-phonemic contrast between [d] (as in *day*) and [d̥] (as in *stay*), whereas 6- to 8-month-olds do discriminate this contrast (Pegg and Werker 1997). Hence, during the first year of life, the capacity to discriminate non-phonemic contrasts already diminishes.

2.2 *Non-native language perception*

As shown above, our speech perception system is optimized for processing our native language. On the one hand, allophonic contrasts – which never induce differences in meaning – appear perceptually less salient than phonemic ones, even when acoustic distance is equated. On the other hand, the information provided by allophonic variation concerning neighboring segments is exploited for the purposes of word recognition. This optimization comes at a price, though, when we learn a second language later in life. Indeed, we experience difficulty perceiving non-native sounds and sound sequences, just as we find it hard to produce them.

That speech perception depends upon the phonological properties of the listener's native language was already observed in the 1930s (Polivanov 1931; Bloomfield 1933; Swadesh 1934; Trubetzkoy 1939). In the words of Trubetzkoy, the native phonology acts as a sieve during speech perception, in that non-native sounds and sound sequences are perceived as native ones, a phenomenon also called *perceptual assimilation*. Starting in the 1970s, experimental research has documented many cases of perceptual assimilation. Most of this research uses discrimination paradigms, the logic being that if a non-native sound is perceived as a native one, these two sounds are hard to distinguish; similarly, it is difficult to distinguish two non-native sounds if they are perceived as the same native one. An example of the former case is provided by German listeners, who have difficulty perceiving the contrast between the Polish phonemes /ʃ/ and /ɕ/, since they perceive non-native /ɕ/ as an instance of their native consonant /ʃ/ (Lipski and Mathiak 2007). An example of the latter is provided by what is probably the best-known instance of perceptual assimilation: Japanese listeners have difficulty perceiving the contrast between English /l/ and /ɹ/ (Goto 1971; Miyawaka *et al.* 1975). Indeed, Japanese has only one liquid consonant, /ɾ/, which acoustically falls in between [l] and [ɹ]; they thus perceive both English phonemes as instances of their native category /ɾ/ (see also CHAPTER 30: THE REPRESENTATION OF RHOTICS and CHAPTER 31: LATERAL CONSONANTS). According to the Perceptual Assimilation Model (PAM; Best *et al.* 1988; Best 1995), this English contrast is particularly difficult for Japanese listeners because [l] and [ɹ] are equally bad (or good) exemplars of Japanese /ɾ/. By contrast, the Polish contrast between /ʃ/ and /ɕ/ should cause less difficulty for German listeners, since it contrasts one good ([ʃ]) and one bad ([ɕ]) exemplar of the German category /ʃ/.

As already noted by Polivanov (1931), perceptual assimilation is not limited to segments, but also concerns suprasegmentals, syllable structure, and phonotactics. As for suprasegmentals, much research has been devoted to the perception of tones by speakers of non-tonal languages (CHAPTER 45: THE REPRESENTATION OF TONE). Kiriloff (1969) showed, for instance, that English listeners have difficulty perceiving certain lexical tones used in Mandarin. More recently, the perception of stress and length contrasts has started to be investigated. For instance, it was shown that French listeners have difficulty perceiving vowel length as well as word stress (Dupoux *et al.* 1997; Dupoux *et al.* 1999), neither of which is lexically contrastive in French. Concerning syllable structure (CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE), an example is again provided by Japanese, which has a very simple syllable structure, disallowing branching onsets and allowing only for codas consisting of a moraic nasal or the first half of a geminate. Japanese listeners find it hard to distinguish between consonant clusters and the same clusters broken up by the vowel /ɯ/

(Dupoux *et al.* 1999). Hence, they perceive non-native clusters with an illusory epenthetic vowel (CHAPTER 67: VOWEL EPENTHESIS). Finally, perceptual assimilation of non-native phonotactic structure has been studied, for instance, in French listeners; they have difficulty perceiving the contrast between the word-initial clusters /tl/ and /kl/ (Hallé *et al.* 1998; Hallé and Best 2007). In French, /tl/ is not a legal onset cluster (CHAPTER 55: ONSETS), and French listeners tend to perceive it as /kl/.

Perceptual assimilation arises during very early processing stages and does not require the listener's conscious attention. This is evidenced by the fact that compared to native contrasts, non-native contrasts give rise to a reduced mismatch negativity (Näätänen *et al.* 1997); this electrophysiological response component is a neural marker of automatic change detection that arises 150 to 250 msecs after the onset of the change when a deviant stimulus interrupts a sequence of repetitive stimuli.

One caveat is in order when interpreting the results of experiments purporting to demonstrate perceptual assimilation. Specifically, it is important to use experimental designs that tap a phonological processing level, where speech is coded in an abstract, language-specific format. Indeed, if listeners are given the opportunity to perform a task using low-level acoustic response strategies, no perceptual assimilation will be observed. For instance, Dupoux *et al.* (2001) used a very robust sequence recall task that yielded a complete separation between French and Spanish listeners as far as the perception of stress is concerned (that is, even the Spanish listener who had the most difficulty with perceiving stress performed better than the best French listener). However, they showed that in the absence of phonetic variability in the stimuli, French listeners performed equally well as Spanish listeners on the very same task. Differences between phonological and acoustic perception can also be shown by transforming the speech stimuli in such a way as to make them sound like non-speech. For instance, after removal of all acoustic information except the F3 component, Japanese listeners perceived the [l–ɹ] contrast as well as native English speakers (Miyawaki *et al.* 1975).

Research with infants has shown that many aspects of the native phonology, including segments, suprasegments, syllable structure, and phonotactics, are acquired early in life; one-year-old infants thus show many of the perceptual assimilation effects that can be observed in their parents. The question as to whether perceptual assimilation can be overcome at a later age has been studied in two types of populations, i.e. bilinguals, and monolinguals who have received specific perception training. In both cases, improvement appears to be possible, with native-like performance generally remaining out of reach, though. To start with training studies, Japanese monolinguals can improve their perception of the [l–ɹ] contrast over several weeks of intensive computerized training, but their performance remains significantly below that of English monolinguals (Lively *et al.* 1993). Similar results have been reported for monolinguals who were trained to improve their perception of non-native suprasegmental contrasts (e.g. Wang *et al.* 1999).

Concerning bilinguals, we can broadly distinguish *late* bilinguals, who have started to learn a second language during or after adolescence, and *early* bilinguals, who have learned their second language during childhood. There is evidence that late bilinguals exhibit the same difficulty as monolinguals, although possibly to a lesser extent (Takagi and Mann 1995; Flege *et al.* 1997; Dupoux *et al.* 2008). For instance, Takagi and Mann (1995) found that Japanese–English bilinguals who had lived for at least 12 years in the United States did not yield the same performance as native speakers of English on the [l–ɹ] contrast; their performance was better, though, than that of Japanese monolinguals. Dupoux *et al.* (2008) found that advanced French–Spanish bilinguals were actually indistinct from French monolinguals as far as the perception of stress is concerned. However, these bilinguals had lived on average only four years in a Spanish-speaking country; the possibility that their perception of Spanish stress would improve with more exposure cannot be excluded. Perhaps more surprisingly, even early bilinguals sometimes fail to show native-like perception of their second language (Mack 1989; Pallier *et al.* 1997; Sebastián-Gallés and Soto-Faraco 1999). For instance, Pallier *et al.* (1997) tested fluent Spanish–Catalan adult bilinguals, who had started to learn Catalan before the age of six and who use Catalan on a daily basis. Spanish has only one mid front vowel, whereas Catalan has two, /e/ and /ɛ/. The early bilinguals were found to have difficulty perceiving the Catalan vowel contrast.

Three factors have been proposed that influence the strength of perceptual assimilation in bilinguals. First, in the realm of segmental perception, the Speech Learning Model (SLM) of Flege (1995) states that the greater the perceived dissimilarity of an L2 sound from the closest L1 sound, the better the L2 sound is learned, both in perception and production. Second, L2 features that play a role – even a non-contrastive one – in the native language have been argued to be less difficult than others. For instance, McAllister *et al.* (2002) compared native speakers of English and Spanish living in Stockholm. They showed that the former perceive Swedish length better than the latter. Indeed, in English, length is used allophonically, making it easier for native speakers of English to learn the Swedish length feature. Finally, the amount of usage of the native language appears to be correlated negatively with performance in the second language. That is, the less bilingual speakers use their native language, the better they perceive the sounds of their second language, possibly becoming even indistinguishable from native speakers of the latter language (Flege and MacKay 2004).

2.3 *Perception and loanword adaptations*

One area of *phonological* research in which perceptual assimilation of non-native sounds and sound sequences appears to be relevant is that of loanword adaptations. In language contact situations, words of one language that are introduced into another language will typically (though not always) be adapted to the phonological pattern of the latter. Loanword adaptations thus transform non-native segments and suprasegments, as well as phonotactic structures, into native ones (see CHAPTER 95: LOANWORD PHONOLOGY). Given the fact that words must be perceived before they can be produced, it seems necessary that loanword phonology take into account the perception of non-native words by speakers of the borrowing language.

In phonological accounts of loanword adaptations, it is generally assumed that the input to the adaptations is constituted by the surface form of the source language, and that the adaptations are computed by the phonological grammar of the borrowing language. In processing terms, this means that during perception, the phonetic form of the source words is faithfully copied onto an abstract underlying form, and that adaptations are transformations produced by the standard phonological processes in production. Some researchers explicitly argue that foreign words can indeed be perceived without distortions and that speech perception is hence irrelevant to loanword adaptations (Paradis and LaCharité 1997; Jacobs and Gussenhoven 2000; LaCharité and Paradis 2005; see also CHAPTER 76: STRUCTURE PRESERVATION: THE RESILIENCE OF DISTINCTIVE INFORMATION). They are correct in as much as – as shown in §2.2 – at some level of representation non-native contrasts are indeed discriminated (after all, such contrasts present – sometimes quite large – acoustic differences). However, whether discrimination at an acoustic level of representation suffices to faithfully import non-native forms into one's lexicon is a different question. Given the phonetic variability due to inter- and intra-speaker variation, as well as the environmental noise in which speech is embedded, it seems reasonable to expect that listeners should discriminate non-native contrasts at a more abstract level of representation in order to store them faithfully. This is where the psycholinguistic literature becomes relevant, since it is exactly at this abstract level that listeners have been shown to experience difficulty perceiving non-native contrasts.

Several researchers have argued that at least *certain* loanword adaptations take place during perception, due to difficulties with the perception of non-native sound patterns. The first explicit proposal was made by Silverman (1992), based on the adaptations of English loanwords in Cantonese. Silverman proposed a two-stage model of loanword adaptations, containing a perceptual and an operative stage. During the first stage, the surface form of the source language is mapped onto an underlying form in the borrowing language on a segment-to-segment basis. This context-free mapping involves a first part of transformations that are due to misperception of non-native segments. In particular, segments from the source language that are illegal in the borrowing language are perceived as legal ones (for instance, *bus* would be perceived as [pas], the Cantonese phoneme inventory containing /p/ and /a/, but not /b/ and /ʌ/). The perceptual level is also held responsible for the deletion of segments with a low acoustic saliency (for instance, *lift* would be perceived as [lif]). The output of the perceptual stage serves as the input to the operative stage, during which entire word forms rather than individual segments are evaluated. Adaptations take place whenever the phonological structure of the borrowing language is not respected. For instance, given that fricatives cannot occur in Cantonese syllable codas, the output [lif] of the perceptual level is transformed into [lip]. Other non-native coda consonants, as well as consonant clusters, undergo vowel epenthesis; for instance, the output [pas] of the perceptual level is transformed into [pasi].

Silverman's (1992) article, along with Yip's (1993) constraint-based reinterpretation of his model, has provided the impetus for an extensive literature on loanword adaptations, much of which is partly or wholly devoted to the issue of the role of perception (see CHAPTER 95: LOANWORD PHONOLOGY and references therein). Two questions in particular have been debated, concerning how much of loanword adaptation is due to perception, and how the role of perception is to be modeled. As to the first question, most researchers nowadays defend an intermediate position, arguing that perception does play a role in loanword adaptations, but that it cannot explain all effects. Oftentimes, arguments in favor of a perceptual account of adaptation patterns are impressionistic. For instance, word-final stop consonants are typically argued to lack acoustic saliency (especially if they are unreleased) and therefore to be prone to deletion during speech perception. Explicit reference to the psycholinguistic literature, however, has also been made. Specifically, comparing this literature to loanword data, Peperkamp and Dupoux (2003) observed a number of correspondences between adaptation patterns on the one hand and perceptual distortions of non-native phonological structure on the other hand. Some of these correspondences concern transformations applying at the level of individual segments. For instance, Japanese listeners' difficulty with the perception of the English consonants /ɹ/ and /l/ (Goto 1971) is reflected in loanwords from English, where these consonants are both adapted as /ɾ/ in Japanese (Lovins 1975). Others concern suprasegmental structure. For instance, French listeners' difficulty in perceiving stress contrasts (Dupoux *et al.* 1997) is reflected in loanwords in French, with stress being systematically word-final, regardless of the position of stress in the source word. Still others concern syllable structure. For instance, the illusory vowel that Japanese listeners perceive within consonant clusters (Dupoux *et al.* 1999) is reflected in the epenthetic vowel by which such clusters are broken up in loanwords in Japanese (Lovins 1975). This last example is especially revealing, since it contrasts with the intuition first expressed by Silverman (1992) that perception plays a role in cases of segment *deletion*, but not in those of segment *insertion*.

Speech perception experiments that specifically aim at comparing loanword data to the perception of non-native sound patterns have also been carried out (Takagi and Mann 1994; Kim and Curtis 2002; Peperkamp *et al.* 2008). For instance, Peperkamp *et al.* (2008) examined an asymmetry in French and English [n]-final loanwords in Japanese: the former but not the latter are adapted with a final epenthetic vowel (cf. [kannu] < Fr. *Cannes* [kan] and [pen] < Eng. *pen*). They showed that Japanese listeners perceive a vowel at the end of [n]-final non-words produced by French speakers but not when they are produced by English speakers. Based on these findings, they argued that the asymmetry in the loanword adaptation pattern originates in the perception of French and English words by Japanese listeners.

Concerning the question as to how the role of perception on loanword adaptations should be modeled, Silverman (1992) considered perceptually driven adaptations to be pre-grammatical, in the sense that they are computed before the phonological grammar per se comes into play. That is, they are influenced by the phonology of the native language, but not computed by the phonological grammar (see also Yip 1993; Peperkamp *et al.* 2008). Other researchers, however, have modeled perception-driven adaptations with grammatical tools, either as part of a perception grammar (Kenstowicz 2004; Boersma and Hamann 2009), or by incorporating constraints demanding perceptual similarity into the production grammar (Kang 2003; Adler 2006; Kenstowicz and Suchato 2006). The former appear to have the advantage that they can account straightforwardly for the fact that loanword adaptations sometimes conflict with native alternations.[2] For instance, in Korean, obstruents turn into nasals before nasal consonants (e.g. /kukmul/ → [kuŋmul] 'soup'), but in loanwords, obstruent + nasal sequences undergo epenthesis (e.g. [pʰikʰɨnik] < *picnic*). If loanword adaptations are computed by the phonological production grammar, this grammar thus needs to be able to distinguish between native words and loanwords; by contrast, if loanword adaptations are

[2] See Peperkamp *et al.* (2008) for an extensive list of examples, and CHAPTER 95: LOANWORD PHONOLOGY for discussion.

computed by a perception grammar, conflicts between native alternations and loan-word adaptations are a natural consequence of the fact that they are computed by distinct mechanisms, one in production and the other in perception.

3 Perceptual influences in phonology

3.1 *The influence of perception on phonology: Evidence*

Many phonological processes reduce the chances of a listener incorrectly perceiving certain sounds, particularly in contexts where they might be easily confused with other, highly similar sounds. Phonologies, in other words, appear to be optimized with respect to speech perception. In this section we survey a number of examples of this optimization; in §3.2 we discuss possible reasons for this typological bias.

In the previous section, we looked at evidence that speech perception is shaped by one's native language experience, making some distinctions easier to perceive than others. Despite the language-specific nature of perception, however, there are universal hierarchies of perceptibility imposed by the nature of the human auditory system. Some distinctions, for example those that involve greater acoustic distances, are simply more perceptible, *ceteris paribus*, than others. Although languages may differ in the phonemic contrasts they make use of, some contrasts are thus less marked than others. It is these universal differences in perceptibility that we refer to throughout this section.

As Flemming (2004) points out, perceptual markedness is best understood as a property of distinctions between sounds rather than of individual sounds themselves. Speech perception involves segmenting raw acoustic input and assigning each segment the appropriate category label. The probability that a given segment will be correctly categorized depends on what other categories it might be confused with, and where precisely the boundary between categories lies. It is thus meaningless to claim that a given sound A is difficult to perceive – it is rather the difference between two sounds A and B that is difficult to perceive. A *weak contrast* is a phonemic contrast involving two such hard-to-distinguish sounds.

The problem presented by a weak contrast can be solved in two ways: the weak contrast may be enhanced, making it easier to perceive, or it may be neutralized, making its correct perception unnecessary (Hayes and Steriade 2004). The remainder of this section discusses several examples of each type of process. We also examine the claim that perceptual similarity plays a role in processes that relate different surface forms to each other.

3.1.1 *Contrast enhancement*

Stevens *et al.* (1986) note that the distinctiveness of contrasts tends to be enhanced by the use of redundant features – high vowels, for example, are typically distinguished not only by their backness but also by rounding (CHAPTER 2: CONTRAST). This is one example of a more general tendency for languages to maximize the acoustic distance between vowels, an observation that has a long history in linguistic theory (e.g. Jakobson 1941; Wang 1971). When the vowels of a language are plotted in F1–F2 space, inventories like that in (1a) are typical, while inventories like the one in (1b) are unattested (although see Lass 1984 for a discussion of possible counterexamples).

(1) a. *Evenly spaced vowel inventory* b. *Unevenly spaced vowel inventory*

i		u				
e		o		e	ə	
	a			æ	a	ɑ

Liljencrants and Lindblom (1972) show that a simple quantitative model, which essentially maximizes the total distance[3] among a given number of points in F1–F2 space, provides a reasonably good fit to attested vowel systems (see also Lindblom 1986; Vallée 1994; Schwartz *et al.* 1997). Actual vowel inventories thus appear to be optimized in the sense that they minimize the risk of misperception by dispersing vowels as widely as possible throughout the available perceptual space.[4] Similar claims have also been made regarding consonant inventories (Padgett 2001; Padgett and Zygis 2003; Flemming 2004; Gallagher 2008).[5]

Dispersion theory, as it has come to be known, has also been used to analyze historical sound change (CHAPTER 93: SOUND CHANGE). Padgett and Zygis (2003) argue that a change in Polish in which palatalized palato-alveolar sibilants became retroflex sibilants (e.g. [ʃʲija] > [ʂija] 'neck') was motivated by the existence of an alveolo-palatal series of sibilants in Polish. Using evidence from perceptual studies, they show that alveolo-palatals are more difficult to distinguish from palato-alveolars than from retroflexes for both native Polish and English speakers. The sound change in Polish was thus perceptually optimizing in the sense that it improved the overall discriminability of sibilants.

Another case of perceptually driven contrast enhancement can be seen in the Sonority Sequencing Constraint, which requires that segments in a syllable's onset and nucleus should occur in order of increasing sonority (Sievers 1881; Jesperson 1904; Saussure 1916; Hooper 1976; Kiparsky 1981; Steriade 1982; Clements 1990; CHAPTER 49: SONORITY). Wright (2004) argues that this cross-linguistically robust generalization can be understood as a way of maximizing the perceptibility of place, manner, and voicing cues in vowels and consonants. A similar analysis is presented by Ohala and Kawasaki-Fukumori (1997), who suggest that perceptibility can also explain the low frequency of the sequences /ji/ and /wu/ in many languages, as well as the fact that the CV syllable is a more common syllable type than VC, because the cues to the place of the consonant are more robust in a transition into a following vowel than in a transition out of a preceding vowel (Fujimura *et al.* 1978).[6]

[3] Technically, their model minimizes the sum, over all possible vowel pairs, of the inverse square of the perceptual distance between each pair.

[4] Other factors, such as homophony avoidance, are presumed to regulate the number of vowels, preventing languages from adopting a single-vowel inventory, which is ideal from a purely perceptual standpoint (Flemming 2004).

[5] For more on dispersion theory, including discussion of possible formalizations, see Flemming (2002, 2004, 2005), Padgett (1997), and Ní Chiosáin and Padgett (2001).

[6] Although see Moreton *et al.* (2008) for experimental evidence of an asymmetry between CV and VC syllables that cannot be explained by this perceptual account.

A perception-based account of the Sonority Sequencing Constraint explains why sibilants, whose cues do not depend heavily on transitions into vowels, often violate the constraint. Several languages employ metathesis (CHAPTER 59: METATHESIS), whether synchronically or diachronically, in stop–sibilant clusters, converting a stop–sibilant–vowel (TSV) sequence into a sibilant–stop–vowel (STV) sequence, as in the examples from Udi in (2).

(2) *Udi metathesis* (Schultze 2002, cited in Hume 2004)

/tad-esun/	→	[tast'un]	'to give'
/t'it'-esun/	→	[t'ist'un]	'to run'
/etʃ-esun/	→	[eʃtʃ'un]	'to bring'

Metathesis of this kind, which places the stop in a position where its place cues are the strongest, is attested in Udi, Faroese, and Lithuanian (Hume 2004), while the opposite pattern (i.e. one that converts STV to TSV) is unattested (but see Blevins 2009 for a critique of this analysis).

3.1.2 *Contrast neutralization*

Perception can also shape phonologies in ways that involve the elimination of contrasts rather than their preservation (CHAPTER 80: MERGERS AND NEUTRALIZATION). One such example is the common process of nasal place assimilation, whereby nasal consonants adopt the place of articulation of the following consonant (CHAPTER 81: LOCAL ASSIMILATION). The ubiquity of this process may be related to the difficulty listeners have in determining a nasal's place of articulation when it occurs before a consonant (Fujimura *et al.* 1978; Hura *et al.* 1992; Ohala 1990; Beddor and Evans-Romaine 1992, 1995). Listeners who know that their language employs a rule of place assimilation do not need to attend to the place cues of the nasal, thus minimizing the risk of misidentification.

Another example comes from Ohala (1990), who argues that in phonological processes, or cases of historical change, in which heterorganic stop clusters become geminate stops, it is the features of the second stop in the cluster that are typically adopted by the geminate (e.g. Sanskrit *bhaktum* > Pali *bhattum*, not **bhakkum*). He explains this asymmetry as the result of the fact that cues to stop place are more salient in the transition from the stop to a following vowel than they are from a preceding vowel to the stop. The place of the second stop in the cluster is therefore more perceptible, and it is not surprising that when geminates are formed from such clusters, this place is chosen over that of the initial stop. The place contrast in stops is thus eliminated precisely where it is difficult to perceive (CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS).

Steriade (1999) shows that voicing contrasts in obstruents tend to be permitted intervocalically, where cues to voicing are strongest, and neutralized preconsonantally, where cues are weak (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). Cues to apical contrasts, such as that between [d] and [ɖ], on the other hand, follow a different pattern, being strongest after vowels and weakest after consonants. This correlates with the typology of processes that affect apicality – languages tend to eliminate such contrasts postconsonantally and preserve them post-vocalically. Place assimilation, for instance in apical consonant clusters, is overwhelmingly progressive ([ɖd] → [ɖɖ]), while major place assimilation is typically regressive ([np] → [mp]) (Steriade 2001).

3.1.3 *Similarity maximization*

A third category of perceptual effects in phonology, in addition to contrast enhancement and neutralization, could be called *similarity maximization*. Steriade (2009) observes that phonological processes tend to make changes to underlying forms that are perceptually minimal. For example, although many languages require adjacent obstruents to agree in voicing, every such language repairs disagreeing obstruent clusters through assimilation, by either voicing or devoicing one of the segments in question. No language deletes one obstruent, or inserts a vowel between the obstruents, even though both of these processes would produce a surface form that no longer violates the prohibition on voicing disagreement. Steriade argues that this is because assimilation results in a surface form that is more similar to the underlying form (or to morphologically related surface forms) than would result from deletion or epenthesis.[7] This bias toward minimal changes demonstrated in the typology of phonological processes can be thought of as optimizing, in that it increases the chances of a listener or learner correctly recognizing multiple realizations of the same morpheme in different contexts (although one must be careful to rule out alternative analyses based on ease of production, which in many cases predict the same types of changes).

Similarity maximization has also been documented extensively by Fleischhacker (2005) in two other phenomena, loanword adaptation and reduplication. Fleischhacker shows that these processes operate in ways that tend to minimize the perceptual difference between the two strings. For example, she shows that in some languages with reduplication, the reduplicant faithfully copies a consonant cluster if it consists of a sibilant and a stop (as in Gothic *ste-stald*), but not if the consonants are an obstruent and a sonorant (as in Gothic *ge-grot*), whereas the opposite pattern is not attested (CHAPTER 100: REDUPLICATION). Using evidence from perception experiments, partial rhymes in poetry (Zwicky 1976; Kawahara 2007), and imperfect puns (Zwicky and Zwicky 1986), Fleischhacker argues that this asymmetry is due to the fact that (to use the Gothic example) *ge-* and *gre-* are more perceptually similar to one another than *se-* and *ste-*.

3.2 *The influence of perception on phonology: Mechanisms*

Although there is widespread agreement that the facts of speech perception shape phonologies, there is less consensus on the causal mechanisms involved. Two types of explanation have been proposed. A *misperception* account explains perceptually optimizing processes as accidental results of language learners' confusing sounds in contexts where the cues needed to distinguish them are weak or absent (Ohala 1981, 1993; Hale and Reiss 2000; Blevins 2004, 2006, 2007). On this view, listeners' errors accumulate in the lexicon, resulting in a state in which a given process is instantiated in the data to which future generations of learners

[7] Steriade defines similarity on the basis of the perceptual confusability of two surface forms, although of course alternative definitions which take into account more abstract properties of the output are possible.

are exposed. These misperceptions could occur in a number of ways (Blevins 2004): co-articulation effects may be misinterpreted as phonological alternations (resulting in assimilation), listeners may attempt to undo a co-articulatory effect that they believe has taken place (resulting in dissimilation), or listeners may hear a range of variant pronunciations of a sound, and misidentify which is the prototypical variant.

Under a misperception account of nasal place assimilation, for example, learners would tend to misperceive [anpa] as [ampa]; once enough of these mistakes have been made, learners will posit a process of place assimilation, even if no such process was employed by the previous generation. Of course, for misperceptions to result in perceptually optimizing phonologies, confusion between sounds must be asymmetric; in the example above, listeners must mishear [np] as [mp] more often than they mishear [mp] as [np]. In an experiment carried out by Ohala (1990), English-speaking subjects did indeed misperceive heterorganic nasal–stop clusters as homorganic clusters 93 percent of the time (see Beddor and Evans-Romaine 1995 for a similar experiment with an 80 percent homorganic error rate). However, it should be noted that the fact that nasal–stop clusters in English are overwhelmingly homorganic (Hay *et al.* 2004) may have biased the subjects' perception; as discussed in §2.2, there is evidence that categorical native-language phonotactics can affect listeners' perception of ambiguous stimuli (Massaro and Cohen 1983; Moreton 1999; Moreton and Amano 1999), and it is plausible that gradient phonotactics may have a similar effect (CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY; CHAPTER 90: FREQUENCY EFFECTS). Further research is called for to identify asymmetries in misperception which are independent of language experience, and to determine the extent to which they align with the directionality of common phonological processes.

Another possible critique of this approach challenges the claim that all perceptual effects are the outcome of misperceptions. Flemming (2005) argues that misperception can only account for processes that eliminate contrasts, as in the nasal place assimilation example, and not for processes that enhance contrasts, as in the dispersion of vowel inventories. Indeed, it is hard to see at first how perceptual errors could result in making a distinction clearer. Much recent work using computational models, however, has shown that under the right conditions *self-organization* (Nicolis and Prigogine 1977) can lead to contrast enhancement even when the perception of categories is disturbed by random noise. De Boer (2001) demonstrates that a group of simulated agents which attempt to imitate each others' vowel productions evolve realistically dispersed vowel inventories over time, even in the absence of any agent-internal biases toward optimization.

In de Boer's model, an agent represents each vowel category as a cloud of tokens, representing all of the vowels heard by the agent which were classified as belonging to that category. It thus strongly resembles *exemplar theory*, the hypothesis that highly phonetically detailed tokens of sounds (or larger units such as words) are stored by listeners, and that phonological categories consist of sets of these tokens (Pierrehumbert 2001). Blevins and Wedel (2009) identify two processes that can maintain category distinctness in an exemplar model: *variant trading*, in which ambiguous exemplars (i.e. those near the border between two categories) are more likely to be misidentified than prototypical exemplars, and *variant pruning*, in which ambiguous exemplars are more likely to fail to be assigned to any category at all. Both processes have the effect of pushing categories apart, since exemplars further from the border between the categories have a higher chance of surviving (i.e. being correctly identified), and thus a greater influence over the center of the category as a whole. This work, together with other recent research using agent-based simulations (Wedel 2006, 2007; Kochetov and So 2007; Boersma and Hamann 2008; CHAPTER 6: SELF-ORGANIZATION IN PHONOLOGY), has shown that surprising and counterintuitive effects can emerge globally in a system of very simple interacting agents, and provides an important "proof of concept" for the misperception approach to perceptual effects on phonology.

An alternative to the misperception account is what we will call the *cognitive bias* approach, which sees phonological processes as being implemented *in order to* minimize the possibility of listener errors (Hayes 1999; Steriade 1999, 2001, 2009; Hayes and Steriade 2004). On this view, optimization is an explicit goal of phonological systems. A cognitive bias account would explain the ubiquity of perceptually optimized phonological processes as the result of knowledge on the part of learners which biases learning in some way. There are several ways these biases might manifest themselves. In the case of nasal place assimilation, for example, learners who know that homorganic NC clusters are "better" than heterorganic clusters, for example, might simply be unable to learn a language in which [np] is legal, but [mp] is not. Alternatively, learners may come equipped with a soft bias against the pathological heterorganic-clusters-only language – they could learn such a grammar, but only when given much more evidence than would be required to learn a homorganic-clusters-only language. If these biases are universal, then even if sound change were random and unconstrained, those languages that conform to learners' expectations would flourish at the expense of those that do not.

One way to formalize these cognitive biases is to posit optimality-theoretic (OT) markedness constraints which prohibit problematic sequences, which interact with faithfulness constraints to produce the observed typological distribution of languages (CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). Nasal place assimilation could be modeled as the effects of a constraint against heterorganic clusters and the absence of a corresponding constraint banning homorganic clusters. This type of analysis, however, would not work in other cases. How could dispersion effects be explained with constraints on output forms? In languages with one high vowel it will be central, but languages with two high vowels ban the central vowel in favor of two peripheral vowels. The grammar cannot "know" whether a front high vowel is marked or not without knowing what other vowels the language uses. The problem stems from the fact, mentioned earlier, that perceptual difficulty is a property of contrasts between sounds, not of individual sounds or sound sequences. Under standard formulations of OT (Prince and Smolensky 2004), however, phonological contrasts are emergent properties of an entire constraint ranking, and as such cannot be referred to directly by markedness constraints. Two examples of ways around this problem are represented by Flemming's (2002) *distinctiveness constraints* and Steriade's *P-map* (Steriade 2009).

Flemming (2002) solves the problem by expanding the definition of OT markedness constraints to allow them to refer to contrasts, or to entire sound inventories. These distinctiveness constraints are of two types, MAXIMIZECONTRASTS constraints, which prefer a greater number of distinct sounds, and perceptually motivated MINIMIMUMDISTANCE constraints, which ban inventories in which sounds are not sufficiently distinct. A learner equipped with these constraints will be able to posit a ranking that will generate a typical, well-dispersed vowel inventory, but no ranking will generate an unattested inventory, such as one in which there are only mid and low vowels.

Steriade's approach is to posit a "map" of perceptibility, realized as a matrix of sound pairs and contexts, with a similarity value assigned to each pair in each context. This "P-map" is essentially a representation of a speaker's knowledge of perceptual similarity (whether innate or somehow deduced by the speaker). The P-map is held to regulate rankings among faithfulness constraints, and it is these fixed rankings that result in perceptually optimizing typological biases. Recall the case discussed in §3.1.3 of voicing assimilation in obstruent clusters, which is preferred over deletion as a solution to voicing disagreement. According to Steriade, speakers faced with a cluster like /kd/ know that the difference between [k] and [g] is smaller than the difference between [k] and Ø. This translates into the fixed constraint ranking MAX-C >> IDENT[voice], with the consequence that no possible ranking will generate the unattested change /kd/ → [d].

Evaluating which of these theories is most likely to be true is made difficult by the fact that their empirical predictions largely overlap. We conclude this section by considering ways in which they may be teased apart. The first concerns parsimony. As many proponents of the misperception approach argue, if typological asymmetries can be shown to emerge without the need for a complex set of synchronic cognitive biases, then the diachronic analysis is the simplest, and should therefore be preferred. It is customary that in the absence of any empirical evidence, simpler explanations should be adopted over more complex ones, and it is surely healthy to question assumptions about the supremacy of synchronic over diachronic analyses. But ultimately, the question will be decided on empirical grounds.

A more empirical way of distinguishing the two approaches begins by independently establishing the plausibility of the mechanisms proposed by each theory. In the misperception case, this can take two forms. First, experiments can be used to determine under what conditions listeners in fact make perceptual errors, and the results can be compared to the typological facts. Several experiments (Fujimura *et al.* 1978; Ohala 1990; Hura *et al.* 1992) have shown, for example, that nasal place cues are difficult to identify before consonants, the same environment which frequently triggers nasal place assimilation. Guion (1998) shows that the sequences [ki] and [ʧi] are highly confusable when compared to similar sequences, which may explain why velar palatalization before [i] is a common process synchronically and diachronically. Kochetov and So (2007) show that the perceptibility of released stops in various environments correlates with cross-linguistic patterns of place assimilation.

Second, because misperception accounts rely on effects that emerge globally within a network of speakers and listeners, rather than inside the head of each individual, computer simulations involving multiple speakers can be used to explore how biases develop and propagate throughout a speech community. As discussed above, de Boer's (2001) work on the evolution of vowel systems falls under this heading, as does Wedel's (2006) research into the emergence and maintenance of linguistic contrasts. Boersma and Hamann (2008) show that optimally dispersed sibilant inventories can evolve over multiple generations when agents produce sounds using the same grammar used for perception. Simulations such as these do not by themselves provide direct evidence of the misperception hypothesis, but they are a crucial step in delineating the precise empirical predictions made by the theory.

Similarly, proponents of the cognitive bias model have sought to establish its plausibility by looking for evidence for these biases in domains for which diachronic explanations do not apply. Kawahara (2007) shows, for example, that the lyrics of Japanese rap songs employ a type of line-final rhyme which has no precursor in Japanese poetic tradition. Kawahara shows that the rhymes attested in a corpus of lyrics adhere to a hierarchy of perceptual similarity which plays no role in the rest of Japanese phonology, suggesting that the Japanese speakers who composed these lyrics possess a knowledge of perceptual similarity relations that is not learned.

Wilson (2006) comes to a similar conclusion using an artificial language learning experiment. He shows that English-speaking subjects who are taught a velar palatalization rule which changes /k/ to [ʧ] before [e] generalize that rule to contexts in which the vowel is [i], while subjects taught a rule that only applies before [i] do not generalize it to [e]. Wilson attributes these results to a learning bias that is motivated by the subjects' (presumably unconscious) knowledge that [ki] and [ʧi] are more perceptually similar than [ke] and [ʧe] (Guion 1998). Skoruppa *et al.* (forthcoming) also use an artificial language learning experiment to show that subjects find it easier to learn phonological rules which implement small changes.

Of course, evidence that adults use biases when faced with novel phonological situations does not itself constitute evidence that children use the same biases when acquiring their native language (CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY). The next step, for both theories, is to look for evidence that their proposed mechanisms are in fact operating in the real world of language learning and use. How often do language learners, outside of a laboratory, actually misperceive sounds, and how frequently must misperception occur in order to allow a given sound change to spread throughout a speech community? What evidence is there that children employ learning biases in the early stages of acquisition?[8] To what extent are typological patterns driven by historical forces other than sound change, for example the preferential adoption or retention of some words over others (Martin 2007)? Answering these questions would be difficult, but not impossible, and the relevant research has for the most part yet to be done. Data could come from detailed examinations of diachronic changes, either those currently in progress or those for which there is a written record, or from experiments involving numbers of interacting subjects that more closely resemble an actual language learning environment.

In a sense, both theories agree on the fact that linguistic sound systems are optimized for human perception – they disagree only on how this optimization takes place. In the misperception approach, optimization takes place globally, as the overall effect of many "innocent misapprehensions." The cognitive bias theory claims that optimization is local, occurring inside the head of each speaker. The two theories are not mutually incompatible (see Moreton 2008 for discussion). It is certainly possible that biases in diachronic change shape phonological systems, and that learners have also become biased so as to more easily learn the types of phonological patterns that they are most likely to encounter.

[8] The experimental studies in Saffran and Thiessen (2003) and Cristià and Seidl (2008) suggest that phonotactic learning in infants is indeed biased in favor of certain types of pattern.

4 Conclusion

Learning the phonology of a language involves, among other things, adjusting one's expectations in accordance with the language one has already heard. This makes the learner better able to perceive the distinctions used by her native language, albeit at the cost of a reduced ability to perceive non-native distinctions. On this view, the fact that the native phonology influences perception is not surprising – it is simply a consequence of the trade-off involved in tuning the perceptual system to expect input of a certain type. In this chapter, we surveyed evidence that the perception of phonemic and allophonic categories is biased by the phonemic categories and phonotactic restrictions of one's native language, and that these factors also influence the adaptation of words borrowed from other languages. Further research on this topic promises to give us a clearer picture of exactly which properties of the native language influence speech perception.

We also discussed evidence of influence in the other direction, from perception to phonology. This influence operates over historical time, changing languages in the direction of eliminating difficult-to-perceive contrasts, whether by neutralizing the problematic contrasts, or by making them more distinct, ultimately resulting in a linguistic typology which appears to be perceptually optimized. Current debate on this topic is centered on the question of what drives this historical change – synchronically active learning biases, or perceptual errors on the part of language learners. Both hypotheses make strong claims regarding the fundamental structure of phonological grammars, and the results of this debate have the potential to shape the direction phonological theory will take in the coming decades.

ACKNOWLEDGMENTS

Research for this chapter was supported by the Agence Nationale de la Recherche (ANR-BLAN-0065-01) and the Centre National de la Recherche Scientifique. We are grateful to Katrin Skoruppa for comments and discussion.

REFERENCES

Adler, Allison N. 2006. Faithfulness and perception in loanword adaptation: A case study from Hawaiian. *Lingua* 116. 1024–1045.

Beddor, Patrice Speeter & David Evans-Romaine. 1992. Acoustic-perceptual factors in nasal place assimilation. *Journal of the Acoustical Society of America* 91. 2473.

Beddor, Patrice Speeter & David Evans-Romaine. 1995. Acoustic-perceptual factors in phonological assimilations: A study of syllable-final nasals. *Rivista di Linguistica* 7. 145–174.

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Strange (1995), 171–206.

Best, Catherine T., Gerald McRoberts & Nomathemba Sithole. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14. 345–360.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns.* Cambridge: Cambridge University Press.

Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32. 117–166.

Blevins, Juliette. 2007. Interpreting misperception: Beauty is in the ear of the beholder. In Maria-Josep Solé, Patrice Streeter Beddor & Manjari Ohala (eds.) *Experimental approaches to phonology*, 144–154. Oxford: Oxford University Press.

Blevins, Juliette. 2009. Phonetic explanations for recurrent sound patterns: Diachronic or synchronic? In Eric Raimy & Charles Cairns (eds.) *Contemporary views on architecture and representations in phonology*, 325–336. Cambridge, MA: MIT Press.

Blevins, Juliette & Andrew Wedel. 2009. Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26. 143–183.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Boer, Bart de. 2001. *The origins of vowel systems*. Oxford: Oxford University Press.

Boersma, Paul & Silke Hamann. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25. 217–270.

Boersma, Paul & Silke Hamann. 2009. Loanword adaptation as first-language phonological perception. In Andrea Calabrese & W. Leo Wetzels (eds.) *Loan phonology*, 11–58. Amsterdam & Philadelphia: John Benjamins.

Boomershine, Amanda, Kathleen Currie Hall, Elizabeth Hume & Keith Johnson. 2008. The impact of allophony versus contrast on speech perception. In Peter Avery, B. Elan Dresher & Keren Rice (eds.) *Contrast in phonology: Theory, perception, acquisition*, 146–172. Berlin & New York: Mouton de Gruyter.

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In Kingston & Beckman (1990), 283–333.

Cristià, Alejandrina & Amanda Seidl. 2008. Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development* 4. 203–227.

Dupoux, Emmanuel, Christopher Pallier, Núria Sebastián-Gallés & Jacques Mehler. 1997. A destressing "deafness" in French? *Journal of Memory and Language* 36. 406–421.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25. 1568–1578.

Dupoux, Emmanuel, Sharon Peperkamp & Núria Sebastián-Gallés. 2001. A robust method to study stress "deafness." *Journal of the Acoustical Society of America* 110. 1606–1618.

Dupoux, Emmanuel, Núria Sebastián-Gallés, Eduardo Navarrete & Sharon Peperkamp. 2008. Persistent stress "deafness": The case of French learners of Spanish. *Cognition* 106. 682–706.

Flege, James Emil. 1995. Second-language speech learning: Theory, findings, and problems. In Strange (1995), 229–273.

Flege, James Emil & Ian R. A. MacKay. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26. 1–34.

Flege, James Emil, Ocke-Schwen Bohn & Sunyoung Jang. 1997. The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics* 25. 437–470.

Fleischhacker, Heidi. 2005. Similarity in phonology: Evidence from reduplication and loan adaptation. Ph.D. dissertation, University of California, Los Angeles.

Flemming, Edward. 2002. *Auditory representations in phonology*. London & New York: Routledge.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Hayes *et al.* (2004), 232–276.

Flemming, Edward. 2005. Speech perception in phonology. In David B. Pisoni & Robert E. Remez (eds.) *The handbook of speech perception*, 156–182. Malden, MA & Oxford: Blackwell.

Fujimura, Osamu, Marian Macchi & Lynn Streeter. 1978. Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech* 21. 337–346.

Gallagher, Gillian. 2008. The role of contrast in laryngeal cooccurrence restrictions. *Proceedings of the West Coast Conference on Formal Linguistics* 27. 177–184.

Gbéto, Flavien. 2000. *Les emprunts linguistiques d'origine européenne en Fon (Nouveau Kwa, Gbe: Bénin).* Cologne: Rüdiger Köppe Verlag.

Goto, Hiromu. 1971. Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia* 9. 317–323.

Guion, Susan. 1998. The role of perception in the sound change of velar palatalization. *Phonetica* 55. 18–52.

Hale, Mark & Charles Reiss. 2000. "Substance abuse" and "dysfunctionalism": Current trends in phonology. *Linguistic Inquiry* 31. 157–169.

Hallé, Pierre A. & Catherine T. Best. 2007. Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop + /l/ clusters. *Journal of the Acoustical Society of America* 121. 2899–2914.

Hallé, Pierre A., Juan Segui, Uli Frauenfelder & Christine Meunier. 1998. Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance* 24. 592–608.

Hay, Jennifer, Janet B. Pierrehumbert & Mary E. Beckman. 2004. Speech perception, well-formedness and the statistics of the lexicon. In John Local, Richard Ogden & Rosalind Temple (eds.) *Phonetic interpretation: Papers in laboratory phonology VI,* 58–74. Cambridge: Cambridge University Press.

Hayes, Bruce. 1999. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In Michael Darnell, Edith Moravcsik, Frederick Newmeyer, Michael Noonan & Kathleen Wheatley (eds.) *Functionalism and formalism in linguistics,* vol. 1: *General papers,* 243–285. Amsterdam & Philadelphia: John Benjamins.

Hayes, Bruce & Donca Steriade. 2004. Introduction: The phonetic bases of phonological markedness. In Hayes *et al.* (2004), 1–33.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology.* Cambridge: Cambridge University Press.

Hooper, Joan B. 1976. *An introduction to natural generative phonology.* New York: Academic Press.

Hume, Elizabeth. 2004. The indeterminacy/attestation model of metathesis. *Language* 80. 203–237.

Hume, Elizabeth & Keith Johnson (eds.) 2001. *The role of speech perception in phonology.* San Diego: Academic Press.

Hura, Susan L., Björn Lindblom & Randy L. Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35. 59–72.

Jacobs, Haike & Carlos Gussenhoven. 2000. Loan phonology: Perception, salience, the lexicon and OT. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: Phonology, syntax, and acquisition,* 193–210. Oxford: Oxford University Press.

Jaeger, Jeri. 1980. Testing the psychological reality of phonemes. *Language and Speech* 23. 233–253.

Jakobson, Roman. 1941. *Kindersprache, Aphasie, und allgemeine Lautgesetze.* Uppsala: Almqvist & Wiksell. Translated (1968) by A. R. Keiler as *Child language, aphasia and phonological universals.* The Hague & Paris: Mouton.

Jespersen, Otto. 1904. *Lehrbuch der Phonetik.* Leipzig & Berlin: Teubner.

Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20. 219–273.

Kawahara, Shigeto. 2007. Half rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics* 16. 113–144.

Kazanina, Nina, Colin Phillips & William J. Idsardi. 2006. The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences* 103. 11381–11386.

Kenstowicz, Michael. 2004. Review article: The role of perception in loanword phonology. (Review of Gbéto 2000.) *Studies in African Linguistics* 32. 95–112.

Kenstowicz, Michael & Atiwong Suchato. 2006. Issues in loanword adaptation: A case study from Thai. *Lingua* 116. 921–949.

Kim, Soohee & Emily Curtis. 2002. Phonetic duration of English /s/ and its borrowing into Korean. In Noriko M. Akatsuka & Susan Strauss (eds.) *Japanese/Korean Linguistics 10*, 406–419. Palo Alto: CSLI.

Kingston, John & Mary E. Beckman (eds.) 1990. *Papers in laboratory phonology I: Between the grammar and physics of speech.* Cambridge: Cambridge University Press.

Kiparsky, Paul. 1981. Remarks on the metrical structure of the syllable. In Wolfgang U. Dressler, Oskar E. Pfeiffer & John R. Rennison (eds.) *Phonologica 1980*, 245–256. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Kiriloff, Constantine. 1969. On the auditory perception of tones in Mandarin. *Phonetica* 20. 63–67.

Kochetov, Alexei & Connie K. So. 2007. Place assimilation and phonetic grounding: A cross-linguistic perceptual study. *Phonology* 24. 397–432.

LaCharité, Darlene & Carole Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry* 36. 223–258.

Lahiri, Aditi & William Marslen-Wilson. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38. 245–294.

Lass, Roger. 1984. Vowel system universals and typology: Prologue to theory. *Phonology Yearbook* 1. 75–111.

Liberman, Isabelle, Donald Shankweiler, F. William Fischer & Bonnie Carter. 1974. Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology* 18. 201–212.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48. 839–862.

Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John J. Ohala & Jeri J. Jaeger (eds.) *Experimental phonology*, 13–44. Orlando: Academic Press.

Lipski, Silvia & Klaus Mathiak. 2007. A magnetoencephalographic study on auditory processing of native and nonnative fricative contrasts in Polish and German listeners. *Neuroscience Letters* 41. 90–95.

Lively, Scott E., John S. Logan & David B. Pisoni. 1993. Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94. 1242–1255.

Lovins, Julie. 1975. *Loanwords and the phonological structure of Japanese.* Bloomington: Indiana University Linguistics Club.

Mack, Molly. 1989. Consonant and vowel perception and production: Early English–French bilinguals and English monolinguals. *Perception and Psychophysics* 46. 189–200.

Martin, Andrew. 2007. The evolving lexicon. Ph.D. dissertation, University of California, Los Angeles.

Massaro, Dominic W. & Michael M. Cohen. 1983. Phonological context in speech perception. *Perception and Psychophysics* 34. 338–348.

McAllister, Robert, James E. Flege & Thorsten Piske. 2002. Influence of L1 on acquisition of Swedish quantity in Spanish, English and Estonian. *Journal of Phonetics* 30. 229–258.

Miyawaki, Kuniko, Winifred Strange, Robert Verbugge, Alvin Liberman, James Jenkins & Osamu Fujimura. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics* 18. 331–340.

Morais, José, Paul Bertelson, Luz Cary & Jesus Alegria. 1986. Literacy training and speech segmentation. *Cognition* 24. 45–64.

Morais, José, Luz Cary, Jésus Alegria & Paul Bertelson. 1979. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* 7. 323–331.
Moreton, Elliott. 1999. Evidence for phonological grammar in speech perception. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.) *Proceedings of the 14th International Congress of Phonetic Sciences*, 2215–2217. Berkeley: Department of Linguistics, University of California, Berkeley.
Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25. 83–127.
Moreton, Elliott & Shigeaki Amano. 1999. Phonotactics in the perception of Japanese vowel length: Evidence for long-distance dependencies. In *Proceedings of the 6th European Conference on Speech Communication and Technology, Budapest*, 2679–2682.
Moreton, Elliott, Gary Feng & Jennifer L. Smith. 2008. Syllabification, sonority, and perception: New evidence from a language game. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 41. 341–355.
Näätänen, Risto, Anne Lehtokoski, Mietta Lennes, Marie Cheour, Minna Huotilainen, Antti Iivonen, Martti Vainio, Paavo Alku, Risto J. Ilmoniemi, Aavo Luuk, Jüri Allik, Janne Sinkkonen & Kimmo Alho. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385. 432–434.
Ní Chiosáin, Máire & Jaye Padgett. 2001. Markedness, segment realization, and locality in spreading. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 118–156. Cambridge: Cambridge University Press.
Nicolis, Gregoire & Ilya Prigogine. 1977. *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*. New York: Wiley.
Ohala, John J. 1981. The listener as a source of sound change. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 17(2). 178–203.
Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. In Kingston & Beckman (1990), 258–275.
Ohala, John J. 1993. Sound change as nature's speech perception experiment. *Speech Communication* 13. 155–161.
Ohala, John J. & Haruko Kawasaki-Fukumori. 1997. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In Stig Eliasson & Ernst Håkon Jahr (eds.) *Language and its ecology: Essays in memory of Einar Haugen*, 343–365. Berlin & New York: Mouton de Gruyter.
Otake, Takashi, Kiyoko Yoneyama, Anne Cutler & Arie van der Lugt. 1996. The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America* 100. 3831–3842.
Padgett, Jaye. 1997. Perceptual distance of contrast: Vowel height and nasality. In Rachel Walker, Mitoko Katayama & Daniel Karvonen (eds.) *Phonology at Santa Cruz* 5, 63–78. Santa Cruz: Linguistics Research Center.
Padgett, Jaye. 2001. Contrast dispersion and Russian palatalization. In Hume & Johnson (2001), 187–218.
Padgett, Jaye & Marzena Zygis. 2003. The evolution of sibilants in Polish and Russian. *ZAS Working Papers in Linguistics* 32. 155–174.
Pallier, Christophe, Laura Bosc & Núria Sebastián-Gallés. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64. B9–B17.
Paradis, Carole & Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33. 379–430.
Pegg, Judith E. & Janet F. Werker. 1997. Adult and infant perception of two English phones. *Journal of the Acoustical Society of America* 102. 3742–3753.
Peperkamp, Sharon & Emmanuel Dupoux. 2003. Reinterpreting loanword adaptations: The role of perception. In M. J. Solé, D. Recasens & J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*, 367–370. Barcelona: Causal Productions.
Peperkamp, Sharon, Michèle Pettinato & Emmanuel Dupoux. 2003. Reinterpreting loanword adaptations: The role of perception. In Barbara Beachley, Amanda Brown & Francis

Conlin (eds.) *Proceedings of the 27th Annual Boston University Conference on Language Development*, 650–661. Somerville, MA: Cascadilla Press.

Peperkamp, Sharon, Inga Vendelin & Kimihiro Nakamura. 2008. On the perceptual origin of loanword adaptations: Experimental evidence from Japanese. *Phonology* 25. 129–164.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam & Philadelphia: John Benjamins.

Pol'ivanov, Yevgeny. 1931. La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* 4. 79–96.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Read, Charles, Zhang Yun-Fei, Nie Hong-Yin & Ding Bao-Qing. 1986. The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition* 24. 31–44.

Saffran, Jenny R. & Erik D. Thiessen. 2003. Pattern induction by infant language learners. *Developmental Psychology* 39. 484–494.

Sapir, Edward. 1933. La réalité psychologique des phonèmes. *Journal de Psychologie Normale et Pathologique* 30. 247–265.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Lausanne & Paris: Payot.

Schultze, Wolfgang. 2002. Functional grammar of Udi. Unpublished ms., University of Munich.

Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée & Christian Abry. 1997. The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics* 25. 255–286.

Sebastián-Gallés, Núria & Salvador Soto-Faraco. 1999. On-line processing of native and nonnative phonemic contrasts in early bilinguals. *Cognition* 72. 112–123.

Sievers, Eduard. 1881. *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. Leipzig: Breitkopf & Härtel.

Silverman, Daniel. 1992. Multiple scansions in loanword phonology: Evidence from Cantonese. *Phonology* 9. 289–328.

Skoruppa, Katrin, Anna Lambrechts & Sharon Peperkamp. Forthcoming. The role of phonetic distance in the acquisition of phonological alternations. *Papers from the Annual Meeting of the North East Linguistic Society* 39.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT.

Steriade, Donca. 1999. Phonetics in phonology: The case of laryngeal neutralization. *UCLA Working Papers in Linguistics 2: Papers in Phonology* 3. 25–146.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Hume & Johnson (2001), 219–250.

Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Studies in honor of Paul Kiparsky*, 151–179. Cambridge, MA: MIT Press.

Stevens, Kenneth N., Samuel J. Keyser & Haruko Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In Joseph S. Perkell & Dennis H. Klatt (eds.) *Invariance and variability in speech processes*, 426–449. Hillsdale, NJ: Lawrence Erlbaum.

Strange, Winifred (ed.) 1995. *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press.

Swadesh, Morris. 1934. The phonemic principle. *Language* 10. 117–129.

Takagi, Naoyuki & Virginia Mann. 1994. A perceptual basis for the systematic phonological correspondences between Japanese loan words and their English source words. *Journal of Phonetics* 22. 343–356.

Takagi, Naoyuki & Virginia Mann. 1995. The limits of extended naturalistic exposure on the perceptual mastery of English /ɹ/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics* 16. 379–405.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht.

Vallée, Nathalie. 1994. Systèmes vocaliques: De la typologie aux prédictions. Thèse de Doctorat, Université Stendhal, Grenoble.

Wang, William S.-Y. 1971. The basis of speech. In Carroll E. Reed (ed.) *The learning of language*, 267–306. New York: Appleton-Century-Crofts.

Wang, Yue, Michelle M. Spence, Allard Jongman & Joan Sereno. 1999. Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America* 106. 3649–3658.

Wedel, Andrew. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23. 247–274.

Wedel, Andrew. 2007. Feedback and regularity in the lexicon. *Phonology* 24. 147–185.

Whalen, Douglas, Catherine T. Best & Julia R. Irwin. 1997. Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics* 25. 501–528.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30. 945–982.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Hayes *et al.* (2004), 34–57.

Yip, Moira. 1993. Cantonese loanword phonology and Optimality Theory. *Journal of East Asian Linguistics* 2. 261–291.

Zwicky, Arnold M. 1976. Well, the rock and roll has got to stop. Junior's head is hard as a rock. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 12. 676–697.

Zwicky, Arnold M. & Elizabeth Zwicky. 1986. Imperfect puns, markedness, and phonological similarity: With fronds like these, who needs anemones? *Linguistica* 20. 493–503.

99 Phonologically Conditioned Allomorph Selection

Andrew Nevins

Since Jakobson's (1948) analysis of Russian verbs as having one stem rather than two, it has been an attempt of generative phonology to minimize distinctly listed allomorphs in favor of phonological rules, some of which may be morphologically specific. Nonetheless there are certain allomorphs that cannot be derived one from the other, leading to "multiple URs" for the same morphemic category, such as the Moroccan 3rd person object clitic, which varies between the allomorphs *-h* (chosen after vowel-final stems) and *-u* (chosen after consonant-final stems), as will be discussed in §1.3 below. Such cases require distinct suppletive allomorphs, whose distribution is determined according to their phonological environments. The division of labor in dealing with allomorphy, then, is taken up both by purely phonological rules (such as those responsible for the voicing alternations of the English plural morpheme) and by morphological selection among separately listed allomorphs competing for insertion. In this chapter, we focus on the latter type of alternation between morphemes, while emphasizing that many cases of allomorphy not included in the present discussion are still best handled by phonological derivations from a single underlying form.

While a number of instances of allomorph choice depend on morphosyntactic and lexical factors (including conjugation or declension class), many cases of allomorph selection can be predicted on the basis of phonological configuration. Carstairs (1988) points out the relevance of phonological conditions on allomorph selection to morphological theory, but does not provide an implementation of how such allomorph selection should work. The tradition of analyzing allomorphy as multiple allomorphic input candidates that compete for the same morphemic realization and that are chosen among them for the output based on phonological selection begins with Mester's (1994) treatment of Latin stem augments in the perfect.[1] One of the first goals in discussing the role of phonological well-formedness in allomorph distribution is an explicit connection of these phenomena to well-established categories of phonological well-formedness

[1] However, this case will not be reviewed below, since Lapointe (1999) and Embick (2009) raise important caveats about the appropriateness of Mester's foot-based analysis for a full treatment of Latin.

based on cross-linguistic research. In this chapter I will present six different types of phonological conditions that force choice of one listed allomorph over another. I will then discuss cases which appear to be arbitrary but still reference phonological conditions, and thus require the mechanism of phonological subcategorization. Finally, I will proceed to a more general theoretical discussion about the level of phonological representation at which allomorph selection takes place, and the mechanism of choice.

1 Six phonological conditions forcing allomorph choice

I have attempted to categorize below a wide range of cases from typologically diverse languages into six basic categories of phonological markedness. As the case studies are chosen to make more general points of contact between the languages, in some cases I have chosen only the clearest examples, and direct the interested reader to the original articles for a fuller discussion of some of the complexities. The argumentation for particular analyses is necessarily abbreviated in this overview, and the reader is advised to consult the references provided for detailed argumentation in favor of certain analyses and against possible other analyses. In addition, I should mention that my choice of citations for these phenomena often reflect the most recent, or most theoretically relevant to the current discussion, rather than choosing the *locus classicus* on a particular case of allomorphy per se.

The six conditions are organized in terms of segment-level phenomena, syllable-level phenomena, and prosodic-level phenomena. I begin with segmental dissimilation, as it arises in the familiar case of English as well as many other languages, and then to segmental phonotactics. I then turn to syllable structure, whereby preference for onsetful, codaless, and less complex syllables plays a role in allomorph selection, and subsequently to morphological alignment of stems with syllables and syllabic constituents, which causes resistance to resyllabification in the relevant cases. Finally, I turn to stress-to-sonority and peak-prominence effects, which prefer sonorous stressed syllables and less sonorous unstressed syllables, and then to the optimization of foot structure.

1.1 *Segmental dissimilation*

The English Saxon genitive 's, often thought to be one of the most stoic and flexible of all inflectional markers in English, has in fact two allomorphs: 's and a zero allomorph, Ø (with devoicing and epenthesis processes well known from the plural also automatically occurring for the former, which is underlyingly /-z/; Pinker and Prince 1988). The zero allomorph is chosen when the head noun to which the genitive marker attaches contains the plural -s (Zwicky 1987; Lapointe 1996). As (1a) and (1b) show, choosing -s is ungrammatical with a pluralized head noun in -s, though not with other plural forms (1c). As (1c) shows, this is the result of a dissimilatory pressure between two *affixes* with identical segmental content: when both are [-s] (or more likely, when both are [-z]), one of them – in this case, the outermost – has to go unexpressed, but when it is part of the stem, no such constraint holds, as shown in (1d).

(1) *English possessive clitic's zero allomorphy*

a. the cats' feet are dirty ([kæts], *[kætsɪz])
b. the pigs' hooves are clean ([pɪgz], *[pɪgzɪz])
c. the oxen's hooves are dirty
d. Katz's deli (*[kæts], [kætsɪz])

Interestingly, the dissimilatory zero allomorphy of the Saxon genitive is not enforced when the pluralized noun is not the head of the entire noun phrase – in other words, when the element undergoing possessive marking is not identical to the phonological edge at which the clitic -s is placed. In such cases, some speakers (including myself) variably allow fully-fledged -s-marking even on a noun which bears the plural -s:

(2) *Zero allomorphy optional when plural is not on head noun*

a. the lady with the cats' name is Tinuviel ([kæts], [kætsɪz])
b. the man in front of the pigs's son won the competition ([pɪgz], [pɪgzɪz])

The statement of these facts requires reference to the head *vs.* non-head status of the noun phrase bearing the external marking of genitive case, and as such presents interesting challenges for fully monostratal theories of morphology–phonology interaction. Returning however to our primary concerns in (1), the choice of allomorphs and indeed the phenomenon as a whole can clearly be understood in terms of dissimilatory pressures (see CHAPTER 60: DISSIMILATION) against adjacent identity, a set of pressures we will group here under the label of the Obligatory Contour Principle (Goldsmith 1976; Yip 1988; Suzuki 1998), relativized to adjacent sibilants.

A second case of dissimilatory pressures driving allomorphy is found in the interaction between masculine gender allomorphy and plural marking in Catalan (Bonet *et al.* 2007). Like the English case discussed immediately above, it involves avoidance of adjacent sibilants. However, unlike the English case, in which -Ø is the exceptionally chosen allomorph recruited in case of potential adjacent identity, -Ø is the default allomorph in Catalan, and avoided through choice of [-u] in order to prevent adjacent identical sequences of [s], even though the regular epenthetic vowel in Catalan is [ə].

In Catalan, masculine gender on nouns can be expressed by three separate allomorphs: -Ø, the most widespread, and, according to Bonet *et al.* (2007), unmarked allomorph, is used for nouns such as (3a), which have no overt expression of masculine gender. The second-choice allomorph, [-u], is found with nouns such as (3b) and (3c), in both the singular and the plural, as a result of lexical specification to select this allomorph. Interestingly, however, there are nouns such as (3d) which clearly select the zero allomorph, as shown in the singular, but recruit the second-choice allomorph [-u] specifically in the plural, in order to avoid an otherwise adjacent-identical sequence of stridents caused by the root-final /s/ of the noun and the [s] of the plural.

(3) *Catalan theme vowel allomorphy* (Bonet *et al.* 2007)

a.	gɔt	'glass'	gɔt-s	'glasses'
b.	awt-u	'car'	awt-u-s	'cars'
c.	mos-u	'lad'	mos-u-s	'lads'
d.	gos	'dog'	gos-u-s	'dogs'

The Catalan case thus represents another instance of allomorph selection driven by the phonological pressure of dissimilation. Interestingly, as mentioned above, the choice of allomorphy is one in which an otherwise default zero allomorph is skipped over in favor of an overt allomorph, exactly the opposite of the English possessive case above. One crucial difference between the two is that the English possessive marker realized as zero occurs as the outermost suffix: avoiding identical stridents could only be resolved by jettisoning one of them. By contrast, in the Catalan case, choice of an overt masculine gender marker before the plural marker is attached involves a "prophylactic" allomorph choice, ensuring a buffer between two non-negotiably unchangeable stridents.

A third case of sibilant-OCP-driven allomorph choice, quite similar to those discussed above, is found in Hungarian (Carstairs 1988), in which the 2nd singular indefinite present indicative is normally *-as*, but is *-ol* after sibilants (and affricates whose right edge is a sibilant). In short, a number of typologically unrelated segmental dissimilatory pressures dictating allomorph choice can be found involving sibilant consonants.

While dissimilation is in general more common among consonants than among vowels (Nespor *et al.* 2003), within vowels the most common type of dissimilation is that between low vowels (Suzuki 1998). This type of pressure can be seen at work in choice of the Spanish feminine definite article, which is ordinarily *la* (4a), but which recruits the masculine allomorph *el* (4b) in the case of feminine nouns beginning with stressed *á* (Harris 1987), such as (4c). This allomorph selection to avoid the sequence *a á* is only in case of identical vowels (cf. (4d)) and in fact is only in case of stressed *á*, as can be seen in the diminutive form in (4e). In all of the examples below, stress falls on the penultimate vowel in the noun.

(4) *Spanish article allomorphy*

a.	*la mesa*	'the table (FEM)'
b.	*el libro*	'the book (MASC)'
c.	*el agua*	'the water (FEM)'
d.	*la isla*	'the island (FEM)'
e.	*la aguita*	'the water (FEM DIM)'

A similar kind of vowel dissimilation occurs with the Dutch agentive suffix, which has the allomorphs [-aar] and [-ər]. According to Smith (1976) and van Oostendorp (2009), [-ər] is the default, and [-aːr] is chosen when it follows a syllable that contains [ə], to avoid adjacent instances of schwa:

(5) *Dutch agentive suffix allomorphy*

a.	*dans*[ə]*r*	'dancer'	b.	*wand*[ə]*laar*	'walker'
	schrijv[ə]*r*	'writer'		*bewond*[ə]*raar*	'admirer'
	voorzitt[ə]*r*	'chairperson'		*tek*[ə]*naar*	'illustrator'

What is unusual about the [-aːr] allomorph is that it does not attract stress, as most other superheavy suffixes do. Van Oostendorp (2009) suggests that [-ər] is inserted first, before stress is assigned, and that after stress is assigned, [-aːr] is re-inserted as a dissimilatory repair when there are two adjacent schwas.

A third kind of dissimilatory pressure in allomorph selection is avoidance of wholesale identity between two homophonous but distinct morphemes. For example, the English demonstrative *that* and the complementizer *that* are distinct morphemes, but have identical phonologies. However, the complementizer *that* independently has a zero allomorph. Walter and Jaeger (2008), employing corpus studies, show that the incidence of complementizer *that* as opposed to its zero allomorph is much lower than otherwise expected when the distal demonstrative follows, e.g. *Among those two, I think that* <u>*that*</u> *one is better*, where the zero allomorph of the complementizer is preferred.

1.2 Segmental phonotactics

While dissimilation is one kind of segmental interaction whereby segmental similar allomorphs "repel" each other, there are kinds of segmental phenomena that involve avoiding incompatible sequences of consonant + vowel or incompatible combinations of subsegmental features. These forces, too, can drive allomorph selection, leading to avoidance of particular allomorphs when they would incur violations of combinatorial phonotactics.

In Romanian, some *k*-final nouns have an alveopalatal-final stem allomorph in the plural, while others do not (Steriade 2008), e.g. [kolak, kolaʧʲi] 'bagel (SG/PL)' *vs.* [fok, fok-uri] 'fire (SG/PL)'. Steriade (2008) shows that verbal formations based on the same root avoid the verbalizing suffix *-i* if they do not have an existing affricate-final stem allomorph available (e.g. one from the plural). Thus the denominal verb based on /fok-/ must use the suffix *-a* to avoid the consonant + vowel sequence [ki]:

(6) *Romanian availability of stem allomorphs determines denominal suffix*

singular		*plural*		*denominal verb*	
kolak	'bagel'	kolaʧ-i	'bagels'	iŋ-kolaʧʲ-i	'to roll up'
fok	'fire'	fok-uri	'fires'	in-fok-a	'to fire up'

This case is interesting because it shows that allomorph selection of an *affix* is dependent on the existence of availability of appropriate allomorphs of the *stem*, a notion that Steriade calls lexical conservatism – in other words, allomorphy is opportunistic, but it is conservative in that it depends on recruiting existing allomorphs. That is, the default verbalizing suffix (according to Steriade 2008) *-i* can only be chosen if a [k]-final stem has an independently available [ʧʲ]-final allomorph, so as to avoid the phonotactically illicit configuration of velar stop before front vowel. Otherwise, if no [ʧ]-final allomorph is available, the verbalizing suffix *-a* will be chosen.

Turning to subsegmental phonotactic restrictions, in Udihe (Bye 2008), the perfective is expressed on stems ending with [−high] vowels by laryngealizing the vowel – in this case, the allomorph is thus smaller than a full segment. However, since the language does not allow laryngealized high vowels, when the stem ends in a high vowel, there is a suffix *-ge* concatenated, instead of laryngealization:

(7) *Udihe perfective allomorphy*

a.	etetḛ	'work-PERF'	b.	dogdi-ge	'hear-PERF'
	zawa̰	'grab-PERF'		bu-ge	'give-PERF'
	olokto̰	'cook-PERF'			

In this case, therefore, a feature co-occurrence phonotactic (banning [+high] together with [+constricted glottis]) drives allomorph selection: the ordinary exponence process is overridden by a phonotactic one.

1.3 Syllable structure

Arguably some of the most widespread instances of phonologically conditioned allomorphy arise in the domain of syllable structure (see also CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE). When there are two or more allomorphs, the choice among them often is based on yielding a syllable structure that either avoids codas, avoids hiatus, or avoids complex codas without a sufficient sonority drop.

To illustrate this phenomenon it might be interesting to take a familiar case, the English past tense, in a slightly unfamiliar context: its use by second-language learners of English, in this case those whose native language is Brazilian Portuguese (Baptista and da Silva 2006). According to these authors (and to my own observations), speakers of Brazilian Portuguese English (henceforth BPE) choose between the allomorphs *-d* and *-ed* not based on whether the verbal stem ends in a coronal consonant, but on whether it ends in an obstruent:

(8) *BPE past tense allomorphy for obstruent-final verbs*

a.	*added*	[ædəd]	
b.	*packed*	[pækəd]	
c.	*act*	[ækt]	*[ækət]
d.	*leaned*	[ɹind]	*[ɹinəd]

As (8b) shows, in BPE the vowel-initial allomorph of the past tense suffix is extended beyond its original context and is chosen following an obstruent, in order to avoid the sonority plateau of two adjacent stops in a syllable coda. As (8c) shows, this is not a general process of epenthesis, as it occurs only in heteromorphemic contexts, and hence really is about choosing the allomorph *-ed* and not about general purpose vowel insertion. Finally, as (8d) shows, this epenthesis is sensitive to the nature of the stem-final consonant, and does not apply after sonorants. L2 phonology of this sort can be interesting, as it illustrates a case in which the distribution of allomorphs may be "opportunistically" generalized beyond their original (or historical) contexts in order to improve syllable structure.

The English indefinite article *a/an* is a case of hiatus avoidance (CHAPTER 61: HIATUS RESOLUTION), in which choosing *an* in vowel-initial contexts allows one to avoid a sequence of vowels. Stem allomorphs may also be chosen in order to avoid hiatus; Rubach and Booij (2001) cite the example of *Plato* and *hero*, which have an identical syllabic template, but which differ in their adjectival forms *Platonic* and *heroic*. In the case of *Platonic*, according to Rubach and Booij (2001), there is already an existing stem allomorph *Platon-* that can be recruited in order to avoid hiatus; in the case of *heroic*, there is not. One way of putting things is that allomorph selection is resourceful but not omnipotent: avoiding hiatus is nice when possible, and existing allomorphs can be recruited for that purpose, but wholly new stem allomorphs cannot spring into existence for this sole purpose.

Cases of phonologically conditioned allomorph selection (as opposed to general purpose epenthesis, cf. (8c), or the implausibility of lexically specific *-n-* epenthesis in *Platonic*) are clearest when the two allomorphs are quite distant from one another. In Korean, the nominative case ending is chosen according to whether the stem ends in a vowel or consonant, as shown in (9).

(9) *Korean nominative case suffix chosen based on final segment of stem* (Suh 2008)

mom-ɨ	'body-NOM'
k^ho-ka	'nose-NOM'

The vowel-initial allomorph allows for resyllabification, thus removing a coda from the representation. Both allomorphic choices result in a CV.CV profile for the suffixed stems.

The choice of phonologically unrelated allomorphs can sometimes be the result of historical divergence (CHAPTER 93: SOUND CHANGE). For example, Mascaró (2007) discusses the fact that the Moroccan Arabic object clitic used to be /-hu/, but now has developed into two distinct (and synchronically unrelatable) allomorphs, [-h], chosen after vowel-final stems (e.g. [xt^{ʕ}a-h] 'his error'), and [-u], chosen after consonant-final stems (e.g. [ktab-u] 'his book'). Each of these allomorphs is clearly chosen in order to avoid marked syllable structure: choosing the opposite contexts for these allomorphs would result in either hiatus or in a complex coda.

Somewhat similar to the BPE past tense case discussed above is the pattern of allomorphy with the Swedish definite suffix (Löfstedt 2008), in which an allomorph originally intended for one set of environments is recruited for another. The two allomorphs are [-n] and [-en]:

(10) *Swedish definite suffix allomorphy* (Löfstedt 2008)

	stem	*definite*	
a.	byː	byːn	'village'
b.	sykːel	sykːeln	'bicycle'
c.	gruːp	gruːpen	'hole'
d.	pilgrim	pilgrimen	'pilgrim'
e.	ɛlsklɪŋː	ɛlsklɪŋːen	'love (DIM)'
f.	hymːn		'hymn'

As (10a) and (10b) show, [-n] is the allomorph ordinarily chosen after sonorants, while *-en* is chosen after obstruents (10c). However, [-en] is also chosen after nasals (10d) and (10e), despite the fact that nasal + nasal sequences are tolerated tautomorphemically in Swedish (10f). This represents an emergent dispreference for sonority plateaus in codas that is blocked when possible by recruiting another allomorph (see also CHAPTER 49: SONORITY).

In French, masculine and feminine adjectives have two allomorphs, e.g. [pəti]/[pətit] 'small (MASC/FEM)'. However, the latter may also be used when preceding *masculine* adjectives if they are vowel-initial. This is a case in which hiatus resolution (or perhaps the preference for a syllable onset in the noun) recruits an adjectival allomorph of the "wrong" gender:

(11) *French adjectival allomorphy* (Tranel 1997; Perlmutter 1998; Steriade 1999)

petit canard	[pətikanar]	'small duck (MASC)'
petite bête	[pətitbɛt]	'small beast (FEM)'
petit animal	[pətitanimal]	'small animal (MASC)'

The consonant present in the feminine form is recruited wholesale for masculines simply for syllable-structure considerations. Similar patterns can be found with demonstratives (e.g. *ce* [sə], *cette* [sɛt]). While some authors treat patterns such as (11) in terms of phonological derivation (where /pətit/ is in the underlying form of both masculine and feminine, with a rule of final consonant deletion operative in the masculine), cases of wholesale suppletion, such as (12) below, clearly call out for an analysis in terms of allomorph selection, arguably driven by the same requirements of onset-furnishing as in (11) above:

(12) *French adjectival allomorphy with wholesale suppletion*

beau canard	[bokanar]	'beautiful duck (MASC)'
belle bête	[bɛlbɛt]	'beautiful beast (FEM)'
belle animal	[bɛlanimal]	'beautiful animal (MASC)'

In sum, a wide range of allomorph selection process are driven by bread-and-butter syllable structure well-formedness constraints, such as preference for onsets, dispreference for codas, and dispreference for sonority drops in codas that are not steep enough. In the next section, we will turn to a competing force in allomorph selection, where considerations of alignment of morphological and syllabic constituents may in fact create marked syllable structure.

1.4 Morphological alignment

Since at least Dressler (1977) it has been realized that there is a tendency in natural language to align morphological constituents with syllabic constituents (CHAPTER 103: PHONOLOGICAL SENSITIVITY TO MORPHOLOGICAL STRUCTURE). In fact, the tendency for C-initial allomorphs to occur after vowel-final stems and V-initial allomorphs after consonant-final stems that we have seen in the preceding section is occasionally trumped by precisely such factors, resulting in the opposite pattern. The best-known case of this is the Haitian definite article (Klein 2003), in which a C-initial suffix occurs after C-final stems and a V-initial suffix after V-final stems:

(13) *Haitian definite article allomorphy* (Klein 2003)

liv-la	'book-the'
papa-a	'father-the'

According to Klein, the default allomorph is *-a*, and *liv-la* is preferred to *li.va* because the latter would resyllabify the stem. Apparently, the morphological alignment of the right edge of the stem with the right edge of a syllable is important to maintain.

The Galician definite article (Kikuchi 2006) shows that morphological/syllabic alignment may enter into opaque interactions. The allomorphs of the definite article are *o/lo* (MASC SG), *a/la* (FEM SG), *os/los* (MASC PL), and *as/las* (FEM PL). The onsetless forms, illustrated in (14a), are the default, but following a continuant *r*, *l*, *s* in the coda of a preceding word (in combinations that are adequately close, such as infinitives or prepositions plus their complements), the liquid-initial allomorph is chosen, as shown in (14b):

(14) *Galician article allomorphy* (exemplified for feminine *a(s)* vs. *la(s)*)

a.	*a xente*	'the people'	
	as mulleres	'the women'	
	para as mulleres	'for the women'	
	sobre a xente	'about the people'	
b.	*ve-la xente*	'I saw the people'	/ver/
	'toda-las mulleres	'all the women'	/'todas/

Kikuchi (2006) proposes that, like the Haitian Creole case discussed above, the choice of an onsetless article in Galician is motivated by a preference for morphological word edges to be aligned with syllable edges. When the preceding word ends in a consonant, *ver.la* is preferred to *ve.ra*, because the latter moves the definite article *a* inside, rather than at the edge of, a syllable.

Notice that choice of the liquid-initial onset is conditioned by a preceding segment that is deleted on the surface; in other words, the allomorphic choice shows opacity (since otherwise surface ['toda] (underlying /'todas/) should pattern like *para*). The deletion at this intermediate level involves an OCP effect: *ver-la* or *todas-las* would create adjacent identical continuant consonants, which Kikuchi (2006) posits are banned across word boundaries in prosodically close domains, such as prepositions or verbs plus their complements. Hence, once the allomorph *la* is chosen, the preceding continuant deletes.

A similar case is the Korean conjunctive suffix realized as [-wa] or [-kwa] (Suh 2008). The interest of this allomorphy is the fact that within the same language, the nominative (discussed in (9)) and the conjunctive have apparently contradictory distributions.

(15) *Korean conjunctive suffix allomorphy* (Suh 2008)

kho-wa	'nose-CONJ'
mom-kwa	'body-CONJ'

According to Suh (2008), this pattern results from the need for right-alignment of the stem with the right edge of a syllable. As *mw-* is tolerated as an onset, onset maximization with *mo-mwa* would result in resyllabification of the stem *mom*, while *mom.kwa* does not. What about *mo.mi*, then, the nominative form seen above? Suh's (2008) actual constraint is about alignment with the right edge of any subsyllabic constituent. In *mo.mi* the stem-final [m] is aligned with the right edge of the onset (which is simplex), but in **mo.mwa* it is not, being blocked from the right edge of the onset by the following glide. By contrast, in *mom.kwa*, the stem-final [m] is comfortably at the right edge of the coda constituent.

Mòcheno is a German dialect spoken in the Trentino region of Italy, with two allomorphs of past participial prefix, according to Alber (forthcoming): [ga-], and a subsegmental [−voice, −cont] feature matrix. The distribution of these allomorphs depends on the initial consonant of the verb stem. With voiceless stops and supralaryngeal fricatives, the [−voice, −cont] allomorph is chosen (16a), which may result in Ø realization or affricate formation. For voiced stops and all sonorants, [ga-] is chosen (16b):

(16) *Mòcheno participial allomorphy* (Alber, forthcoming)

	stem		*participial form*
a.	tondərn	'to thunder'	tondərt
	krɐtsn	'to scratch'	krɐtst
	viern	'to conduct'	pfiert
	ʃlo:ŋ	'to beat'	ʧlo:ŋ
b.	o:tnən	'to breathe'	gao:tnt
	rɛarn	'to cry'	garɛart
	mɐxen	'to make'	gamɐxt
	bisn	'to know'	gabist
	griezən	'to greet'	gagriest

According to Alber, the principle that a stem should not be resyllabified with a prefix (i.e. that the verb stem should be aligned with its own syllable) is what drives the allomorph selection. Voiceless stops and fricatives can form a doubly linked representation with the [−voice, −cont] segment, leaving the stem at the left edge of its own syllable. The rest choose [ga-] so as not to resyllabify the stem with a preceding consonant.

1.5 *Stressedness and vowel quality*

Stressed vowels like to be more sonorous, and unstressed positions tolerate fewer vowel contrasts (Crosswhite 1998; CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE). A number of allomorph selection processes seem to be motivated by the distribution of non-peripheral vowels (i.e. mid vowels, in the cases below) in unstressed positions.

Due to the Stress-to-Weight principle, stressed syllables like to be heavy, and hence may undergo processes such as Iambic Lengthening (Hayes 1995; CHAPTER 44: THE IAMBIC–TROCHAIC LAW) and glottal stop insertion. Conversely, unstressed syllables may undergo glottal stop deletion, in order to reduce their relative prominence with respect to stressed syllables.

Somewhat like the case of the Spanish definite article allomorphy, which recruits "the wrong gender" in order to avoid hiatus of identical low vowels, an allomorphy process in Hebrew plural formation occurs that borrows the wrong gender for phonological reasons. The Hebrew feminine suffix *-ot* (17b) is exceptionally extended to masculine nouns, such as (17c) (Becker 2009), which otherwise take the masculine suffix *-im* (17a).

(17) *Hebrew plural allomorphy* (Becker 2009)

a.	'jelad	jelad-'im	'boy (MASC SG/'MASC PL)'
b.	xat'ser	xatser-'ot	'backyard (FEM SG/FEM PL)'
c.	xa'lon	xalon-'ot	'window (MASC SG/MASC PL)'

One defining feature of Hebrew plural formation is that it takes stress away from the nominal stem, and this has consequences for the marking of masculine nouns whose final vowel is [o]. Becker (2009) argues that the mid-round vowel [o] is dispreferred in unstressed syllables in Hebrew (and more generally in weak positions (Beckman 1997), but can receive support by a kind of vowel harmony that licenses it: a weak unstressed [o] can be tolerated when adjacent to a stressed [o]. For this reason, the feminine *-ot* is recruited specifically for masculine nouns of which the last syllable contains [o]. Becker (2009) shows that this allomorphic recruitment is not a historical quirk, but is actively used in novel plural formations in a wug test.

Surmiran Rumantsch verbs have two sets of stem allomorphs, chosen on the basis of stress placement. The unstressed variant is not predictable from the stressed variant, ruling this out as a case of straightforward vowel reduction. According to Anderson (2008), the alternations are not predictable, and require listing of both stem allomorphs. It is a common pattern in Romance that the infinitive, 1st plural, and 2nd plural will bear distinct allomorphs from the rest of the paradigm, as in the Italian verb conjugation for *andare* 'to go', which is wholly suppletive.

(18) *Italian present tense conjugation for* andare, *with stress marked and agreement endings separated*

	singular	*plural*
a.	*'vad-o*	*and-i'amo*
b.	*'v-ai*	*an'd-ate*
c.	*'v-a*	*'v-anno*

As in Italian, which has the longer endings for 1st plural and 2nd plural, Surmiran shows a pattern which is understandable in terms of preference for unstressed syllables to be low-sonority [i u ə] rather than [e a]. The choice of allomorphs thus aligns prosodic weakness with segmental weakness: the low-sonority stem allomorphs [ləv-] and [fit-] are chosen when stress is on the inflectional ending rather than on the root.

(19) *Surmiran stem allomorphy* (Anderson 2008)

inf	lə'var	'get up'	fit'tar	'finish'
1sg	'lev		'fet	
2sg	'levas		'fettas	
3sg	'leva		'fetta	
1pl	lə'vagn		fit'tagn	
2pl	lə'vez		fit'tez	
3pl	le'van		'fettan	

Anderson argues that these alternations do not involve unstressed vowel reduction, as the same stressed [e] corresponds to both [i] and [ə]. As no deterministic rule of reduction is at work here, this must be handled in terms of allomorph selection, in this case based on stress-to-sonority.

In a somewhat similar vein, Spanish mid vowels display an alternation whereby they diphthongize under stress. However, while certain instances of diphthongization in unstressed syllables are allowed, others are not.

(20) *Spanish mid-vowel diphthongization stem allomorphy* (Bermúdez-Otero 2009)

a. *encontrár* 'to meet'
b. *encúentro* 'a meeting'
c. *encuentrón* 'meeting (AUG)' (word-level: 'one hell of a meeting')
d. *encontrón* 'someone who bumps into others' (stem-level: deverbal)

According to Bermúdez-Otero (2009), whether or not a mid-voweled root will diphthongize or not under stress is unpredictable, and hence two stem allomorphs must be listed, e.g. /eŋkwent/ and /eŋkont/. The difference between denominal derivation in (20c) and deverbal derivation (20d) is that, while denominal derivation (e.g. diminutives, augmentatives) is word-level (within the Lexical Phonology/Stratal OT conception of levels; see also CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME), the deverbal derivation is stem-level. Given such a division, Bermúdez-Otero (2007) argues that there is a relevant stem-level constraint in Spanish that disallows unstressed dipththongs. As a result, given the deverbal derivation which places stress on the suffix, the stem allomorph /eŋkont/ will be chosen, again due to the phonological preference for stress and syllabic prominence to line up.

1.6 Foot structure

In this section we examine cases of allomorph selection which involve improving or maintaining the preferred foot structure in the language (CHAPTER 40: THE FOOT).[2] The cases are similar in spirit to the alignment of segmental prominence to stress, but often involve syllable counting in a way that requires footing in order to determine allomorph selection.

González (2005) analyzes allomorph selection in languages of the Panoan family as choosing segmental allomorphs based on foot structure. For example, the Shipibio repetitive suffix has the allomorphs [ɾiba] and [ɾibi], with the former appearing after odd numbers of syllables and the latter appearing after even numbers of syllables. Since this is a trochaic language, the effect of this allomorph choice places the vowel [a] in strong syllables and [i] in weak syllables.

(21) *Shipibio syllable-counting allomorphy:* [ɾibi] *vs.* [ɾiba] (González 2005)

a. pima 'eat-CAUS' + REPETITIVE + PAST
(pi.ma)(ɾi.bi)ki 'He made him eat it again'.
b. pi 'eat' + REPETITIVE + PAST
(pi.ɾi)(ba.ki) 'He ate it again'.

In Yaminahua, the allomorphs [tiʃo] and [toʃi], a directional perfect suffix meaning 'on arriving', are chosen on the same basis: the appearance of syllable counting, which is actually based on foot structure. Higher sonority [o] is thus lined up with the stronger half of the foot.

[2] All of the cases discussed here involve trochaic languages, which may be an accident, or may be part of a deeper generalization.

(22) *Yaminahua syllable-counting allomorphy:* [tiʃo] *vs.* [toʃi] (González (2005)

a. fiʧi 'find' + ARRIVE + PAST + PLURAL
(fi.ʧi)(to.ʃi)(afo) 'found on arriving'

b. fa 'say' + ARRIVE + yesterday
(fa.ti)(ʃo.i)(ta) 'said on arriving'

Somewhat like the cases of peak prominence discussed in §1.5 above, these cases illustrate allomorph selection based on syllable counting that makes sense once integrated with the fact that this is a trochaic language, and hence odd-numbered stems will want the more sonorous syllable to come second, while even-numbered stems will want it to come first.

Similar to the Panoan cases above is the selection of allomorphy in Estonian, where the genitive plural has two allomorphs, *-te* and *-tte*, and the partitive plural has two allomorphs, *-sit* and *-it*. When these are added to vowel-final bases, the choice of allomorph is determined by what looks like syllable count: *-it* and *-tte* are chosen with odd-numbered bases. Estonian is trochaic, and Kager (1996) argues that it is foot structure that determines the choice of allomorphy: the head of a foot should be heavy if possible (CHAPTER 57: QUANTITY-SENSITIVITY):

(23) *Estonian odd-numbered stem allomorphy:* [-tte] *and* [-it] (Kager 1996)

paraja	'suitable'	('pa.ra)(ˌjat.te)	(GEN PL)
		('pa.ra)(ˌjait)	(PART PL)

By contrast, when the stem is even-numbered, no suffixation can possibly improve the weight of the head of the foot, and so affixes which cause no resyllabification or realignment are chosen:

(24) *Estonian bisyllabic stem allomorphy:* [-te] *and* [-sit] (Kager 1996)

'visa	'tough'	'(vi.sa)te	(GEN PL)
		'(vi.sa)sit	(PART PL)

Kager (1996) argues that it is feet and not syllables that are counted, as bisyllabic bases with superheavy initial syllables group with [paraja], e.g. [aasːta-tte] 'year-GEN PL'. Hence this allomorph selection cannot be reduced to syllable counting, but must be stated in terms of heads of feet and making them heavier.

Greek is a language with mostly antepenultimate stress in trisyllabic words (Drachman *et al.* 1996). Its syllable-counting allomorphy with action nominals involves [-ma] and [-imo]. Monosyllabic stems take [-imo], thereby achieving antepenultimate stress unproblematically, whereas polysyllabic stems take [-ma], keeping the stress on the initial syllable while changing word length minimally:

(25) *Greek action nominals and syllable count* (Drachman *et al.* 1996)

'vreks-imo	'wetting'
'skupiz-ma	'sweeping'

The case above shows a kind of complementarity effect: the shorter stems take the longer allomorph and the longer stems take a shorter allomorph. A similar case can be found in Dutch, whose feet are preferably disyllabic trochees. The plural suffixes *-en* and -s are chosen in order to form trochees at the right edge.

(26) *Dutch plural allomorphy: based on trochees at right edge* (Booij 1998)

knie	*'knieën*	'knee'
bal	*'ballen*	'ball'
'natie	*'naties*	'nation'
ge'nie	*ge'nieën*	'genius'
'kanon	*'kanons*	'canon'
ka'non	*ka'nonnen*	'cannon'

Notice above that a complementarity effect obtains with monosyllables *vs.* disyllables, and also in terms of distance from the right edge: the longer *-en* is chosen when there is a shorter distance of the stress from the right edge, and the shorter -s when there is a longer distance. A similar case can be found in Spanish with the suffix that creates abstract nouns from adjectives and has two allomorphs:

(27) *Spanish* [-ez]/[-eza] *allomorphy* (Aranovich and Orgun 2006)

a.	*vil-eza*	'vileness'	b.	*rigid-ez*	'rigidity'
	nobl-eza	'nobility'		*madur-ez*	'maturity'
	trist-eza	'sadness'		*tirant-ez*	'tenseness'

According to Aranovich and Orgun (2006), there appears to be a requirement that derived nouns be larger than a foot, but not larger than necessary. This pattern also shows the complementarity effect discussed above. However, consonant-final stems appear to contradict this pattern, e.g. *gentil-eza* 'gentleness', *sutil-eza* 'subtlety'. This apparent misbehavior of allomorph selection can be resolved, however, given that Spanish vowel sequences delete heteromorphemically. With such an analysis available, the stems above are thus in fact *vilo-eza, triste-eza vs. rigido-ez, maduro-ez*, at the point before which vowel deletion applies. The more abstract underlying representation, with vowel-final stems as the conditioning environment, enables a consistent statement of allomorph selection in terms of foot structure. Aranovich and Orgun (2006) argue that allomorph selection takes place at a level of representation before vowel deletion, and that the choice is motivated by a goal of forming two perfectly binary feet at that level of representation, e.g. (vi.lo)(e.za), (ri.gi)(do.ez). This pattern thus shows allomorph selection conditioned by foot structure, but at a level removed from the surface, a point to which we will return.

2 Arbitrary cases that still reference phonology

Dealing with phonologically conditioned suppletive allomorphy, Paster (2005, 2006) proposes that certain cases of allomorphy are sensitive to phonology but do not optimize anything, and proposes instead a mechanism of subcategorization in the grammar. Based on a number of cases such as Axininca and Italian, Bye (2008) echoes this conclusion, i.e. that allomorphy is selection, and that phonological optimization is due to historical or coincidental factors. Embick (2009) endorses a similar viewpoint to these authors, eschewing a model of grammar with "global" interaction between morphology and phonology. While we will discuss their models further below, it is first worth showing a few cases of these that indeed reference phonology but in which the actual choice seems to have no phonotactic/markedness-based motivation.

Two such cases involve syllable- or mora-counting allomorphy of the type that was relevant for optimization of foot structure above, but in ways which do not seem to improve anything, and cannot be easily understood in terms of existing grammatical well-formedness constraints unless one wants to "pollute" the constraint set with a host of parochial constraints unrelated to the core intuitions of markedness. They are shown below, for Kaititj and Axininca Campa, both of which reference the phonological factor of syllable or mora count in order to determine selection of allomorphs, but choose in a way that still leaves the connection between the structural description (syllable count) and structural change (insertion of allomorph) arbitrary.

(28) *Kaititj ergative suffix allomorphy:* [-ŋ] *after bisyllabic stems,* [-l] *after trisyllabic stems* (Paster 2006)

a.	aki-ŋ	'head-ERG'	b.	aliki-l	'dog-ERG'
	iltʲi-ŋ	'hand-ERG'		aṭuji-l	'man-ERG'
	ajnpni-ŋ	'pouch-ERG'		aɣiki-l	'sun-ERG'

(29) *Axininca Campa genitive allomorphy:* [-ni] *after bimoraic stems,* [-ti] *elsewhere* (Bye 2008)

a.	no-jorʲa-ni	'my manioc worm'	b.	i-wisiro-ti	'his small toucan'
	i-çaa-ni	'his anteater'		no-jairo-ti	'my termite'
	a-sari-ni	'our macaw'		a-jaarato-ti	'our black bee'

Another case that is not obviously amenable to an optimization analysis is the plural definite article selection in Italian, which chooses [ʎi] before consonant-initial stems, but [i] before vowel-initial stems.[3] There is no obvious advantage to having [ʎi] as opposed to [i] before a vowel-initial noun.

(30) a.	*i vini*	'the wines'	b.	[ʎ]*i alberghi*	'the hotels'
	i padri	'the fathers'		[ʎ]*i inglesi*	'the Englishmen'
	i ragazzi	'the boys'		[ʎ]*i uccelli*	'the birds'

While one might imagine certain possibilities (e.g. one violation of ONSET is tolerated with vowel-initial stems, but having a vowel-initial article before it tips the scales with two violations of ONSET, perhaps due to constraint self-conjunction), these are far from obvious. Paster (2006) discusses a similar case from Jivaro, in which consonant-final stems take the suffix *-cha*, while vowel-final stems take the suffix *-cho*.

These cases clearly reference phonology, but cannot be the result of automatic grammatical choices based on well-formedness. Instead, they require a mechanism

[3] The allomorph [ʎi] is also chosen before geminate-initial and sC-cluster-initial stems. Under certain analyses (e.g. Kaye 1992), these can be reduced to the vowel-initial context. Whether or not this reduction is made does not change the point in the text about the seeming arbitrariness of the distribution of [i] *vs.* [ʎi].

called "subcategorization," which states in the lexical entry of Kaititj ergative allomorphs that /-ŋ/ is subcategorized for, or is chosen in the case of, disyllabic stems, and /-l/ in other cases:

(31) *Sample subcategorization for Kaititj ergative*

ERG ↔ /-ŋ/ in the context σσ __
ERG ↔ /-l/

Given a schema like the one above and the Elsewhere principle (see e.g. Halle 1997) for vocabulary insertion/allomorph selection, the more specific contextual specification will always block the less specific one, if met.

Although Paster (2005, 2006) and Bye (2008) argue, on grounds of parsimony, that if the language faculty needs subcategorization anyway, we need not bother with optimization at all, other authors take a stance in which the case of phonologically optimizing allomorph selection can be determined entirely by constraint satisfaction, with brute-force subcategorization arising only as a "last resort" (e.g. Lapointe 1999) for the non-optimizing cases (CHAPTER 106: EXCEPTIONALITY). Presumably such a division of labor would lead to the prediction that, as subcategorization is more costly (requiring lexical listing, whereas constraints otherwise active in universal grammar come "for free"), systems might evolve over time into becoming phonologically optimizing.

Wolf (2008) makes the interesting point that the mechanism of subcategorization is better for stating positive conditions than negative conditions. In other words, it works well for saying "choose [-ŋ] when the stem is bisyllabic," but would not work well for cases like "choose Spanish [la] when the following noun does not start with ['a]." The logic of subcategorization conditions like (31) certainly allows reference to natural classes, even those formed by negative values of features (e.g. choose English *an* before stems beginning with [-consonantal] segment), but when such specifications involve disjunctions or miss generalizations linking the structural description to the structural change (i.e. insertion of the allomorph), they become less appealing.

3 Issues for theoretical models

Having established that a great deal of allomorph selection is phonologically conditioned, important consequences arise for models of morphology–phonology interaction, whether couched in terms of constraint satisfaction as the means for choosing among allomorphs (e.g. Kager 1996; Mascaró 2007)) or subcategorization/vocabulary insertion (e.g. Halle 1997; Embick 2009). While I will not choose among these models here, I will identify two important issues for any class of models: the question of when (i.e. at what stage in a phonological computation) allomorph selection takes place, and what the mechanism of choice is.

3.1 When does allomorph selection take place?

Stratal (e.g. Bermúdez-Otero 2007), derivational (e.g. Wolf 2008), and cyclic (e.g. Embick 2009) models of grammatical computation differ from monostratal models in that the first three potentially allow multiple levels of intermediate representation, with processes like allomorph selection occurring according to the well-formedness principles of one level, but subsequently obscured by the operations of a later level. However, monostratal (or globalist) models, which allow information from various modules to be present either simultaneously or preserved across different stages of computation, have a distinguishing trait: they allow, for example, phonological processes such as allomorph selection to freely refer to syntactic boundaries and constituency. By contrast, stratal and/or cyclic models may limit such information from entering later stages of phonological computation. In the following two subsections we turn to these two issues.

3.1.1 Opacity: Allomorph selection at intermediate levels

As mentioned above in three case studies, not all allomorph selection occurs at the surface. We review the relevant facts, four of which suggest that allomorph selection occurs at an earlier level of representation (therefore requiring the interaction with tools for opacity such as rule ordering, stratal Optimality Theory or OT-CC); and one of which suggests the possibility of lexical reinsertion after certain phonological processes have applied.

Reviewing the facts in the Galician definite article case, the allomorph *la* was chosen over *a* to avoid syllabification of the definite article with a preceding onset from a different morpheme, e.g. *ver.la* 'see the' instead of *ve.ra*. However, the verb-final *r* was subsequently deleted due to an OCP effect, yielding *ve.la* (cf. *para a* 'for the', to show hiatus is clearly not at stake). Thus the phonological conditioning of allomorph selection is clearly taking place before *r*-deletion.

Similarly, in the Spanish deadjectival suffix case, the allomorph [-eza] is chosen for adjectival stems that are disyllabic as the input to this affixation, and thus *vilo, triste, gentil* all pattern the same in taking [-eza] in order to form two feet. However, the former two undergo final vowel deletion. Thus, the statement of syllable count in terms of allomorph selection is clearly taking place before vowel deletion.

Finally, in a somewhat different vein, the Dutch agentive suffix discussed above according to van Oostendorp (2009) requires selection of one allomorph, [-ɔr], before stress is assigned, as the default. However, after stress assignment has already happened, if the suffix immediately follows another syllable containing schwa, van Oostendorp (2009) proposes another cycle of allomorph selection in which [-aːr] can be chosen to resolve the schwa-based OCP.

There are two other cases I will mention here, described in Gibson (2008), which relate to allomorphs chosen after C-final and V-final stems, both of which occur derivationally prior to a process of consonant deletion. In Japanese, for example, the non-past suffix is [-u]/[-ru] and the inchoative suffix is [-oo]/[-joo]. However, [w]-final stems delete their final consonant when preceding high or mid vowels, and do so after allomorph selection. For comparison, the negative suffix [-anai]/[-nai], which does not trigger [w]-deletion, is shown.

(32) *Japanese opaque allomorphy with* [w]-*final stems*

	jom 'read'	ne 'sleep'	iw 'say'	jow 'get drunk'
non-past	jom-u	ne-ru	i-u	jo-u
inchoative	jom-oo	ne-joo	i-oo	jo-oo
negative	jom-anai	ne-nai	iw-anai	jow-anai

A very similar case occurs in Turkish, where a process of intervocalic *k*-deletion applies after C-final *vs.* V-final allomorph selection. The 3rd person possessive suffix [-i]/[-si] surfaces after C-final *vs.* V-final stems, but *k*-final nouns choose [-i], even though the conditioning stem-final consonant subsequently deletes:

(33) *Turkish opaque allomorphy with* k*-final stems*

bedel	bedel-i	'price (NOM/POSS)'
fire	fire-si	'attrition (NOM/POSS)'
bebek	bebe-i	'baby (NOM/POSS)'

In both of these cases, allomorph selection is applying to a level of representation prior to consonant deletion, which means that the optimization is on an intermediate form. If the consonant deletion processes can be assigned to different strata of phonological computation (e.g. post-lexical, or word-level), then these phenomena can be easily dealt with in Lexical Phonology or Stratal OT (CHAPTER 85: CYCLICITY; CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME).

The notion that intermediate representations are evaluated for well-formedness (i.e. a derivational model that includes negative well-formedness constraints) is entirely reasonable. It is now largely accepted that monostratalism is untenable, and also known that grammars without statements of negative constraints miss generalizations. Future models that develop a series of ordered computations with intermediate levels of representations should definitely incorporate treatment of allomorphy phenomena, particularly since they interface with interesting questions of lexical insertion (cf. Embick 2009; Wolf 2008).

3.1.2 *Reference to syntactic factors*

As mentioned above, an important issue that arises in the study of allomorphy and "where/when" it takes place is the issue of syntactic sensitivity. For example, Russian 3rd person pronouns have two allomorphs, one vowel-initial and the other containing an initial [n-], with the latter chosen after prepositions (presumably for the same reasons as the Galician case above: namely to avoid resyllabification and allow stems to begin their own syllable). Hence the pronoun [ix] is realized as [nix] when a preposition precedes, as prepositions closely prosodify with their complements in Russian (34a). However, when the pronoun is not the head of the prepositional complement, but rather a possessive modifier, no such allomorphy occurs (34b).

(34) *Russian prepositional complement allomorphy*

a. bez nix 'without them'
b. bez ix brata 'without their brother'

In fact this is the same kind of pattern we observed in the first example in this chapter: the English possessive -*s* does not display zero allomorphy when following a plural -*s* that is *not* the head of the NP. What is interesting about both of these cases is that the affected linear string is the same. Hence "syntactic brackets" must be visible. The simplest model to envision is one in which they are literally visible to morphophonological representations. However, one might also experiment with the possibility, in cyclic models, that the possessive modifier in (34) is submitted to its own closed-off morphophonological computation prior to the cycle in which the preposition is visible to its complement. In other words, the idea might be that in *bez nix*, the equivalent of [without 3PL]$_{PP}$, the preposition is visible at the moment the pronoun is spelled out; by contrast, in *bez ix brata*, the equivalent of [without [3PL brother]$_{DP}$]$_{PP}$, the pronoun is spelled out at an earlier cycle in which the preposition is not yet visible. Whether this hypothesis can be integrated with other facts about the order of spell-out in Russian prepositional complements and noun phrase structure awaits a fuller investigation.

A second case of reference to syntactic factors concerns the -s/Ø allomorphy found in dialects of Catalan, described by Bonet *et al.* (forthcoming). In these varieties, plural -s on adjectives is realized as zero between two consonants (a process which the authors argue is not due to a general process of strident deletion). However, this process can only apply prenominally:

(35) *Catalan prenominal/postnominal allomorphy*

aquelØ	bonØ	vins	blancs	dolços
those-PL	good-PL	wine-PL	white-PL	sweet-PL

The statement of these allomorphy conditions must thus take into account the hierarchical/linear position of the adjective with respect to the noun, which may be a few words away. Again, a simple model would be one in which the entire noun-phrase structure is still visible at the point of allomorph selection, and somehow the linear position with respect to the noun is relevant for -s/Ø allomorphy. An approach within a cyclic model might be one in which there is a sequenced computation in which postnominal elements do not have the same syntactic relation with the noun as prenominal elements – for example, postnominal elements being spelled out in a different cycle than the noun, and hence required to bear overt exponents of plurality. Again, while this account is potentially much more interesting (and restrictive) than a globalist model, it would require integration with independent facts about prenominal and postnominal asymmetries in the spell-out of adjectives.

In sum, much exciting work lies ahead particularly in the domain of syntactically sensitive phonologically conditioned allomorphy, as it raises important analytic challenges as to how the syntactic structure is made accessible – whether syntactic brackets are literally visible, or whether their effects arise as epiphenomenal due to the timing of spell-out.

3.2 *How are allomorphs chosen?*

Perhaps one of the most widely debated issues within the study of allomorph selection is the mechanism of choice. The existence of phonological conditioning and, in particular, phonological optimization, suggests that grammatical, rather than morpholexical, mechanisms are at work in accomplishing the selection – more specifically, that when a given allomorph is better for segmental, syllabic or prosodic structure, then the grammar will somehow "automatically" choose that allomorph without having to list these contexts in its lexical representation.

The choice of the allomorph that *does* lead to some kind of phonological optimization is relatively straightforward, particularly in models such as Optimality

Theory, or in fact any constraint-based models which encode well-formedness principles as an active grammatical force that can compel choice of one input over another. For example, a constraint such as *Coda will prevent choice of Korean [-ka] after consonant-final stems such as [mom], and will thus lead to choice of the phonologically optimizing allomorph, [-i].

However, in most of the cases described above there is one allomorph – call it "the optimizer" – chosen for particular phonotactic reasons, and another that is simply a default. For example, in the case of English *a/an* allomorphy, *a* is simply a default that does not optimize anything when it is chosen – it is only *an* that is recruited to improve phonotactics. The question thus arises why the optimizer *an* is not *always* chosen – it will provide an onset when needed for vowel-initial words that follow, and is otherwise seemingly harmless. The intuition to be captured is that the other allomorph, *a*, is the default, and is chosen "elsewhere," when providing an onset is not at stake one way or the other.

In the discussion that follows we will be concerned with a pattern as schematized in (36), (keeping in mind that in some cases *both* allomorphs are optimizers (e.g. Moroccan Arabic [-h] *vs.* [-u]), and neither is clearly a default).

(36) *Optimizer*
The allomorph chosen in order to satisfy a particular phonotactic, e.g. *an* to provide an onset, in a particular set of environments (e.g. before vowel-initial words).
Default
The allomorph chosen otherwise.

The pressing issue for models of morphology–phonology interaction thus becomes how to grammatically state that one allomorph is the *default* within different models of grammar – in other words, that it should be inserted *unless* some phonotactic pressure demands the other one. Interestingly enough, this question becomes of more relevance to precisely those models in which all the action is encoded in terms of negative constraints: how does one force the default allomorph to be used over the other one?

A number of distinct answers have been posited in the literature, and it is not yet clear which is the best. One class of proposals holds that the default is always phonologically more unmarked than the optimizer, which leads specifically to the question of context-free markedness. If the default can be shown to be generally unmarked (either in segmental or structural terms) compared with the optimizer, this approach will lead to selection of the default by unmarkedness criteria alone when phonotactic optimization with the stem + affix is not at stake.

Let us begin by considering cases such as the Djabugay genitive, in which [-ŋun] is chosen after consonant-final stems, while [-n] is chosen after vowel-final stems (a pattern which is presumably driven by avoidance of complex consonant clusters ending in [n]).

(37) *Djabugay genitive*

a. gu.lu.du-n 'dove'
b. ga.ɲal.ŋun 'goanna'

While the choice of [-ŋun] in (37b) can be understood in terms of avoidance of *[ga.ɲal.n], in order to enforce selection of default [-n] in (37a), Kager (1996) posits the constraint 'Gen=-*n*', which we can call a *violable exponence requirement constraint* (see Russell 1995 for elaboration of this type of constraint). These constraints demand a particular exponent for a particular morphological category and are thus one way of encoding the notion of a certain exponent being the default, but it might strike some as brute force to include and rank a violable constraint dedicated to every exponent in the grammar.

Along the lines of context-free unmarkedness arbitrating in favor of the default, Rubach and Booij (2001) propose that instead of stipulating that [-n] is the default in Djabugay, one should rather state that [-ŋun] is *dis*preferred, except in special cases. They appeal to segmental markedness (see CHAPTER 4: MARKEDNESS), and specifically claim that, in the absence of other competing factors, [-ŋun] will be dispreferred due to a markedness constraint against velar nasals. This formulation thus chooses the default over the optimizer when coda phonotactics are not at stake via context-free (but violable) *ŋ. This proposal is an interesting way of encoding the default selection, but it should always be kept in mind whether such constraints are consistent with the full-blown grammar of the language.

A second class of proposals involving context-free markedness in order to favor the default, applied to this same case, is an appeal to "shorter is better." For example, Wolf (2008) posits that constraints such as *Struc – which penalize structure in general, and thus prefer shorter outputs whenever possible – will lead to the preference for /n/ over /-ŋun/, or *a* over *an*, unless trumped by phonotactic factors. In cases involving complementarity of length effects, in which for example disyllabic stems take [-eza] to accomplish two full feet and all longer stems take [-ez], the default can clearly be seen to be the shorter allomorph. Appeals to *Struc, however, are not without problems, as discussed by Gouskova (2003), and further work is needed to capture the intuition that shorter is more unmarked for default allomorph selection within an implementation that does not wreak typological havoc under factorial reranking.

Not all cases of default choice seem amenable to markedness, however, particularly when inflectional, rather than derivational, morphology is involved. Bonet *et al.* (2007) and Mascaró (2007) posit pairwise preference constraints, e.g. "prefer *a* to *la*" in Haitian Creole. The more complicated use of such constraints are in cases such as the Catalan gender markers, which Bonet *et al.* posit are Ø > /u/ > /ə/ for masculine, and /ə/ > Ø for feminine. Clearly, as the relative preference for /ə/ and Ø is reversed in each gender, it is not possible to reduce these to general segmental markedness hierachies in the language. The implementation of these preference constraints involves pairwise rankings which can be used when there are more than two allomorphs, and a clear potential advantage of this implementation is that it allows *sequences* of defaults.

Stepping outside of OT models, in Distributed Morphology, e.g. Embick's (2009) model, statement of defaults is accomplished by a list of vocabulary items, which are specified for insertion in certain contexts. As schematized in (31) for subcategorization models in general, the default item is thus one with the least amount of contextual specification. Although the optimizer is often not explicitly listed as performing an optimizing function in such models, one can clearly envision a variant in which the contextual specification of vocabulary items mentions *removing* violations to phonotactic constraints. A schematic proposal along these lines for the English indefinite article appears in (38).

(38) *"Positive licensing" of non-default vocabulary item*

INDEF ↔ /æn/ if it removes a violation of ONSET
INDEF ↔ /a/

Such models can be likened to Kager's (1996) proposal, in which the default has an explicit statement "GEN=*n*," but differ in their implementation, in that the default allomorph in Distributed Morphology is precisely one about which nothing special needs to be said in the grammar.

In many cases of allomorph selection involving "recruiting" allomorphs from the "wrong" context, such as Spanish definite article allomorphy or Hebrew plural allomorphy, the default is chosen because it matches the gender features of the head noun, and that is enough to normally drive selection of *la* over *el* for feminine nouns, without need for appealing to phonological markedness in order to make *la* the default. Contextual specification – in terms of morphosyntactic features – is enough to choose the right allomorph, and this is sometimes implemented in OT with constraints referring to matching morphosyntactic features (e.g. Steriade 1999; Becker 2009). In such cases, the default allomorph is not the one that is radically underspecified, but rather the one that realizes the correct morphosyntactic features. This intuition is not straightforward to translate into vocabulary items without negative statements, but one way to accomplish it, while preserving the notion that /el/ is "recruited," is to actually view *it* as the default, used for both masculine gender and for cases where the feminine *la* fails:

(39) *"Negative licensing" of non-default vocabulary item*

DEF, +FEM ↔ /la/ if it does not create a violation of *á*-hiatus
DEF ↔ /el/

Clearly, future work will be needed in order to examine whether a true merger of the elsewhere-notion of Distributed Morphology can be made fully compatible with capturing the phonotactic generalizations that govern allomorph distribution, or whether the priority constraints of Bonet *et al.* (2007) merge more seamlessly with a grammatical treatment of phonological optimization.

4 Conclusion and outlook

Phonologically conditioned allomorph selection refers to any case in which allomorphs are chosen based on the phonology of the stem, affix, or phonological word to which they attach. We have seen that a very healthy number of these cases not only refer to phonology, but seem to involve allomorph distribution that is actively connected with improvement in (or avoidance of declination in) phonological well-formedness, at the levels of segmental, syllabic, and prosodic structure. The most active questions of debate thus revolve around whether, given that *some* allomorph selection is phonologically optimizing, this should be built into the architecture of the grammar, or whether on the contrary, given that some allomorph selection is *not* phonologically optimizing, a single mechanism that simply lists phonological environments is all that is needed. Potentially one way of resolving the question of whether allomorph selection is the result of grammatical computation, rather than morpholexical listing, would involve demonstrating infants' and adults' preferences in (artificial) language acquisition, and finding that in the face of sparse or limited background evidence, such learners demonstrate a preference for phonologically optimizing patterns of allomorph selection, before they have even had a chance to attempt or fall back on lexical listing.

Certainly one might argue that patterns of spontaneous allomorph recruitment, such as BPE speakers' use of [- əd] after all obstruent-final verbs in English, despite never having heard this in the input, reflect pure grammatical biases brought to the task of allomorph selection rather than rehearsed morpholexical listing or historical residue. Similarly, the study of Becker (2009), discussed above, demonstrates that Hebrew speakers apply phonologically conditioned allomorph selection to novel "wug" stems they have never had to inflect before.

A second issue, mentioned in the introduction, is that a great deal of allomorphy need not involve multiple URs, but rather, like the English plural alternations, can be captured entirely in terms of a single UR coupled with phonological rules. Given this division of labor between phonological derivations as one mechanism of yielding allomorphy and morphological selection as another, we may also see divergent acquisition profiles for these two.

In phonological theory, many debates that involve recurrent grammatical generalizations in some languages and exceptions to these generalizations in other languages often devolve into discussions of whether everything is the result of diachrony or not, and end up as philosophical stalemates. Certainly, while Carstairs (1990) is right in raising the point that "The existence of a phonologically conditioned alternation does not by itself prove the existence of some synchronic phonological process giving rise to it," the existence of synchronic grammatical control over allomorph selection *can* be demonstrated when it spontaneously arises in (artificial) language acquisition experiments in which the diachrony is fully controlled by the experimenter and nonetheless the learner demonstrates the emergence of a preference for phonological optimization based on sparse or insufficient evidence.

My own contention is that enough evidence is beginning to accumulate that phonologically conditioned allomorph selection *is* within the purview of the phonological grammar and not merely the lexicon, and that some of the most important questions that therefore arise involve the levels of representation at which allomorph selection occurs and the mechanism of choice, including guaranteeing the default over the optimizer in cases in which euphony is not at stake. The cases of opaque allomorph selection make clear that it needs to happen at intermediate levels of representation, but leave open many possibilities in terms of whether the right model of serial morphology–phonology interaction is stratal, cyclic, or derivational. A much harder and unresolved question is the mechanism for allomorph selection, where many theoretical alternatives are good at capturing some generalizations but in doing so may fail to capture others. A clear dichotomy in this respect is the trade-off between relying on segmental markedness to allow the default to "emerge" without explicit statement and explicit default status of an allomorph through either brute force constraints or by the Elsewhere condition.

In the above paragraphs, I have provided a critical comparison of a wide range of current options (and apologize for ones I have no doubt overlooked in such a large field of inquiry), but most importantly, I have attempted to outline new directions for synthesizing the advantageous aspects of some of these models with those of others. One of the best ways for this study to advance is by figuring out how to incorporate the seemingly incompatible but useful assumptions of one model into those of another.

ACKNOWLEDGMENTS

I am grateful to the editors and two anonymous reviews, and to Birgit Alber, Michael Becker, Ricardo Bermúdez-Otero, and Moira Yip for their comments and suggestions on an earlier draft.

REFERENCES

Alber, Birgit. Forthcoming. Past participle in Mòcheno: Allomorphy, alignment and the distribution of obstruents. In Michael T. Putnam (ed.) *German-language speech islands: Generative and structural approaches*. Amsterdam & Philadelphia: John Benjamins.

Anderson, Stephen R. 2008. Phonologically conditioned allomorphy in Surmiran (Rumantsch). *Word Structure* 1. 109–134.

Aranovich, Raúl & Orhan Orgun. 2006. Opacity in *-ez/-eza* suffixation. In Timothy L. Face & Carol A. Klee (eds.) *Selected proceedings of the 8th Hispanic Linguistics Symposium*, 116–122. Somerville, MA: Cascadilla.

Baptista, Barbara O. & Jair da Silva Filho. 2006. The influence of voicing and sonority relationships on the production of English final consonants. In Barbara O. Baptista & Michael Watkins (eds.) *English with a Latin beat: Studies in Portuguese/Spanish–English interphonology*, 73–89. Amsterdam & Philadelphia: John Benjamins.

Becker, Michael. 2009. Phonological trends in the lexicon: The role of constraints. Ph.D. dissertation, University of Massachusetts, Amherst.

Beckman, Jill N. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14. 1–46.

Bermúdez-Otero, Ricardo. 2007. Morphological structure and phonological domains in Spanish denominal derivation. In Fernando Martínez-Gil & Sonia Colina (eds.) *Optimality-theoretic studies in Spanish phonology*, 278–311. Amsterdam & Philadelphia: John Benjamins.

Bermúdez-Otero, Ricardo. 2009. The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features. Unpublished ms., University of Manchester.

Bonet, Eulàlia, Maria Rosa Lloret & Joan Mascaró. 2007. Lexical specifications and ordering of allomorphs: two case studies. *Lingua* 117. 903–927.

Bonet, Eulàlia, Maria Rosa Lloret & Joan Mascaró. Forthcoming. The prenominal allomorphy syndrome. In Bernard Tranel (ed.) *Understanding allomorphy: Perspectives from Optimality Theory*. London: Equinox.

Booij, Geert. 1998. Phonological output constraints in morphology. In Wolfgang Kehrein & Richard Wiese (eds.) *Phonology and morphology of the Germanic languages*, 143–163. Tübingen: Niemeyer.

Bye, Patrik. 2008. Allomorphy: Selection, not optimization. In Sylvia Blaho, Patrik Bye & Martin Krämer (eds.) *Freedom of Analysis?*, 63–92. Berlin & New York: Mouton de Gruyter.

Carstairs, Andrew. 1988. Some implications of phonologically conditioned suppletion. *Yearbook of Morphology 1988*. 67–94.

Carstairs, Andrew. 1990. Phonologically conditioned suppletion. In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.) *Contemporary morphology*, 17–23. Berlin & New York: Mouton de Gruyter.

Crosswhite, Katherine. 1998. Segmental *vs.* prosodic correspondence in Chamorro. *Phonology* 15. 281–316.

Drachman, Gaberell, René Kager & Angeliki Malikouti-Drachman. 1996. Greek allomorphy: An optimality account. *University of Trondheim Department of Linguistics Working Papers* 28. 345–361.

Dressler, Wolfgang U. 1977. Phono-morphological dissimilation. In Wolfgang U. Dressler & Oskar E. Pfeiffer (eds.) *Phonologica 1976*, 41–48. Innsbruck: Innsbrucker Beiträge zur Sprachwissenchaft.

Embick, David. 2009. Localism versus globalism in morphology and phonology. Unpublished ms., University of Pennsylvania.

Gibson, Masayuki. 2008. Opaque allomorph selection: The need for intermediate forms. Paper presented at the 82nd Annual Meeting of the Linguistic Society of America, Chicago.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.

González, Carolina. 2005. Phonologically-conditioned allomorphy in Panoan: Towards an analysis. *UCLA Working Papers in Linguistics* 11: *Papers in Phonology* 6. 39–56.

Gouskova, Maria. 2003. *Deriving economy: Syncope in Optimality Theory*. Ph.D. dissertation, University of Massachusetts, Amherst.

Halle, Morris. 1997. Impoverishment and fission. *MIT Working Papers in Linguistics* 30. 425–450.

Harris, James W. 1987. Disagreement rules, referral rules, and the Spanish feminine article *el*. *Journal of Linguistics* 23. 177–183.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Jakobson, Roman. 1948. Russian conjugation. *Word* 4. 155–167.

Kager, René. 1996. On affix allomorphy and syllable counting. In Ursula Kleinhenz (ed.) *Interfaces in phonology*, 155–171. Berlin: Akademie Verlag.

Kaye, Jonathan. 1992. Do you believe in magic? The story of s+C sequences. *SOAS Working Papers in Linguistics and Phonetics* 2: 293–313.

Kikuchi, Seiichiro. 2006. On Galician definite article allomorphy. *Onin Kenkyuu* 9. 41–48.

Klein, Thomas B. 2003. Syllable structure and lexical markedness in Creole morphophonology: Determiner allomorphy in Haitian and elsewhere. In Ingo Plag (ed.) *The phonology and morphology of Creole languages*, 209–228. Tübingen: Max Niemeyer.

Lapointe, Steven G. 1996. Comments on Cho and Sells. *Journal of East Asian Linguistics* 5. 73–100.

Lapointe, Steven G. 1999. Stem selection and OT. *Yearbook of Morphology 1999*. 263–297.

Löfstedt, Ingvar. 2008. Perceptually driven allomorphy: The Swedish non-neuter singular definite article. In *Papers from the Annual Meeting of the North East Linguistic Society* 38. 87–98.

Mascaró, Joan. 2007. External allomorphy and lexical representation. *Linguistic Inquiry* 38. 715–735.

Mester, Armin. 1994. The quantitative trochee in Latin. *Natural Language and Linguistic Theory* 12. 1–61.

Nespor, Marina, Marcela Peña & Jacques Mehler. 2003. On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio* 2. 203–230.

Oostendorp, Marc van. 2009. How I learned how to stop worrying and love the derivation. Paper presented at the 17th Manchester Phonology Meeting.

Paster, Mary. 2005. Subcategorization vs. output optimization in syllable-counting allomorphy. *Proceedings of the West Coast Conference on Formal Linguistics* 24. 326–333.

Paster, Mary. 2006. Phonological conditions on affixation. Ph.D. dissertation, University of California, Berkeley.

Perlmutter, David. 1998. Interfaces: Explanation of allomorphy and the architecture of grammars. In Steven G. Lapointe, Diane Brentari & Patrick M. Farrell (eds.) *Morphology and its relation to phonology and syntax*, 307–338. Stanford: CSLI.

Pinker, Steven & Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28. 73–193.

Rubach, Jerzy & Geert Booij. 2001. Allomorphy in Optimality Theory: Polish iotation. *Language* 77. 26–60.

Russell, Kevin. 1995. Morphemes and candidates in Optimality Theory. Unpublished ms., University of Manitoba (ROA-44).

Smith, Norval. 1976. -Aar. *Leuvense Bijdragen* 65. 485–496.

Steriade, Donca. 1999. Lexical conservatism in French adjectival liaison. In J.-Marc Authier, Barbara E. Bullock & Lisa A. Reed (eds.) *Formal perspectives on Romance linguistics*, 243–270. Amsterdam & Philadelphia: John Benjamins.

Steriade, Donca. 2008. A pseudo-cyclic effect in Romanian morphophonology. In Asaf Bachrach & Andrew Nevins (eds.) *Inflectional identity*, 313–359. Oxford: Oxford University Press.

Suh, Yunjuh. 2008. Korean suffix allomorphy in Optimality Theory. *Proceedings of SUNY/CUNY/NYU Mini-Conference* 8. Available (June 2010) at http://web.gc.cuny.edu/dept/lingu/liba/CSN2006.html.

Suzuki, Keiichiro. 1998. A typological investigation of dissimilation. Ph.D. dissertation, University of Arizona.

Tranel, Bernard. 1997. Suppletion and OT: On the issue of the syntax/phonology interaction. *Proceedings of the West Coast Conference on Formal Linguistics* 16. 415–429.

Walter, Mary Ann & Florian Jaeger. 2008. Constraints on optional that: A strong word form OCP effect. In *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 41. 505–519.

Wolf, Matthew. 2008. Optimal interleaving: Serial phonology–morphology interaction in a constraint-based model. Ph.D. dissertation, University of Massachusetts, Amherst.

Yip, Moira. 1988. The Obligatory Contour Principle and phonological rules: A loss of identity. *Linguistic Inquiry* 19. 65–100.

Zwicky, Arnold M. 1987. Suppressing the Zs. *Journal of Linguistics* 23. 133–148.

100 Reduplication

Eric Raimy

1 Introduction

Reduplication is the phonological repetition of segmental material triggered by a morphological source. This definition captures the core aspects of this linguistic phenomenon and distinguishes it from other phenomena that cause the surface repetition of phonological material. This chapter provides a synthesis of questions about reduplication based on classic and contemporary models of reduplication. The most important questions about reduplication can be summarized as the *strong hypothesis for reduplication* (SHR).

(1) *The strong hypothesis for reduplication*

a. *Architectural modularity*
There is a morphology module that is distinct and prior to a phonology module. Both modules have internal structure.

b. *Bipartite reduplication*
The morphology module creates a reduplicated structure, but segmental copying occurs later in the phonology module.

c. *Identity is synchrony*
The source of identity effects in reduplicated forms is the fact that the repeated segments in the output are a single synchronous representation prior to copying. After copying has occurred, the repeated segments are distinct representations and can diverge in identity on the basis of the general application of phonological rules.

The proposed answers to these questions identified by the SHR demonstrate that reduplication informs us about grammatical architecture in general, the relation between the morphology and phonology components, and phonological identity. Two important topics in reduplication are not represented in the SHR. The first is the question of global *vs.* local computation in grammar and the second is reduplicative templates. Each of these topics will be discussed, when they are raised with respect to particular models of reduplication.

This chapter is laid out as follows. First, classic models of reduplication that help define the SHR are presented in §2. Part of the discussion of each classic model is identifying the contribution that it makes to the SHR. §3 turns to contemporary models of reduplication. Fundamental differences in contemporary models can be identified by each model's orientation to the SHR. §4 identifies open questions about reduplication that deserve attention because they raise fundamental questions about what linguistic phenomena should be treated as reduplication.

2 Classic models of reduplication

Classic models of reduplication are no longer being actively pursued. The importance of knowing about them, though, resides in the contribution that each of them made to the *strong hypothesis for reduplication.*

2.1 *Unusual phonology in reduplication*

Wilbur (1973) marks the beginning of the formal investigation of reduplication in generative grammar. Wilbur defines R_o as "the portion of the unreduplicated form . . . of which a copy is made" and states: "the part which is the copy will be referred to as . . . R_r" (1973: 7). This nomenclature also identifies, but does not name, the region of the unreduplicated base that is not part of R_o. Using this notation a reduplicated form can be delimited into sections, as in (2b, 2c). Both R_o and R_r are separated by a dash. R_o is enclosed in square brackets and R_r is underlined. Any remaining segments are part of the unreduplicated base.

(2) *VC reduplication in Chumash* (Wilbur 1973: 7)

a.	/ʔas/ + redup ($C_1V_1C_2$-V_1C_2)	ʔas-as	'chin'
b.	suffixing reduplication	ʔ-[as]-as	
c.	infixing reduplication	ʔ-as-[as]	

Wilbur notes that in cases of total reduplication when R_o and R_r are identical, it cannot be determined whether the reduplicated structure is R_o-R_r or R_r-R_o. In cases of partial reduplication, R_r can be determined by considering the unreduplicated form that consists of R_o and the rest of the base. R_r will then be the remaining segmental material that is repeated. This parsing, when applied to (2b, 2c), identifies one of the [as] sequences as R_r, but we cannot distinguish between (2b) and (2c). One way to distinguish between these options is that the placement of R_r should not alternate among prefixing/suffixing/infixing patterns, and should match the general morphological processes of the given language. For example, if a language does not have any infixes, then R_r should not be analyzed as an infix, (2c).

Wilbur (1973) demonstrates that there are non-trivial interactions between reduplication and phonological rules, because of the Chomsky and Halle (1968: 236) hypothesis that all morphological rules precede all phonological rules, reduplication being a morphological rule. Wilbur (1973) describes three different types of interactions between reduplication and phonology.

(3) *Reduplication and phonological rule interaction* (Wilbur 1973)

a. *normal application*
"Forms where R_r and R_o are not identical at the surface are the result of the normal application of any phonological rule to a form which meets its structural description." (1973: 15)

b. *failure to apply*
"those forms in which either R_o or R_r meets the structural description of a phonological rule and yet has not undergone that rule." (1973: 18)

c. *overapplication*
"Overapplication of a phonological rule refers to the fact that in many reduplicated forms, the structural change applies to a form that does not meet the structural description of the rule." (1973: 26)

The terms "normal application" and "overapplication" have been retained to this day, but "failure to apply" is now normally termed "underapplication." Only normal application is accounted for if the morphology before phonology hypothesis from Chomsky and Halle (1968) is strictly adopted. The reason for this is that, if reduplication is a morphological rule, the copying of the segments to create R_o and R_r will occur before any phonological rule has the opportunity to apply. Consequently, all phonological rules will apply after reduplication, and thus there can only be transparent application of a phonological rule.

Over- or underapplication of a phonological rule suggests that either the morphology before phonology hypothesis, (1a) of the SHR, or some aspect of the model of reduplication must be modified. The basic ordering problem at issue is that R_r must be created at some point in the derivation. Overapplication is problematic, because this situation requires ordering a phonological rule before a morphological process. If the relevant phonological environment appears in the R_o, then ordering the phonological rule prior to reduplication will solve the problem. Underapplication is a more difficult problem, because there is no way to warrant blocking the application of a rule if a relevant structural environment appears in R_o during the derivation.

Wilbur's response to over- and underapplication phenomena was to keep the morphology before phonology hypothesis intact, but to import the concept of global rules (Lakoff 1970) into phonology and propose an Identity Constraint (IC) (CHAPTER 74: RULE ORDERING).

(4) *The Identity Constraint* (Wilbur 1973: 58)

There is a tendency to preserve the identity of R_o and R_r in reduplicated forms.

The IC acts as a diacritic on a phonological rule that allows the rule to over- or underapply to ensure that R_o and R_r are identical at the surface.

Wilbur presents an example from Serrano (Hill 1967) as a case that crucially requires the IC. The important interaction is between an optional phonological rule that inserts a homorganic high vowel between a consonant and glide sequence and an adjectival reduplication rule. Hill (1967: 223) notes: "the reduplication involved in [yi'a:yi'a'n] ~ [yi'a:iyi'ai'n] 'be beautiful' (< #yə'a# 'beautiful') must be introduced *at the same time as* the rule whereby an anticipatory *i* is optionally introduced before *y*" (emphasis and transcription as in original).

(5) *Ordering in Serrano*

a.	*order 1*		b.	*order 2*	
	UR	REDUP+jəʔa+ʔn		UR	REDUP+jəʔa+ʔn
	Glide insertion	*n/a*		Reduplication	jəʔaːjəʔaʔn
	Reduplication	jəʔaːjəʔaʔn		Glide insertion	jəʔaːi̲jəʔaʔn
	Other rules	jiʔaːjiʔaʔn		Other rules	jiʔaːi̲jiʔaʔn
	SR	jiʔaːjiʔaʔn		SR	*jiʔaːi̲jiʔaʔn
				???	jiʔaːi̲jiʔai̲ʔn

(5) demonstrates that local ordering cannot derive the data from Serrano. Order 1 (5a) causes the glide insertion rule to not apply, because the rule's structural description is not met. The surface effect is equivalent to when the glide insertion rule optionally does not apply. Thus this ordering tells us nothing. Order 2 (5b) has the glide insertion rule apply after reduplication. The glide insertion rule inserts an [i] only between the R_r and R_u. The SR form in (5b) shows normal application of the glide insertion rule, but Hill does not list it as a possible form in Serrano. Neither of the possible rule orderings can produce the Serrano form, which shows overapplication of the glide insertion rule.

The key to understanding the importance of the type of pattern is Hill's insight that reduplication and the rule must apply "at the same time." Synchrony of reduplication and rule application is required because reduplication creates the environment that triggers the phonological rule. McCarthy and Prince (1995: 289) term this type of interaction *back-copying*. Back-copying interactions provide the strongest evidence for some form of global computation in phonology, such as Wilbur's IC. Wilbur's model of reduplication assumes strict local computation. This means that the rule and reduplication will each only apply once and that they must be ordered with respect to each other. (5) shows both possible orderings of the rule and reduplication; neither ordering produces the correct forms. Consequently, the data from Serrano provide strong evidence for Wilbur's IC, which explains why the glide insertion rule overapplies in (5b).

The main conclusion from Wilbur on phonology reduplication interactions is that the hypothesis that reduplication is a morphological process and all morphology precedes all phonology can be maintained if a limited amount of global computation is added to the phonological component. Global computation in the phonology is limited to the IC, which can cause phonological rules to over- or underapply to maintain identity between R_u and R_r.

2.2 *Reduplication as morphology*

Carrier (1979) is in essence a direct response to Wilbur (1973) with respect to reduplication and argues against global computation in phonology. Carrier demonstrates that global computation in phonology is not sufficient to account for reduplication patterns. Carrier (1979) points to the interaction between syncope and reduplication (presented as R2 in (6)) in Tagalog as a case where an IC approach is insufficient.

(6) *Syncope and reduplication interaction in Tagalog* (Carrier 1979: 174)

a.		/sunud-in/	b.		/sunud-in/
	Syncope	sund-in		R2	sunudsunud-in
	R2	sundinsundin		Syncope	*sunudsundin
				(overapplies)	*sundsundin

Carrier argues that only by ordering syncope before reduplication can the correct forms in Tagalog be produced. Carrier's R2 pattern of reduplication in (6) copies a foot's worth of phonological material (CHAPTER 40: THE FOOT) from the left edge of the stem. The syncope rule deletes the final vowel of a root that has been suffixed. (6a) shows that if syncope applies before reduplication, then the correct surface form is produced where the vowel of the suffix is copied. If reduplication applies before syncope, as in (6b), then there is no way to produce the correct surface form, because the vowel of the suffix is not copied. The IC can only affect whether the syncope rule over- or underapplies and cannot cause reduplication to re-copy segmental material.

Carrier develops the implications of an analysis of Tagalog where syncope precedes reduplication. She retains the strong hypothesis that all morphology precedes all phonology and argues that reduplication is the result of a [+reduplication] feature being added as part of a word formation rule. This morphological feature triggers the application of a transformational rule that causes the copying of phonological material at the end of the morphology. The bipartite nature of reduplication in Carrier's system allows for morphological rules to apply to a stem with the [+reduplication] feature prior to reduplicative copying. An immediate benefit of the bipartite hypothesis is that it predicts morphological rules can produce over- and underapplication effects, because they will necessarily be ordered before reduplicative copying.

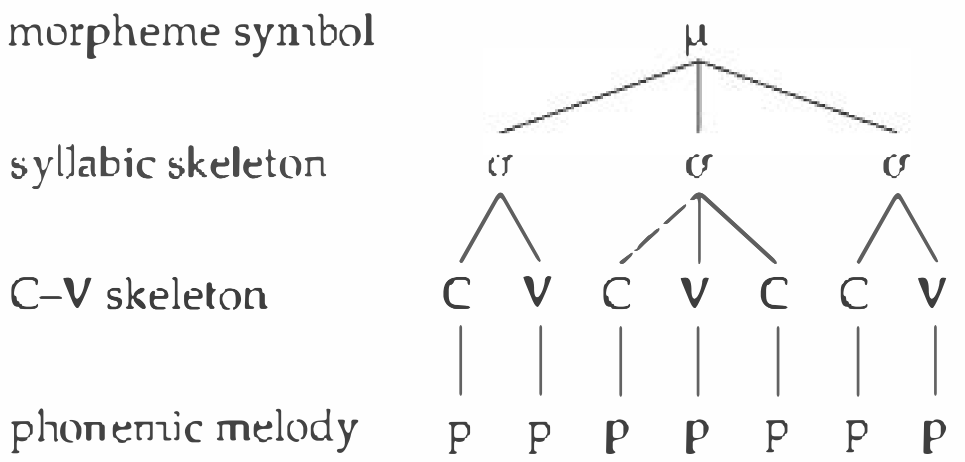
Carrier develops the bipartite hypothesis by arguing that the rules in Tagalog that show overapplication effects (e.g. syncope and nasal substitution) are morphological rules, not phonological. Carrier defines a morphological rule as one that cannot be specified in purely phonological terms. For example, syncope applies to some roots (e.g. /sunud/ → [sund-in]) but not others (e.g. /li:nis/ → [li:nis-li:nis-in]), and nasal substitution does not apply to all /ŋ/-final prefixes and obstruent-initial stems (Carrier-Duncan 1984: 274).

Carrier's contribution to our understanding of reduplication is to develop the morphological aspects of reduplication. Specifically, the hypothesis that reduplication is a bipartite process where the morphology marks a representation as reduplicated but the copying process occurs later in the derivation is extremely important. The bipartite hypothesis allows morphological rules to apply before reduplicative copying, which provides one source of explanation for over- and underapplication effects with local computation.

2.3 *Reduplication as phonology*

Marantz (1982) is a watershed for the study of reduplication. The main proposal extends ideas from McCarthy (1981) that propose prosodic morphology, (7).

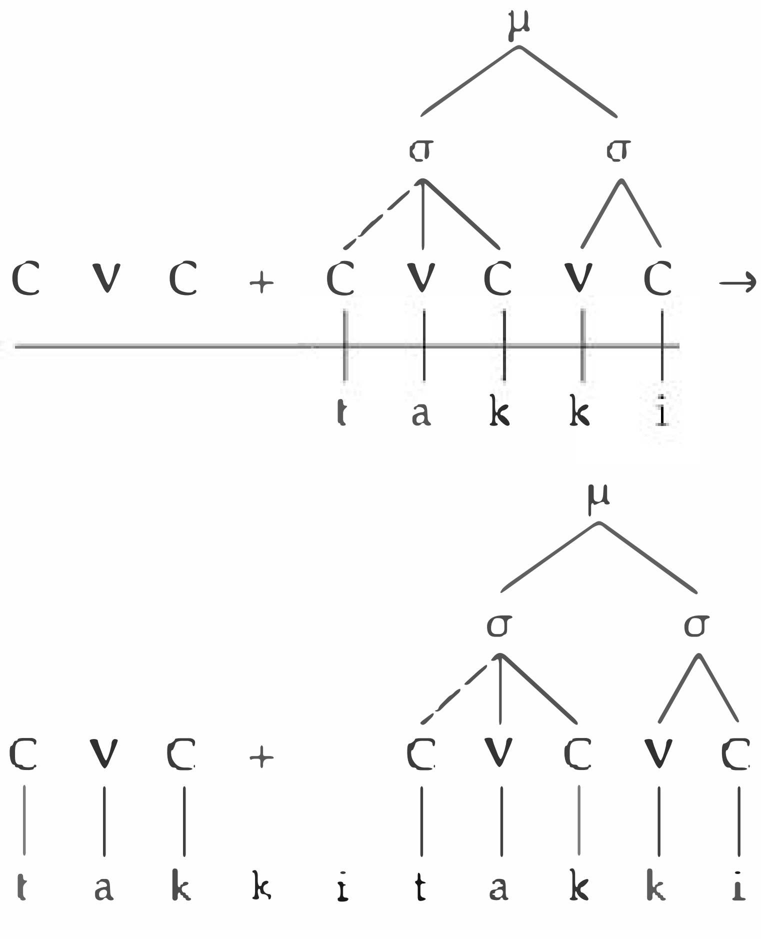
(7) *Prosodic structure of morphemes* (Marantz 1982: 454)



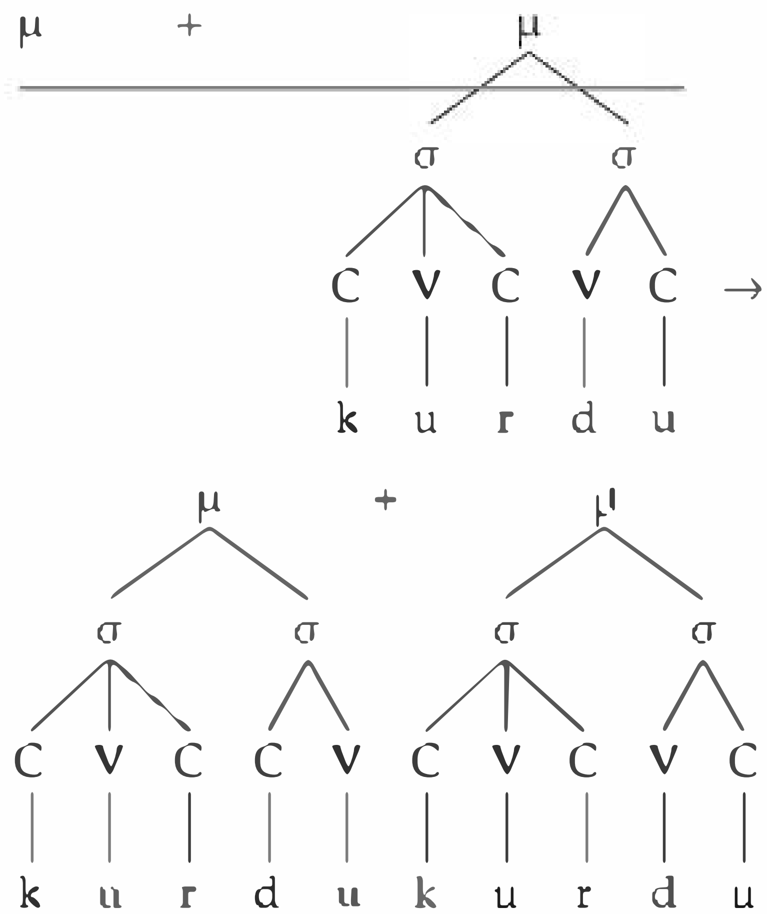
Marantz (1982) proposes that reduplicative morphemes are different from other morphemes in two ways. The first is that they are not fully specified and the second is that they trigger the copying of phonological material of the stem. The lack of specification at some level of representation allows for different patterns of reduplication to be specified by the "prosodic skeleton," which determines how much and what type of phonological material is copied. (8) presents a case of partial reduplication and total reduplication from a Marantzian perspective.

(8) *Marantzian skeletons*

a. *CVC reduplication: Agta* /takki/ → [tak-takki] (Healy 1960)



b. *total reduplication: Walpiri* /kurdu/ → [kurdu-kurdu] (Nash 1980)



Marantz focuses on expressing reduplication patterns at the C–V skeletal level due to Moravcsik's (1978: 307) observation that "reduplicated phonetic strings I found [are] invariably defined in reference to consonant–vowel sequence and absolute linear position." Moravcsik's observation suggests that all of the work being done by templates is thus at the C–V level. (8a) demonstrates the affixation of a CVC prefix that accounts for one pattern of reduplication in Agta. Total reduplication can be produced by affixing a higher level of prosodic structure, the morpheme, which will cause copying of all prosodic structure under the morpheme level, (8b). Deriving different R_r shapes from general phonological representations is a major advance in the understanding of reduplication, because questions about reduplicative templates are now questions about phonological representations and processes.

The other major insight from Marantz (1982: 436) is the proposal that reduplication is a normal affixation process. The phonology of reduplication should mirror the phonological behavior of other affixes (CHAPTER 104: ROOT–AFFIX ASYMMETRIES). By integrating reduplication into the general theories of phonology and morphology, Marantz argues that over- and underapplication effects are accounted for. One source of underapplication effects is the cyclic application of phonological rules to morpheme-internal environments (CHAPTER 85: CYCLICITY). Cyclic rules do not apply in non-derived environments (CHAPTER 88: DERIVED ENVIRONMENT EFFECTS), and if the R_r is an affix (thus a morpheme), then the effect of underapplication will be produced, because R_r will not be a derived environment. Morphological rules will also produce over- and underapplication effects, as in the Carrier bipartite hypothesis.

Marantz develops the hypothesis that phonological aspects of reduplication should be accounted for by general phonological means. Consequently, approaches to reduplication changed as phonological representations have changed. There are two important changes to phonological representations that affected the basic aspects of the Marantzian model of reduplication.

The first change is the single melody model proposed by Mester (1986) (CHAPTER 54: THE SKELETON). Note that proposals in Clements (1985) are extremely similar to Mester's, the difference being whether melodic or templatic effects are the main focus. Clements focuses on templatic effects while Mester focuses on melodic effects. Mester (1986: 172–173) proposes that reduplicative morphemes are attached to the stem in a synchronous manner. This can be viewed as an extension of autosegmental representation (Goldsmith 1976; CHAPTER 14: AUTOSEGMENTS), where different aspects of phonological representations are not linearly ordered with respect to each other. A key aspect of the synchronous nature of reduplicative template and stem is that the melodic tier is shared between them.

Mester's (1986: 190–196) single melody analysis of the interaction of the *ruki* rule (e.g. alveolars become retroflexed after /r/, /u/, /k/, or /i/) and reduplication in Sanskrit demonstrates the advances of this model (see CHAPTER 119: REDUPLICATION IN SANSKRIT). When the root /saɲʤ/ 'hang' is reduplicated and affixed with a *ruki*-triggering prefix (e.g. /ab^{h}i/), then the initial /s/ in the root /sanj/ shows overapplication of the *ruki* rule because the second retroflexed s does not occur after a segment triggering the *ruki* rule (e.g. *[ab^{h}i-ʂa-saɲʤ]). Without the *ruki*-triggering prefix, no retroflexion occurs.

(9) *Overapplication of the* ruki *rule in Sanskrit reduplication*

saɲʤ 'hang' ab^{h}i-ʂa-ʂaɲʤ 'cursed' *ab^{h}i-ʂa-saɲʤ
cf. sa-saɲʤ

(10) demonstrates the parts of a single melody analysis for these forms. (10a) presents the unaffixed root form. The prefixing CV reduplication pattern affixes a synchronous C–V skeleton to the melodic level, (10b). (10c) shows the further affixation of the prefix /ab^{h}i/ and the application of the *ruki* rule (following Mester's 1986: 192 formulation).

(10) *Single melody analysis*

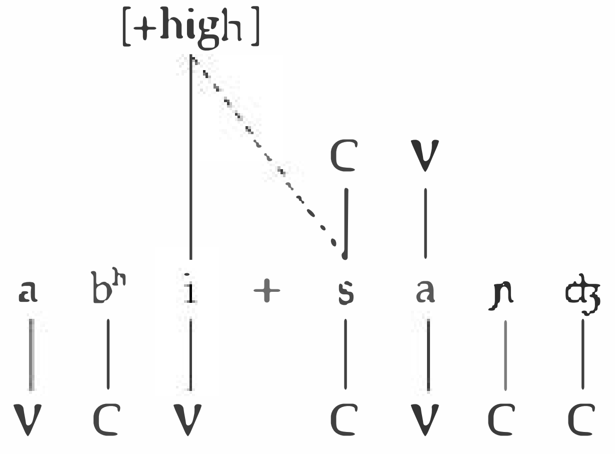
a. *root*

```
s  a  ɲ  ʤ
|  |  |  |
C  V  C  C
```

b. *root and synchronous reduplicative affix*

```
C  V
|  |
s  a  ɲ  ʤ
|  |  |  |
C  V  C  C
```

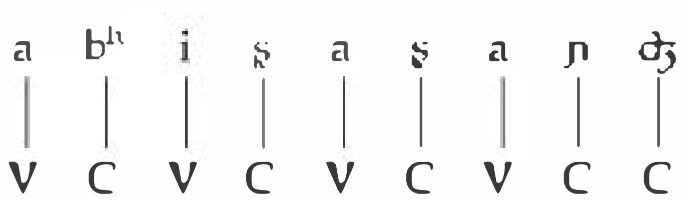
c. *prefixation and application of the* ruki *rule*



Understanding the representation in (10c) is the key to the advances from the single melody model. At this point in the derivation, there is only a single /s/ melody, which was retroflexed by the *ruki* rule. This /s/ is associated to two CV skeletons, one from the stem and one from the reduplicative morpheme. This is the strongest possible identity relationship between R_b and R_r, in that the melodic content of both is a single representation. Any melodic change to one must occur to the other. This is the source of explanation for overapplication effects in the single melody model. Any rule changing the melodic content of a synchronous representation can show overapplication effects. The single melody model also captures Carrier's insight about the bipartite nature of reduplication. The morphology builds a reduplicated structure that is later dealt with by the phonology. In this case, the structure produced by the morphology is one in which R_b and R_r are synchronous.

Non-linear representations like (10c) must be converted to a strictly linear representation. Tier Conflation (McCarthy 1986; Mester 1986: 176–177) converts (10c) into the fully linear representation in (11). Note that concatenation will create a relationship on the phonological skeleton between the /abʰi/ prefix and the reduplicated root in (10c) that is sufficient to trigger the *ruki* rule even though full linear order may not exist between the morphemes at this point in time.

(11) *Linearization as Tier Conflation*



Tier Conflation is not a reduplication-specific device, in that it is the general device to ensure that phonological representations are strictly linear, so the phonetics component can use them. McCarthy (1989) argues that planar segregation is very common in phonological representation. Consequently, a process like Tier Conflation that will provide a strict linear ordering to a representation is required in phonology and will apply to both concatenative and non-concatenative morphologies. Consequently, all phonological representations must undergo some process analogous to Tier Conflation. This is a very important observation, because it accomplishes the complete naturalization of reduplication in phonology. (See also CHAPTER 105: TIER SEGREGATION.)

The publication of Mester (1986) coincided with fundamental changes in prosodic morphology. McCarthy and Prince (1996) drastically revise the prosodic hierarchy to eliminate the CV level of representation. This change represents the debate between the x-slot and moraic theories of the syllable (see Kenstowicz 1994: 425–431; CHAPTER 54: THE SKELETON). McCarthy and Prince (1996) side with the moraic model and provide analyses of reduplicative templates and other phenomena using only legitimate prosodic categories. (12) presents the proposed prosodic categories from McCarthy and Prince (1996: 6).

(12) *Prosodic categories*

Wd	prosodic word
F	foot
σ	syllable
σ_{μ}	light (monomoraic) syllable
$\sigma_{\mu\mu}$	heavy (bimoraic) syllable
σ_{c}	core syllable

Hayes and Abad's (1989) analysis of Ilokano heavy syllable reduplication represents the fundamental arguments for the McCarthy and Prince (1996) approach to reduplicative templates.

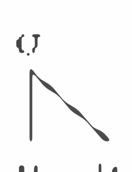
(13) *Heavy syllable reduplication for Ilokano* (Hayes and Abad 1989: 360)

a.	tra'baho	'to work'	ʔag-trab-tra'baho	'is working'
b.	'daʔit	'to sew'	ʔag-daː-'daʔit	'is sewing'
			ʔag-dad-'daʔit	
c.	'pjek	'chick'	pjeː-'pjek	'chicks'
			piː-'pjek	

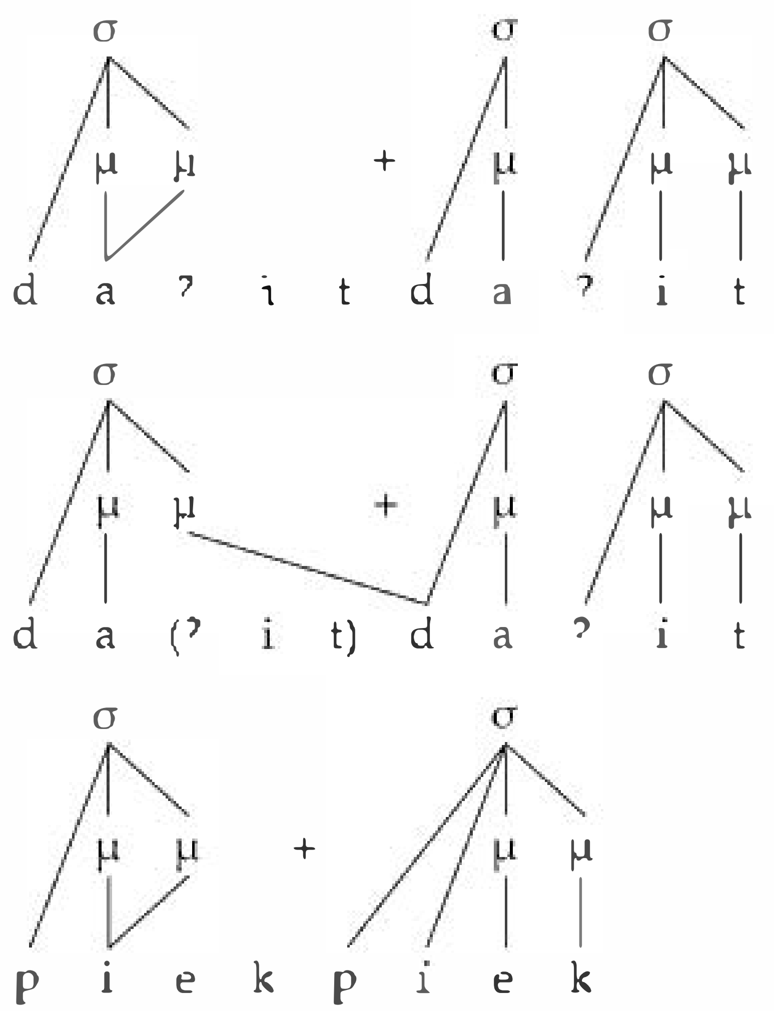
The complexity in this reduplication pattern lies in defining what satisfies the heaviness requirement. (13a) shows that a coda consonant is acceptable, and (13b, 13c) show that lengthening the first vowel copied from the R_0 also satisfies the requirement. There is variation in the forms in (13b) and (13c), though. (13b) shows dialectal variation between lengthening the vowel from R_0 and geminating the first consonant in the R_0 in order to satisfy the branching rhyme requirement. Finally, (13c) shows that a glide in an onset from R_0 can be vocalized and lengthened as a possible form of this reduplication pattern. Hayes and Abad suggest that all of these forms can be captured by specifying the reduplicative template as a bimoraic syllable.

(14) *Bimoraic syllable for Ilokano* (Hayes and Abad 1989: 360–361)

a. *heavy syllable template*



b. *examples*



[da:-daʔit]

[dad-daʔit]

[pi:-pjek]

(14) demonstrates that a single generalization at the syllable level can describe the forms that fall into the class of heavy syllable reduplication in Ilokano. All of the R_rs are a single syllable with a branching rhyme, but melodic association varies in the different forms for both phonological and morphological reasons. The variation seen in (13b) and (13c) is morphological in nature, because it cannot be predicted by phonological considerations.

Extending the original proposals from Marantz (1982) with the single melody model proposed in Mester (1986) and revisions to the prosodic hierarchy from McCarthy and Prince (1996) formalize a strong organic theory of reduplication. This theory is organic because the source of explanation for reduplication is solely from the general theory of morphology and phonology. The aspects of reduplication that make it unique are its bipartite and synchronous characteristics. Reduplication is bipartite because the morphology builds a synchronous phonological structure that is linearized later in the phonology. Reduplication is synchronous because a single phonemic melody is associated with multiple distinct prosodic structures. These two characteristics provide both morphological and phonological sources of over- and underapplication effects. Both morphological and cyclic phonological rules will apply before Tier Conflation and thus can be the source of over- and underapplication effects. Normal application effects are produced by rules that apply after Tier Conflation. Kiparsky (1986: 83) summarizes this situation as

> in principle, as strong a hypothesis as one could hope for. But our present picture of the articulation of phonology and morphology being as tentative as it is, [their proposals on morphology and cyclic phonology] are not easy to verify or falsify.

Although Kiparsky is cautionary, reduplication has been reduced to a unique synchronous representation and to the general question of what the nature of phonology and morphology is.

2.4 The full copy model

Steriade (1988) presages many of the themes dominating work on reduplication at the present time. Steriade's (1988: 146) main proposal is "that templates are not strings of concrete, fillable slots, but rather abstract conditions on prosodic weight and syllabic organization of strings." The reflex of this is the *full copy* model of reduplication, where all reduplication patterns start with R_r consisting of a complete copy of the segmental and prosodic structure of R_u and the rest of the base. Partial reduplication patterns are then produced by eliminating structure from the R_r on the basis of different prosodic parameter settings. Steriade argues that the full copy of prosodic structure remedies inadequacies in the way syllabic transfer effects are accounted for in the Marantzian model of reduplication.

Steriade's main demonstration of the full copy model involves the analysis of two reduplication patterns from Sanskrit. The intensive reduplication pattern in Sanskrit is a prefixing CVX pattern with a prespecified /a/ as the nucleus of the R_r.

(15) *Intensive reduplication in Sanskrit*

a. *Full and zero grade forms in Sanskrit* (Steriade 1988: 108)

root	*intensive full grade*	*intensive zero grade*	
svap/sup	saː-svap-	sau-sup-	'sleep'
		[sošup-]	

b. *Full copy parameter settings* (Steriade 1988: 107)

parameters
  weight parameters
    monosyllabic foot (heavy monosyllable)
  syllable markedness parameters
    obligatory onset: unmarked setting (onset is obligatory)
    complex onset: unmarked setting (onset may not be complex)
    sonorant coda: unmarked setting (coda must be a sonorant)
insertion rule
  Insert /a/ in the intensive stem
  Insertion site: first syllable, rhyme

The two intensive forms in (15a) for the root /svap/ 'sleep' provide the background on understanding how the parameters in (15b) ensure that the R_r will end up being a CVX sequence with a prespecified /a/. Note that the difference between the full grade and zero grade forms in (15a) is based on whether the vowel *a* is deleted from the root. When the *a* is deleted in the zero grade, the *v* vocalizes to *u*. The zero grade form in brackets shows the application of other rules in Sanskrit. The weight parameter is the primary source to truncate R_r to produce patterns of partial reduplication. The particular setting in (15b), monosyllabic foot, for the Sanskrit intensive pattern will ensure that the R_r has a branching rhyme. The syllable markedness parameters provide the source of other modifications to the R_r. Banning complex onsets will cause the R_r to delete segments if there is a complex onset.

The obligatory onset parameter bans the R_r from beginning with a vowel, which accounts for vowel-initial roots in Sanskrit not having a reduplicated intensive form. The sonorant coda parameter captures the observation that the R_r in intensive forms ends with either a long vowel or a sonorant consonant. Finally, there is an insertion rule that adds the prespecified /a/ vowel into the rhyme of the R_r.

(16) demonstrates how the full copy model accounts for the intensive forms for the root /svap/. To begin, the zero grade rule that deletes an unstressed /a/ in the stem occurs before full copy (Steriade 1988: 94–95) in the intensive forms. There are two distinct stems (i.e. one that has undergone zero grade formation and one that has not) to start the derivation of intensive forms.

(16) *Intensive reduplication in Sanskrit* (Steriade 1988: 109)

	full grade	*zero grade*
input	svap	sup
copy	svap-svap	sup-sup
/a/-insertion	*blocked* (OCP)	saup-sup
removal of unlicensed material		
complex onset	sap-svap	*n/a*
obstruent coda	sa-svap	sau-sup
prosodic weight		
rhyme lengthening	saː-svap	*n/a*
output	saː-svap	sau-sup

(16) requires only a few comments. The copy process takes as input the relevant full grade or zero grade form and the left copy is identified as R_r and thus subject to modification. The /a/-insertion rule applies first and is blocked if the R_r already contains an /a/ melody (as in the full grade form). Following this, the syllable markedness parameters remove any material that is not licensed. This causes the complex onset /sv/ to be simplified to /s/ and the non-sonorant /p/ to be deleted in R_r. Finally, the weight parameter ensures that R_r is a monosyllabic foot and lengthens the /a/ in the full grade form. The zero grade does not need this lengthening, because the diphthong /au/ satisfies this weight requirement.

The forms at the end of (16) are not the final output of the phonology. In particular, the output from reduplication of the zero grade /sau-svap/ must undergo the *ruki* rule to convert the /s/ following the R_r to a retroflex, and the /au/ diphthong must be converted to /o/. These additional changes produce the final form [soṣup]. The important aspect of this is that phonological processes have another chance to apply to the reduplicated form and further obscure relationships between R_r and R_o.

Steriade (1988) does not provide an exhaustive list of parameters and insertion rules for the full copy model. This is prudent, because Steriade draws the connection between modifications to R_r in the full copy model and general morphological processes found in non-reduplicative contexts. Examples of non-reduplicative truncation in Madurese and French hypocoristics and segment insertion in English strong verbs and Kaingang verbal formation are provided as examples of processes producible by similar parameters applied in a non-reduplicative context. Thus the question of what the parameters are in the full copy model is the question as to what a possible morphological or phonological rule is.

When a full copy may occur in the derivation is an important issue, because it determines what kinds of interactions between reduplication and phonological rules can occur. Steriade (1988: 141), following Kiparsky (1986), suggests that full copy will occur either at the lexical level or at the end of the cyclic level, with these levels being defined in Kiparsky's (1982) terms. Full copy in intensive reduplication in Sanskrit presumably applies at the cyclic level, which would allow zero grade formation to occur prior to reduplication, while full copy applies at the lexical level in the perfect reduplication pattern in Sanskrit, since Steriade (1988: 123–124) has the zero grade syncope rule apply after reduplication. This position basically echoes the claims about Tier Conflation in Mester (1986), with the modification that Tier Conflation can occur at different points in the derivation.

2.5 *Classic models of reduplication and the SHR*

The SHR in (1) can be seen as the result of the arc of research on reduplication spanning from Wilbur (1973) to Steriade (1988). Each of these approaches to reduplication assumes architectural modularity (1a), in that reduplication is a morphological process that interacts with phonology. Carrier (1979) introduces the bipartite (1b) aspect of reduplication, where a reduplicated structure created by the morphology is interpreted later by the phonology. The main gain here is that morphological rules can apply before reduplicative copying occurs, which is one source of over- and underapplication effects. Adoption of (1b) allows the "identity is synchrony" clause (1c) of the SHR to be adopted, providing a general understanding of the interaction between reduplication and phonological rules. Any morphological or phonological rule that is ordered before reduplicative copying occurs can produce over- and underapplication effects because R_o and R_r are a single representation at this point in time. Rules that apply after reduplicative copying occurs will produce normal application effects because at this point R_r and R_o are distinct. Although Wilbur (1973) adopted limited global computation in the Identity Constraint (4), all other classic models of reduplication favored local computation.

3 Contemporary models of reduplication

All contemporary models of reduplication are reactions to the SHR in (1). There is an interesting cyclic nature to the contemporary models, since they begin by returning to Wilbur's work on reduplication and the question of global computation in grammar.

3.1 *The parallel Correspondence Theory model*

McCarthy and Prince (1995) begin the contemporary era of work on reduplication by completely breaking away from the SHR in (1). The rejection of the SHR is based on the importance of parallel computation in reduplication. McCarthy and Prince (1995: 258) state that:

> In particular, most versions of Optimality Theory assume that constraints on all aspects of phonological structure are applied in parallel (Prince and Smolensky 1993). Inputs are mapped directly to outputs, in an essentially flat derivation.

Because a flat derivation is being assumed, the necessarily local and serial aspects of the SHR, such as separate morphology and phonology modules, the resulting bipartite nature of reduplication from these distinct modules, and the explanation of over- and underapplication effects due to the "before and after" aspects of the "identity is synchrony" hypothesis are all lost in a parallel computational model. Note that strict parallelism is not a necessary component of OT (e.g. Harmonic Serialism; McCarthy 2010), so the tenets of the SHR could be followed in OT, but McCarthy and Prince (1995) explicitly reject them.

As part of the rejection of the SHR and the adoption of global computation for reduplication, McCarthy and Prince (1995) suggest that the conclusions about reduplication should be applied to phonology in general. Correspondence Theory (CT) adopts Wilbur's IC and global computation as the core insight into reduplication and phonology. With global computation, a surface representation can be calculated from a memorized representation in a single computation, with any grammatical aspect potentially affecting the results. CT in practice eliminates all of the components of the SHR, because global computation removes the local distinctions between morphology and phonology. (17) presents the full model of reduplicative identity.

(17) *The full model of reduplicative identity* (McCarthy and Prince 1995: 273)

input	/Af_{RED} + Stem/	
IR-faithfulness	⇗ ⇅	IB-faithfulness
output	R ⇄ B	
	BR-identity	

Reduplication remains the result of the affixation of an underspecified morpheme, but the way phonological content is assigned to this affix is different. There is no copying process in CT, because a copying process operates in a computationally local way, where R_o (B in CT) can determine aspects of R_r (R in CT), but not the other way around (see McCarthy and Prince 1995: 292–294). Instead, reduplicative constructions have two direct correspondences and one indirect correspondence. The two direct correspondence relationships are BR-identity, which allows R_r and R_o to affect each other, and IR-faithfulness, which allows the input stem to influence the realization of R_r. The indirect correspondence for R_r is the general input–output faithfulness relationship between the stem and base. These relationships are all formally the same and operate in a global manner, which allows analyses of reduplicative phenomena unavailable to the classic models.

The bulk of McCarthy and Prince (1995) focuses on overapplication effects. The Malay nasal harmony example presents the strongest case for global computation in reduplication, because it is a case of back-copying.

(18) *Reduplication and vowel nasalization in Johore Malay* (Onn 1976: 180)

underlying	*stem*		*doubled stem*	
/hama/	hamə̃	'germ'	hãmə̃-hãmə̃	'germs'
/waŋi/	waŋĩ	'fragrant'	wãŋĩ-wãŋĩ	'fragrant (INTENS)'
/aŋan/	aŋãn	'reverie'	ãŋãn-ãŋãn	'ambition'
/aŋin/	aŋĩn	'wind'	ãŋĩn-ãŋĩn	'unconfirmed news'

Nasal harmony in Johore Malay operates in a left-to-right manner, as can be seen in the stem forms. The importance of the reduplicated forms is that the first vowel is nasalized even though it is not preceded by a nasal segment. Local computation has difficulties producing this effect because it would have to copy (/hame-hame/), have vowel nasalization apply (/hamẽ-hãmẽ/), and then re-copy to produce [hãmẽ-hãmẽ]. This is clearly an unattractive scenario.

McCarthy and Prince (1995: 289–294) discuss how the Johore Malay data are naturally captured in the global computation of correspondence theory. The constraints in the tableau below are straightforward (e.g. Ident-BR[nasal] is violated when R_o and R_r differ in the feature nasal, $*NV_{oral}$ is violated when a nasal segment is followed by an oral vowel or glide, $*V_{nasal}$ is violated when a vowel or glide is nasalized, and Ident-IO[nasal] is violated when the input stem and output base differ in the feature nasal).

(19) *CT analysis of Johore Malay* (McCarthy and Prince 1995: 291)

/RED-waŋi/	Ident-BR[nas]	$*NV_{oral}$	$*V_{nas}$	Ident-IO[nas]
☞ a. w̃ãŋĩ$_B$-w̃ãŋĩ$_R$			******	***
b. waŋĩ$_B$-waŋĩ$_R$		*!	**	*
c. waŋĩ$_B$-w̃ãŋĩ$_R$	**!		****	***

Given the constraint ranking in (19), (a) is the most harmonic candidate, showing a back-copying effect where the R_r provides the environment to nasalize the initial segments of the R_o and this alternation is transferred back to R_r. The global computation allows for the transfer of any alternation from the R_o to the R_r regardless of the source of the alternation, even if the cause of the alternation is R_r itself. Notice that normal application of nasal assimilation can be produced by simply ranking Ident-BR[nasal] below $*V_{nasal}$.

Adopting global computation increases the importance of generalizations on outputs, and this has resulted in further investigation of ideas about templates from Steriade (1988). *Generalized template theory* (McCarthy and Prince 1994a; Urbanczyk 2006) proposes that the particular shapes of reduplicative templates can be derived from language-specific prosodic requirements. In other words, the parameter settings from Steriade (1988) should be determined by a language's phonology and morphology instead of having to be set for each reduplicative morpheme.

Urbanczyk (2006) argues that a reduplicative template's prosodic weight can be predicted by its morphological status. Where Steriade (1988: 83) would specify a light syllable by setting a weight parameter to "unfootable," as an idiosyncratic aspect of the particular reduplicative morpheme, Urbanczyk derives the presence or lack of a coda in R_r (producing a weight contrast) from whether the reduplicative morpheme is classed as root or affix. Roots have additional faithfulness relationships (e.g. BR-Max(Rt)) that can support more marked phonological material. Affixes lack additional faithfulness pressures, and are thus subject to the effects of *the emergence of the unmarked* (TETU, McCarthy and Prince 1994b; CHAPTER 58: THE EMERGENCE OF THE UNMARKED).

(20) *Morphological status and reduplicative template shape in Lushootseed* (Urbanczyk 2006)

RED	*morphological status*	*reduplicated*
DIM	affiix	q'i-q'ix^{w}
DIST	root	saq$^{w'}$-saq$^{w'}$

(21) CV *vs.* CVC *status in* R$_r$

/DIM-q'ix^{w}/	BR-MAX(Rt)	NOCODA	BR-MAX
☞ a. q'i-q'ix^{w}		*	x^{w}
b. q'ix^{w}-q'ix^{w}		**!	

/DIST-saq$^{w'}$/	BR-MAX(Rt)	NOCODA	BR-MAX
a. sa-saq$^{w'}$	q$^{w'}$!	*	q$^{w'}$
☞ b. saq$^{w'}$-saq$^{w'}$		**	

The diminutive reduplicative morpheme in Lushootseed is an affix, while the distributive morpheme is a root, according to Urbanczyk. The constraint ranking in (21) has a root-specific BR-faithfulness constraint, BR-MAX(Rt), dominating NOCODA, which then dominates the general BR-faithfulness constraint, BR-MAX. This is a classic TETU ranking, which produces the surface effect that the DIST morpheme can contain a coda (as in [saq$^{w'}$-saq$^{w'}$]), while the DIM morpheme cannot (cf. [q'i-qix^{w}]).

(22) demonstrates how Urbanczyk derives the monosyllabicity of all R$_r$s in Lushootseed. By ranking a constraint that penalizes the number of syllables in the output (*STRUCT-σ) above both BR-faithfulness constraints, R$_r$ will be the minimal number of syllables possible (see also Spaelti 1997; Hendricks 1999). All atemplatic models assume some additional constraint that requires the reduplicative morpheme to occur in the output (see proposals by Gafos 1998).

(22) *Monosyllabic R$_r$ without a template in Lushootseed* (Urbanczyk 2006: 202)

/DIM-hiw-il/	*STRUCT-σ	BR-MAX(Rt)	BR-MAX
☞ a. 'hi-hiwil	***		wil
b. 'hiwil-hiwil	****!		

/DIST-pastəd/	*STRUCT-σ	BR-MAX(Rt)	BR-MAX
☞ a. 'pas-pastəd	***	təd	təd
b. 'pastəd-pastəd	****!		

Each of the optimal candidates in (22) has the characteristic that the R$_r$ contains as much phonological material (taking affix *vs.* root status into consideration, which determines whether a coda is in R$_r$ or not; see (21)) from R$_o$ as is needed to add only a single syllable to the output.

Although CT has had a major impact on phonology as part of the rise of Optimality Theory (Prince and Smolensky 1993), there has been little critical evaluation of whether it has furthered our understanding of reduplication. The crux of the matter is determining whether the empirical gains provided by global computation outweigh the results of the SHR in (1). While SHR models do not offer a natural account of back-copying effects, they do provide a general model of how phonology and reduplication interact. Once the point of copying in a derivation is determined for a reduplicative morpheme, there is a prediction about whether normal application for different types of rules (e.g. morphological, lexical, cyclic, etc.) should be the interaction between reduplication and phonology. Because CT abandons the modularity and local computation aspects of the SHR, it does not make any such predictions. The question that must be answered is whether a model with more empirical coverage but little predictive power is better than a model with less empirical coverage but great predictive power.

The proposals on reduplicative templates from CT are mostly extensions of Steriade (1988) because of their claimed connection to TETU effects. Urbanczyk (2006) presents the most restricted version, where prosodic principles are used to produce the reduplicative templates. Urbanczyk's proposals have not been evaluated cross-linguistically on languages with more than three reduplication patterns. Also, Haugen (2008) demonstrates the need for a syllable template to describe various reduplication patterns in Yaqui and Yapese, and work by Hendricks (1999) on bare consonant reduplication suggests that R_rs can be specified as a single segment, which conflicts with Prosodic Morphology (McCarthy and Prince 2001: 1), where reduplicative templates are to be made of authentic units of prosody (see (12)).

The role of TETU in reduplication needs to be critically evaluated, because of its widespread use as explanation in CT. Alderete *et al.* (1999) propose that fixed segmentism in any R_r has one of two sources: TETU or overwriting. The TETU analyses of Lushootseed, Nancowry, Yoruba, and Tübatulabal offered in Alderete *et al.* have all been shown to be inadequate: for Lushootseed, Fitzpatrick and Nevins (2003) demonstrate that fixed /ˈi/ in Lushootseed R_rs results from general considerations of the metrical system and has nothing to do with reduplication per se; for Nancowry, Raimy (2000a: 79–96) demonstrates that the vowel in the reduplicant is not predictable, but must be fully prespecified by the morphology; for Yoruba, Akinlabi (2004) demonstrates that the reduplicant must have a prespecified tone and that the prespecified vowel of the reduplicant is phonologically distinct from the phonetically identical epenthetic vowel derived by TETU; and for Tübatulabal, Cairns (2008) demonstrates that claimed TETU effects fall out from general phonological processes (see the original papers for the full arguments). Furthermore, the general typological claim that reduplicants/affixes will not contain any material more marked than bases/roots is questioned by the distribution of codas in Lakota (Albright 2004). These facts suggest that the Steriade (1988) position that all segmental prespecification in reduplication is carried out via insertion may be more accurate.

3.2 *Precedence-based phonology*

Raimy (2000a, 2000b) proposes that phonological representations consist of not only segments and prosodic structure but also precedence relations that encode the ordering of phonological elements (see CHAPTER 34: PRECEDENCE RELATIONS

IN PHONOLOGY). These proposals change phonological representations from (23a), where order is derived from graphemic conventions (left-to-right indicates order), to (23b), where order is directly indicated by precedence relations (indicated by '→').

(23) *Precedence in phonology*

a. kæt
b. # → k → æ → t → %

Raimy (2000a, 2000b) argues that, once precedence is encoded in representations, more complicated phonological representations can be considered and investigated. Specifically, reduplication results from a phonological representation that contains a transitive symmetrical precedence relation (i.e. a "loop").

(24) presents the unreduplicated and reduplicated forms from Johore Malay that McCarthy and Prince (1995) present as an argument against local computation in reduplication. (24a) is the memorized form for 'fragrant' and (24b) is the reduplicated version of this form.

(24) *Precedence-based reduplication*

a. # → w → a → ŋ → i → %
b. # → w → a → ŋ → i → %
c. # → w → a → ŋ → i → w → a → ŋ → i → %

The form in (24b) contains a "loop," which is the exponence of the reduplicative morpheme. Raimy reformulates the Tier Conflation of earlier bipartite proposals as serialization; see Idsardi and Raimy (forthcoming). The surface result of serialization is the repetition of segments within the loop, producing (24c) from (24b).

Although the "loops" have no privileged status (see Raimy 2009a: 187, n. 4), they are the locus of much explanation of reduplication. Raimy (2000b: 547) points out that the precedence link needed to account for reduplication in (24b) provides the exact phonological environment required to understand the interaction of reduplication and nasal harmony in Malay in a computationally local manner.

(25) *Vowel nasalization and reduplication in Malay*

a. # → w → a → ŋ → i → %
b. # → w̃ → ã → ŋ → ĩ → %
c. # → w̃ → ã → ŋ → ĩ → w̃ → ã → ŋ → ĩ → %

Non-consonantal segments in Johore Malay are nasalized if they are preceded by a [+nasal] segment, with nasalization spreading until a non-nasal consonant is reached. (25a) shows that there is a precedence relationship where /i/ precedes /w/, and this allows nasalization to spread to /w/ and consequently to /a/, producing the representation in (25b), where all of the segments are nasalized. (25c) is the serialized form where the nasalized /w̃/ and /ã/ appear in an environment that does not include a preceding nasal segment. Serialization has eliminated the environment that allowed nasal harmony to occur. This is a classic example of opacity, which is well known to occur with local computation. The conclusion of Raimy (2000a, 2000b) is that global computation is not necessary to account for back-copying or any other interaction between phonology and reduplication.

Reduplicative templates are derived from how precedence links are concatenated to a stem. Raimy (2000a, 2000b) adopts Distributed Morphology (Halle and Marantz 1994) and assumes that the phonological content of a reduplicative morpheme is simply a precedence link (or links and segmental material) that forms a loop when concatenated to the base. Precedence links that are added by the morphology are defined by *anchor points* (Raimy 2009b; see also CHAPTER 34: PRECEDENCE RELATIONS IN PHONOLOGY), which consist of a limited set of positions where links can be added. Depending on how a precedence link is concatenated to a base, segments from different parts of the base will be "in the loop" and reduplicated. Anchor points develop the idea from Moravcsik (1978: 312) that reduplication patterns can be defined in reference to positions of consonants and vowels.

(26) *Reduplicative "templates" in precedence-based phonology*

a. *prefixing* CVC: *Agta* (Healy 1960: 7)

\# → l → a → b → a → n → g → % [lab-labang] 'patches'

b. *suffixing syllable: Dakota* (Sietsema 1987: 337)

Ons Ons

\# → h → ā → s → k → a → % [hāska-ska] 'are tall'

c. *discontinuous: Chukchi* (Dunn 1999: 108)

\# → w → e → n → i → % [weni-wen] 'bell (ABS SG)'

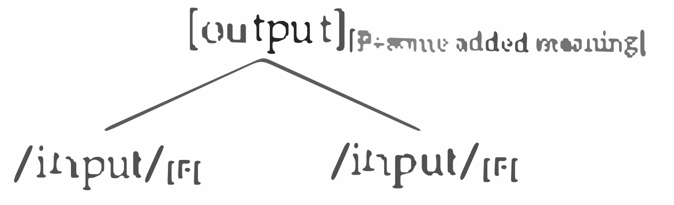
The "loop" in (26a) contains the first CVC of the stem, and these segments will be repeated when serialized. This loop can be defined by the anchor points "after the first vowel" and "first segment." Both of these anchor points are used in infixation (Yu 2007). (26b) presents a suffixing syllable pattern from Dakota that uses the anchor points "last segment" and "last onset." Finally, (26c) shows a discontinuous reduplication pattern from Chukchee that suffixes the first CVC sequence of the base. Discontinuous reduplication patterns that have the R_r separated from R_o in the surface string generally require two precedence links to be added. This pattern requires one precedence link for total reduplication (i.e. "last segment" precedes "first segment") and one precedence link to truncate the R_r to CVC (i.e. "after the first vowel" precedes %). Anchor points are not reduplication-specific, because they provide a general theory of how morphology creates concatenative (e.g. prefixation and suffixation) and non-concatenative (e.g. reduplication, infixation, truncation, root and template, etc.) phonological structures.

The proposals on precedence in phonology in Raimy (2000a, 2000b) provide a local computation solution to back-copying effects. This revitalizes the SHR in (1) as a viable model of reduplication. This makes answering the question about whether global computation in phonology is desirable or not even more important.

3.3 *Morphological Doubling Theory*

Inkelas and Zoll (2005) introduce Morphological Doubling Theory (MDT), which proposes that reduplication is the result of a purely morphological process where the output stem (which shows surface repetition of segments) results from two (or more) daughters that are featurally and semantically identical, as in (27).

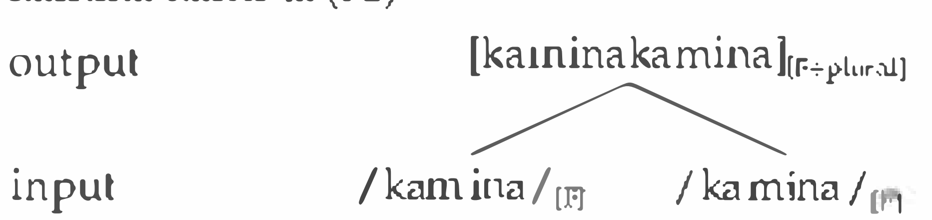
(27) *Morphological Doubling Theory* (Inkelas and Zoll 2005: 7)



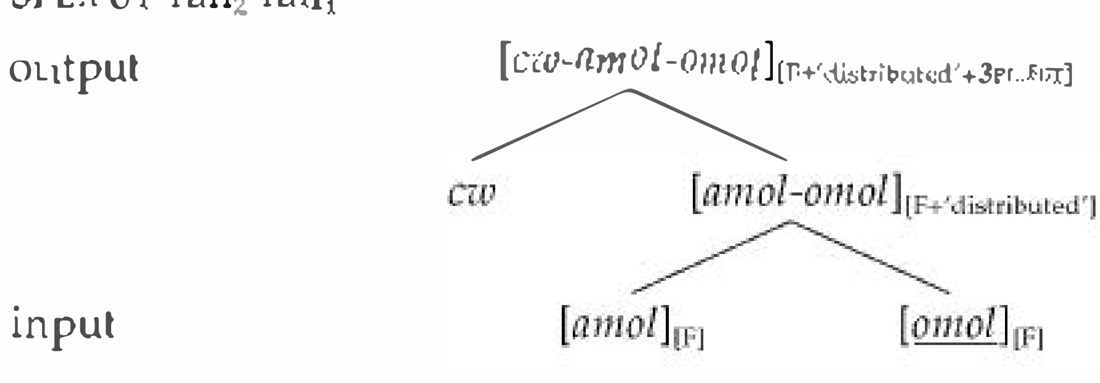
MDT denies the idea that there is phonological copying of any kind (e.g. literal derivational copying, repetition due to serialization or parallel correspondence) in reduplication. Surface repetition of segments is due to multiple independent instances of a stem. One immediate prediction that MDT makes is that total reduplication should be extremely common, since it can be produced immediately from the basic structure in (27). A more interesting prediction that MDT also makes is that a reduplication pattern can consist of two phonologically different allomorphs of a stem.

(28) *MDT basic structure examples*

a. *Warlpiri* (Inkelas and Zoll 2005: 1)
   kamina 'girl'
   kamina-kamina (PL)



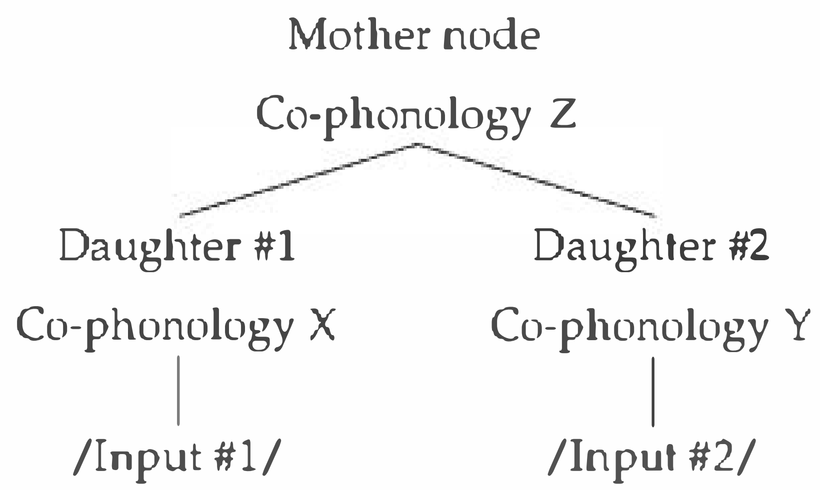
b. *Sye* (Inkelas and Zoll 2005: 54)
   cw-amol$_2$-omol$_1$ 'they will fall all over'
   3PL.FUT-fall$_2$-fall$_1$



(28a) presents an example of total reduplication from Walpiri that is produced by the basic structure of morphological doubling, where two semantically identical stems are inserted and produce total reduplication. (28b) shows that the same structure can produce patterns of reduplication where there is divergence between the two copies. Sye (Inkelas and Zoll 2005: 54, citing Crowley 1998: 79) presents a case where there is total reduplication but different allomorphs of the stem are inserted. /amol/ is what Crowley refers to as a "stem2" and /omol/ is a "stem1." Stem1 is the default form, and Inkelas and Zoll state that the stem2 form appears in a "collection of seemingly unrelated morphological environments" (2005: 52). The phonological difference between the two allomorphs leads to the surface appearance that total reduplication has not occurred.

MDT proposes that there is an analogous phonological side to the basic model in (27), where co-phonologies are associated with each node in the representation.

(29) *Co-phonologies in MDT* (Inkelas and Zoll 2005: 76)



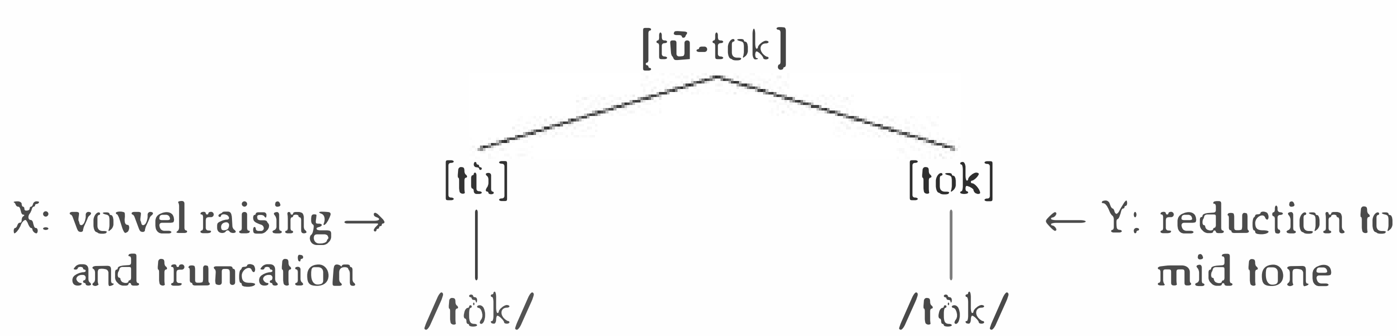
Each node in (29) has its own co-phonology associated with it. Common effects of co-phonologies are truncations that produce partial reduplication patterns. Co-phonologies X and Y will produce phonological changes that are specific to each stem. Co-phonology Z is required to produce juncture effects that hold between the stems only in reduplicative constructions.

Co-phonologies can modify both stems in different ways. Tarok presents an example of this with the reduplication pattern on monosyllabic stems.

(30) *Tarok* (Inkelas and Zoll 2005: 84)

a. ì-gìsàr ì-gìsàr-gisar 'his/her broom'
ì-tòk ì-tù-tok 'his chair'

b.



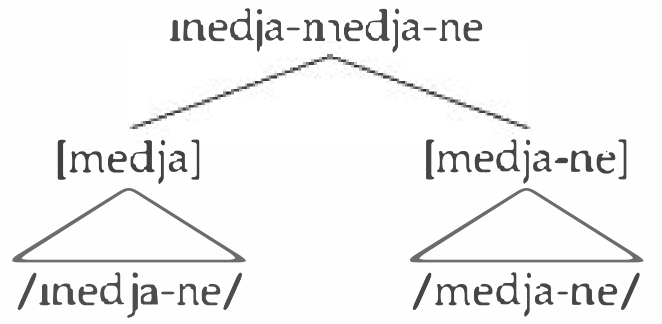
The forms in (30a) demonstrate the processes that are active in reduplicated constructions in Tarok. In all forms, the second stem (governed by co-phonology Y) has its tones reduced to mid regardless of the size of the stem. Co-phonology X must vary, depending on the size of the stem, because, if the stem is larger than a single syllable, no change to the stem occurs. However, if the stem is a single syllable, then the stem is truncated to a CV syllable and the vowel is raised. The example in (30b) lists the processes that each co-phonology is responsible for.

MDT deals with over- and underapplication of phonological processes through morphological truncation. Javanese (see Inkelas and Zoll 2005: 137 for the sources) has an /a/-raising process that interacts with reduplication in an opaque manner.

(31) *Javanese* (Inkelas and Zoll 2005: 136–138)

a. /medja/ medjɔ 'table'
medja-ne 'his/her table'
medjɔ-medjɔ 'tables'
medja-medja-ne 'his/her tables'

b. *Opacity by truncation*



(31) provides the basic template on how over- and underapplication effects are handled in MDT. The data in (31a) demonstrate that /a/ will raise to [ɔ] if it is the last vowel. Suffixation blocks the application of /a/-raising. By comparing the non-suffixed and suffixed reduplicated forms, the opaque interaction can be seen. Whether the first stem undergoes /a/-raising appears to be determined by whether or not the second stem is suffixed. MDT denies any phonological connection between the two stems, and instead proposes that the first stem is actually suffixed, (31b). The presence of the suffix on the first stem will block /a/-raising, and the co-phonology will then truncate the suffix. See Raimy (2006) for discussion of the problems MDT has with opacity in this example.

MDT is conceptually similar to Steriade (1988), in that a full copy of a level of representation is made and there is no transderivational global computation across the two copies. MDT diverges from Steriade's model, though, in disconnecting from a specific model of the phonology–morphology interaction. Co-phonologies are associated with the reduplication constructions, but MDT does not develop any relationship between co-phonologies in reduplicated and non-reduplicated constructions.

Although not working within MDT, Kiparsky (2010) develops a stratal OT model of reduplication that has the characteristics that MDT aspires to. Kiparsky argues that the phonology that an R_r undergoes is determined by the stratal level of phonology that the copy of the stem occurs at. Although there is phonological copying, there is no transderivational identity involved in reduplication, so it is very much in the MDT spirit. The analysis of Sanskrit reduplication presented by Kiparsky (2010) clearly demonstrates that the phonology that the R_r undergoes is predictable from the lexical phonology of Sanskrit.

Proposals by Inkelas and Zoll and by Kiparsky demonstrate that modified versions of the SHR can be developed in OT. These models admit some global computation, but this is limited to modules defined by morphology and phonology.

Reduplication is a bipartite process where R_r is created by the morphology and then subject to the general phonological derivation. Kiparsky differs from Inkelas and Zoll on the issue of identity as synchrony in reduplication. Kiparsky has a phonological version of this hypothesis in which both R_r and R_o are created from the same phonological representation via copying, while Inkelas and Zoll apply this hypothesis to the morphosyntactic level of representation, where R_r and R_o are created from different phonological representations that share a common morphosyntactic identity. Both models agree, though, that the wholesale move to global computation for reduplication in the CT model is unnecessary.

4 Conclusion and future questions

The SHR in (1) provides hypotheses as to how reduplication interacts with morphology and phonology and why reduplication appears to be a unique grammatical phenomenon. Contemporary models of reduplication have more or less followed the original arc of research that created the basis for the SHR. The wheel has not been reinvented by cycling through the evaluation of the parts of the SHR again. On the contrary, contemporary models of reduplication have increased the importance of the SHR, because each model raises more detailed questions about the validity of each component of the SHR.

Evaluation of the *identity is synchrony* clause in (1c) is currently underway in the guise of questioning what kinds of phonology reduplication interactions actually exist. Both Inkelas and Zoll (2005) and Kiparsky (2010) dispute the existence of back-copying. The argument is that the global computation of McCarthy and Prince (1995) and the precedence graphs of Raimy (2000a, 2000b) produce grammars that are more powerful than is necessary to account for reduplication. While the examples of back-copying discussed in the literature are few (e.g. Malay nasal spread, Chaha /x/-dissimilation), more cases do exist (e.g. Serrano). In order to further investigate whether back-copying exists, more examples should be included in the discussion. Obviously, the most fruitful way forward is to develop analyses of the following data in all contemporary models of reduplication to see where the differences in the models arise. Below are additional examples of back-copying that deserve more attention.

(32) *Abkhaz* (Bruening 1997: 325–326)

a. /bʧək/ [bə.ʧək] 'measure of weight'

b. /m/-*reduplication*
[əb.ʧək-m-əb.ʧək]
*[bə.ʧək-m-əb.ʧək]
*[bə.ʧək-m-ə.ʧək]

The Abkhaz data in (32) demonstrate that there is a phonological process that inserts an excrescent vowel between the /b/ and /ʧ/ in the unreduplicated form. When this form undergoes /m/-reduplication, the location of the excrescent vowel changes to precede the /b/ on the basis of the syllabification of the prespecified /m/. It is the syllabification of the prespecified /m/ and the /b/

of the base into a single syllable that places the excrescent schwa before the /b/. This placement of the excrescent schwa is transferred to the first copy (R_1), creating the /əb/ syllable in the reduplicated form. The excrescent schwa appears only in word-initial position, creating a VC syllable in reduplicated forms.

(33) *Korean* (Chung 1999: 177–178)

a.	/hilak/	→	[hirak]	'pleasure'
			[hi-hi-lak-lak] → [hi-hinaŋ-nak]	'rejoicing'
b.	/lwelak/	→	[nwe-rwenaŋ-nak]	'broadminded'
	/l/ → [n]		*[hi-hiraŋ-nak]	
			*[nwe-nwenaŋ-nak]	

The data from Korean in (33) show a complex pattern of normal and over-application of different realizations of the /l/. In (33a) we can see that in the non-reduplicated form the /l/ appears as [r] in an intervocalic environment. The reduplicated form in (33a) shows two interesting effects. First, the form undergoes an AABB reduplication pattern, based on an underlying compound structure to the word. The back-copying effect is present in the interaction between the /kl/ at the juncture between the two copies of the second syllable of the base, /lak/. There is a reciprocal influence on these segments, in that the /k/ preceding an /l/ causes the /k/ to nasalize to [ŋ]. This nasalization process then causes the following /l/ to nasalize to [n]. The nasalization of /l/ to [n] is then transferred to the /l/ of the first copy, [hihi-naŋ-nak], even though it does not follow a nasal segment. Adding to the complexity of these Korean data is the fact that the process that nasalizes the /k/ applies in a normal fashion, so the word-final /k/ is not nasalized. (33b) shows the same derivation for the second syllable /lak/, but the behavior of the /l/ in the first syllable /lwe/ is different. The /l/ in /lwe/ undergoes normal application of the [l ~ r ~ n] distribution, where the /l/ appears as [n] in word-initial position in the first repetition and as [r] intervocalically in the second repetition. The different behavior of the /l/s in the two different syllables creates a very complicated interaction between reduplication and phonology.

(34) *Paamese* (Russell 1997: 109–110)

/muni/	[munu-munu]	'drink'
	*[muni-muni]	
	*[munu-muni]	

The Paamese data in (34) are directly analogous to the Malay nasalization data already discussed in (18) and (25). In Paamese, an /i/ is backed to [u] if it is in non-final position. Reduplication causes the first copy of /muni/ to be in non-final position triggering the backing of /i/ to [u]. This alternation is then copied to the word-final /i/ in the second copy.

All three of these additional instances of back-copying provide examples of the exact base–reduplicant juncture effects that Inkelas and Zoll (2005) and Kiparksy (2010) deny exist. See the original sources for the full details of these examples. One should be cautious about interpreting arguments based on the validity of these data, though. MDT and Stratal OT can produce analyses of these types of data either through additional opaque copying for MDT or by positing special allomorphs in the relevant cases for Stratal OT. Consequently, the existence of back-copying effects (or not) is not as probative in distinguishing models of reduplication as McCarthy and Prince (1995), Inkelas and Zoll (2005), and Kiparsky (2010) suggest.

Another important question about reduplication is whether a repeated string of phonological segments is reduplication or repetition. Gil (2005) discusses this difference and provides examples from a naturalistic corpus of Riau Indonesian.

(35) *Repetition* vs. *reduplication in Riau Indonesian* (Gil 2005)

a. *complete arbitrary reduplication* (2005: 57)
Anton tak sini-sini a
Anton NEG RED-LOC-DEM-DEM:PROX EXCL
[commenting about his friend] 'Anton hasn't been around.'

b. *complete iconic reduplication* (2005: 54)
Naik ojek-ojek-ojek aja
go.up RED-motorbike.taxi just
[group of three persons debating whether to take a single taxi together, or separate motorbike taxis] 'Let's just ride motorbike taxis.'

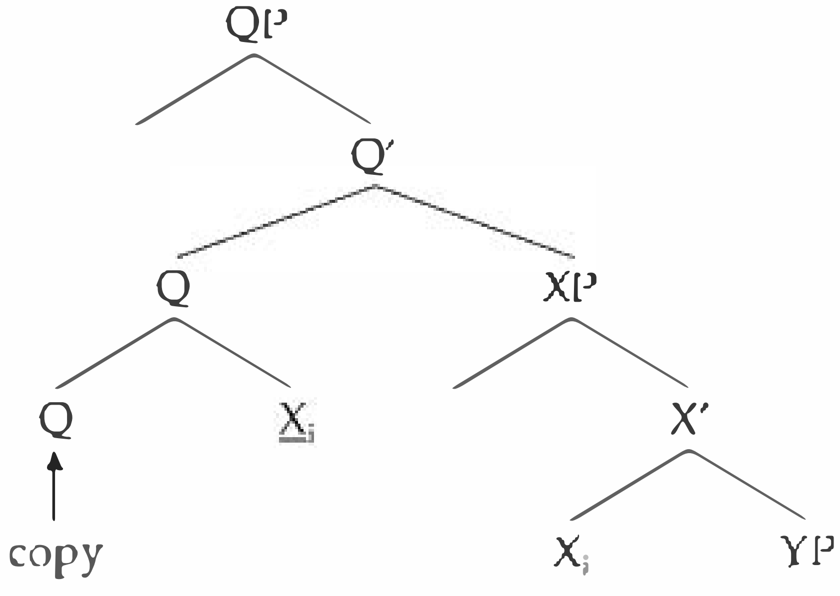
c. *ordinary language: reinforcing repetition* (2005: 43)
Masuk, masuk, masuk, masuk, ah, kelok bola
enter enter enter enter EXCL turn ball
[playing billiards on laptop computer; following the ball as it rolls] 'Go in, go in, go in, go in, damn, the ball curved.'

Gil argues that (35a) is a clear case of reduplication, because the number of repetitions is restricted to two and there is a grammatical function associated with the repetition. (35b) is iconic reduplication, where the number of repetitions of *ojek* 'motorbikes' reflects the number of motorbikes present. This is different from (35a), in that the number of repetitions is variable in (35b), while grammatically fixed in (35a). (35c) presents a case of repetition, not reduplication, because there is no grammatical function of the repetition, only a pragmatic one (i.e. it demonstrates the excitement of the player).

Travis (2003) provides a framework that can potentially distinguish between the examples in (35), on the basis of syntax. Because of the morphological aspects of reduplication, there is the potential for morphosyntactic considerations to play an important role in reduplication. (36) presents the syntax of two different types of reduplicative structures. Q represents a reduplicative morpheme, underlining indicates what syntactic constituent is the target of reduplication, and "copy" indicates where the repetition will occur. (36a) is an example from Tagalog of total reduplication (/lakad/ 'walk' → [mag-lakad-lakad] 'walk a little'), where the capitalized part indicates the R_r of the form. The head of the XP /lakad/ undergoes head movement to adjoin to Q, which is the reduplicative morpheme. Since the X is sister to the Q head, only the content of X is eligible to be copied in reduplication. The syntax of (36b) is different. The Q head copies material from the XP complement and places the copy in the spec position of the QP phrase.

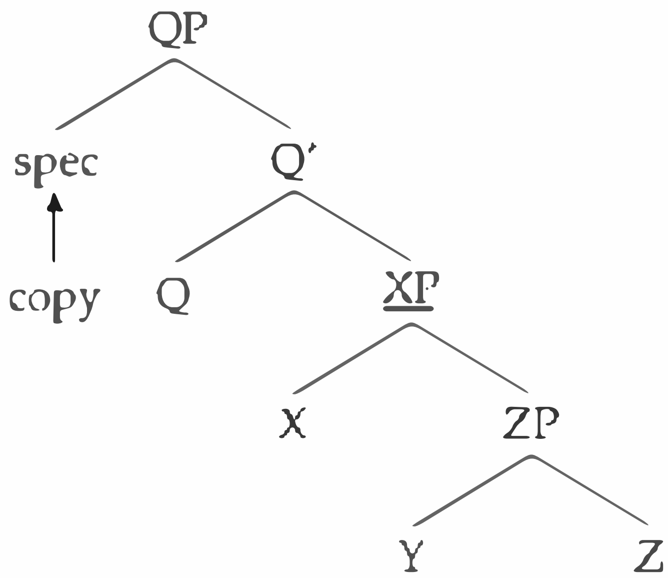
(36) *Reduplication as syntax in Tagalog* (Travis 2003: 240)

a. *phonological*



[magLAKADlakad]

b. *iterative*



CUP after cup of coffee

Travis (2003) captures two important aspects of reduplication: (morpho)syntactic representations limit the amount of phonological material that can be repeated and (morpho)syntax plays a role in determining how many repetitions occur.

Morphosyntax determines the upper bound of reduplicative copying through the sisterhood relationship in syntax. Head movement raises an X to adjoin to Q in (36a). This will limit reduplication to no larger than the phonological content of X and to one copy. X category elements are generally a "word"-sized domain, so this syntactic construction aligns closely with the familiar phonological cases of reduplication. The Q element in (36b) is sister to an XP, which allows for copying of the entire XP. Because the target of copying is an XP, more than a word can be copied in this construction. Furthermore, since the copy in this construction will appear in the spec of QP, more than one repetition can occur if multiple QPs are stacked on top of each other. This accounts for the open-ended number of reduplications in . . . *cup after cup after cup of coffee.* These are only two of the syntactic constructions for reduplication proposed by Travis (2003).

Travis's syntactic differences provide insights into distinctions among the repetition patterns discussed by Gil. (36a) is the morphosyntactic structure for common examples of reduplication such as (35a), while (36b) will account for Gil's (35b) and possibly (35c) examples. (36b) is also appropriate for phrasal reduplication examples like those in (37).

(37) *Phrasal reduplication patterns*

a. *Kannada* (Lidz 2001: 379)
nanrnu baagil-annu much-id-e giigilannu muchide anta heeLa-beeDa
I-NOM door-ACC close-PST-1SG RED that say-PROHIB
'Don't say that I closed the door or did related activities.'

b. *English* (Ghomeshi *et al.* 2004: 326)
Well, he didn't give-it-to-me–*give-it-to-me (he only lent it to me).*

Travis (2003) presents one view of syntactic aspects of reduplication; see Ghomeshi *et al.* (2004), Lidz (2001), and Idsardi and Raimy (forthcoming) for additional relevant data and different perspectives on these syntactic issues. Developing better syntactic analyses of reduplication will improve our understanding of whether there is any connection between semantics and particular reduplication patterns. If syntax plays an important role in reduplication, then there should be constraints on possible reduplication patterns and semantics pairings.

Another burgeoning question about reduplication is whether a surface repetition of segmental material is a reduplicated form or not. Buckley (1998) and Zuraw (2002) provide arguments that morphologically simplex forms can be inherently reduplicated.

(38) *Inherently reduplicated words in Manam* (Buckley 1998: 60)

a.	sa'laga	'be long'	salaga-'laga	'long (SG)'	*salaga-ga
b.	ra'gogo	'be warm'	rago'go-go	'warm'	*ragogo-gogo

Buckley (1998) argues that trimoraic forms in Manam that repeat the last two syllables, (38b) /ra'gogo/, are inherently reduplicated. One reduplication pattern in Manam is to reduplicate the final moraic foot of a form, as in (38a), /sa'laga/ → /salaga-'laga/. Forms like (38b) reduplicate only a single mora in this reduplication pattern, /ra'gogo/ → /rago'go-go/, */ragogo-'gogo/ and Buckley's idea is that if the final syllable is already reduplicated then it will "count" as bimoraic, thus producing only a single syllable for the bimoraic foot reduplication pattern.

(39) *Inherently reduplicated words in Tagalog* (Zuraw 2002: 400)

Intervocalic tapping ([ɾ] / V __ V, [d] *elsewhere*)

transparent	'duːɾi	'loathing'
overapplies	'ɾuːɾok	'acme'
underapplies	'deːde	'baby bottle'

Zuraw (2002) provides examples from Tagalog involving repetition of segmental material that violate general Tagalog phonotactic distribution of [d] and [ɾ]. Zuraw argues that these words can be understood as being inherently reduplicated. Once these forms are designated as reduplicated, their behavior follows documented phonology reduplication interactions such as over- or underapplication of a rule.

The common theme between Buckley and Zuraw's observations is that forms with surface repetition of segmental material that violate otherwise general patterns in a given language can be understood to be inherently reduplicated.

Both Buckley and Zuraw posit an "empty" RED morpheme to provide a grammatical structure to explain the over- or underapplication of some phonological process through reduplication. Fitzpatrick (2006) presents an analysis of the Manam data in the Raimy (2000a, 2000b) model of reduplication by simply allowing "loops" to be parts of an underlying representation.

To summarize this chapter, reduplication as a natural language phenomenon provides insights into grammatical architecture and the workings of grammatical modules. There is a burgeoning consensus about the general nature of reduplication, which can be summarized as the *strong hypothesis for reduplication* in (1). With Optimality Theory turning radically derivational in the form of Harmonic Serialism (McCarthy 2010), it is likely that all contemporary models of reduplication will be in line with the SHR in the near future. This does not mean that there is not disagreement about formal analyses of reduplication; thus further research into reduplication is called for. Future research should refine our understanding of architecture and computation in grammar by developing more explicit analyses of reduplication in more languages. The most difficult question that faces us about reduplication is the parceling out of explanation among potential syntactic, morphological, and phonological sources. By doing this, reduplication will further show its unique status as a natural language phenomenon that involves syntax, morphology, and phonology.

REFERENCES

Akinlabi, Akinbiyi. 2004. Fixed segmentism in Yoruba deverbal nouns. In Kola Owolabi & Ademola Dasylva (eds.) *Forms and functions of English and indigenous languages in Nigeria: A festschrift in honour of Ayo Banjo*, 273–295. Ibadan: Group Publishers.

Albright, Adam. 2004. The emergence of the marked: Root-domain markedness in Lakhota. Paper presented at the 78th Annual Meeting of the Linguistic Society of America, Boston.

Alderete, John, Jill Beckman, Laura Benua, Amalia Gnanadesikan, John J. McCarthy & Suzanne Urbanczyk. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30. 327–364.

Bruening, Benjamin. 1997. Abkhaz Mabkhaz: M-reduplication in Abkhaz, weightless syllables, and base-reduplicant correspondence. *MIT Working Papers in Linguistics* 30. 291–329.

Buckley, Eugene. 1998. Integrity and correspondence in Manam double reduplication. *Papers from the Annual Meeting of the North East Linguistic Society* 28. 59–67.

Cairns, Charles. 2008. Alternate lengthening and stress in Tübatulabal: A modular analysis. Paper presented at the Phonology Forum Conference on the Foot, City University of New York.

Carrier, Jill. 1979. The interaction of morphological and phonological rules in Tagalog. Ph.D. dissertation, MIT.

Carrier-Duncan, Jill. 1984. Some problems with prosodic accounts of reduplication. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound and structure*, 260–286. Cambridge, MA: MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Chung, Chin Wan. 1999. Reduplication in Korean. Ph.D. dissertation, Indiana University.

Clements, G. N. 1985. The problem of transfer in nonlinear phonology. *Cornell Working Papers in Linguistics* 7. 38–73.

Crowley, Terry. 1998. *An Erromangan (Sye) grammar*. Honolulu: University of Hawai'i Press.

Dunn, Michael. 1999. A grammar of Chukchi. Ph.D. dissertation, Australian National University.

Fitzpatrick, Justin. 2006. Sources of multiple reduplication in Salish and beyond. *MIT Working Papers on Endangered and Less Familiar Languages* 7. 211–240.

Fitzpatrick, Justin & Andrew Nevins. 2003. Phonological occurrences: Relations and copying. Unpublished ms., MIT.

Gafos, Adamantios I. 1998. A-templatic reduplication. *Linguistic Inquiry* 29. 515–527.

Ghomeshi, Jila, Ray Jackendoff, Nicole Rosen & Kevin Russell. 2004. Contrastive focus in reduplication in English (the Salad-Salad paper). *Natural Language and Linguistic Theory* 22. 307–357.

Gil, David. 2005. From repetition to reduplication in Riau Indonesian. In Bernard Hurch (ed.) *Studies on reduplication*, 31–64. Berlin & New York: Mouton de Gruyter.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.

Halle, Morris & Alec Marantz. 1994. Some key features of Distributed Morphology. *MIT Working Papers in Linguistics* 21. 275–288.

Haugen, Jason. 2008. The syllable as the base of delimitation of the base for reduplication. Unpublished ms., Williams College.

Hayes, Bruce & May Abad. 1989. Reduplication and syllabification in Ilokano. *Lingua* 77. 331–374.

Healy, Phyllis M. 1960. *An Agta grammar*. Manila: Manila Bureau of Printing.

Hendricks, Sean. 1999. Reduplication without template constraints: A study in bare-consonant reduplication. Ph.D. dissertation, University of Arizona.

Hill, Kenneth Cushman. 1967. A grammar of the Serrano language. Ph.D. dissertation, University of California, Los Angeles.

Idsardi, William J. & Eric Raimy. Forthcoming. Three types of linearization and the temporal aspects of speech. In Theresa Biberauer & Ian Roberts (eds.) *Challenges to linearization*. Berlin & New York: Mouton de Gruyter.

Inkelas, Sharon & Cheryl Zoll. 2005. *Reduplication: Doubling in morphology*. Cambridge: Cambridge University Press.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kiparsky, Paul. 1982. From Cyclic Phonology to Lexical Phonology. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part 1, 131–175. Dordrecht: Foris.

Kiparsky, Paul. 1986. The phonology of reduplication. Unpublished ms., Stanford University.

Kiparsky, Paul. 2010. Reduplication in Stratal OT. In Linda Uyechi & Lian Hee Wee (eds.) *Reality exploration and discovery: Pattern interaction in language and life*, 125–142. Stanford: CSLI.

Lakoff, George. 1970. Global rules. *Language* 46. 627–639.

Lidz, Jeffrey. 2001. Echo reduplication in Kannada and the theory of word-formation. *The Linguistic Review* 18. 375–394.

Marantz, Alec. 1982. Re reduplication. *Linguistic Inquiry* 13. 435–482.

McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12. 373–418.

McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17. 207–263.

McCarthy, John J. 1989. Linear order in phonological representation. *Linguistic Inquiry* 20. 71–99.

McCarthy, John J. 2010. Studying Gen. *Journal of the Phonetic Society of Japan* 13. 3–12.

McCarthy, John J. & Alan Prince. 1994a. Two lectures on prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Rutgers University (ROA-59).

McCarthy, John J. & Alan Prince. 1994b. The emergence of the unmarked: Optimality in prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Rutgers University (ROA-13).

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.

McCarthy, John J. & Alan Prince. 1996. Prosodic morphology 1986. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.

McCarthy, John J. & Alan Prince. 2001. Prosodic morphology: Constraint interaction and satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University (ROA-482).

Mester, Armin. 1986. Studies in tier structure. Ph.D. dissertation, University of Massachusetts, Amherst.

Moravcsik, Edith A. 1978. Reduplicative constructions. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 3: *Word structure*, 297–334. Stanford: Stanford University Press.

Nash, David. 1980. Topics in Warlpiri grammar. Ph.D. dissertation, MIT.

Onn, Farid M. 1976. Aspects of Malay phonology and morphology. Ph.D. dissertation, University of Illinois.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Raimy, Eric. 2000a. *The phonology and morphology of reduplication*. Berlin & New York: Mouton de Gruyter.

Raimy, Eric. 2000b. Remarks on backcopying. *Linguistic Inquiry* 31. 541–552.

Raimy, Eric. 2006. Review of Inkelas & Zoll (2005). *Journal of Linguistics* 42. 478-486.

Raimy, Eric. 2009a. A case of appendicitis. In Raimy & Cairns (2009), 177–188.

Raimy, Eric. 2009b. Deriving reduplicative templates in a modular fashion. In Raimy & Cairns (2009), 383–404.

Raimy, Eric & Charles Cairns (eds.) 2009. *Contemporary views on architecture and representations in phonology*. Cambridge, MA: MIT Press.

Russell, Kevin. 1997. Optimality Theory and morphology. In Diana Archangeli & D. Terence Langendoen (eds.) *Optimality Theory: An overview*, 102–133. Cambridge, MA & Oxford: Blackwell.

Sietsema, Brian M. 1987. Reduplication in Dakota. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 24. 337–352.

Spaelti, Philip. 1997. Dimensions of variation in multi-pattern reduplication. Ph.D. dissertation, University of California, Santa Cruz.

Steriade, Donca. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5. 73–155.

Travis, Lisa. 2003. Reduplication feeding syntactic movement. In Sophie Burelle & Stanca Somesfalean (eds.) *Proceedings of the Canadian Linguistics Association Annual Conference 2003*, 236–247. Montreal: Université de Québec à Montréal.

Urbanczyk, Suzanne. 2006. Reduplicative form and the root–affix asymmetry. *Natural Language and Linguistic Theory* 24. 179–240.

Wilbur, Ronnie. 1973. The phonology of reduplication. Ph.D. dissertation, University of Illinois, Urbana-Champaign.

Yu, Alan C. L. 2007. *A natural history of infixation*. Oxford: Oxford University Press.

Zuraw, Kie. 2002. Aggressive reduplication. *Phonology* 19. 395–439.

101 The Interpretation of Phonological Patterns in First Language Acquisition

Yvan Rose
Sharon Inkelas

1 Introduction

Since the seminal work of Jakobson (1941), phonological patterns observed in child language have been documented and analyzed from a number of perspectives. Throughout the relevant literature, an interesting paradox often manifests itself: while child language is generally characterized as a "simpler" version of the target language, many types of phonological patterns observed in acquisition data create challenges for theories developed to account for more "complex" adult systems. Phonologists within the generative framework have reacted to this problem in a number of ways, from tacitly or conspicuously ignoring acquisition data (e.g. Chomsky and Halle 1968: 331) to elaborating phonological models that account for phenomena observed in child phonology (e.g. Smith 1973; Bernhardt and Stemberger 1998) or, in some extreme cases, rejecting the validity of evidence from child production data for theoretical investigations of phonology (e.g. Hale and Reiss 1998). The disconnect between findings from child phonology and generalizations arising out of generative theories of grammar based on adult phonology has also been recruited in support of alternative, functionalist approaches to child phonology, including Stampe (1969) within Natural Phonology and, more recently, Vihman and Croft (2007) within Construction Grammar.

A general assumption connecting virtually all of these models is that children's phonological abilities are initially impoverished and gradually develop in the face of positive, interpretable evidence from the ambient signal. Regardless of theoretical or philosophical allegiance, researchers acknowledge that theories of linguistics should be learnable and, as such, empirically verified against language development facts. The question, then, is not about whether developmental phonology data should be considered – clearly they should – but about the way they should be incorporated into theoretical debates. As we discuss in this chapter, the main challenge lies in the interpretation of developmental patterns, given that differences between child phonology and adult phonology can be, and have been, attributed variously to perception, grammar, and production.

We begin our discussion with an overview of the types of patterns attested in the literature on phonological development. We then move on to phenomena that have featured in theoretical debates in the field, which we address in light of the main competing approaches to child language phonology. We conclude with desiderata for truly explanatory models of phonological development.

2 A brief survey of phonological patterns

This section presents a survey of the main types of segmental and prosodic patterns found in phonological development. As used here, the term "pattern" refers to any systematic difference between the actual forms produced by a child and the (adult-like) forms the child is evidently attempting. This definition is neutral as to the source of the discrepancy. For example, the child who says [tʰæt] for *cat* might be said to display a process converting target [k] to [t], with /kæt/ as his/her underlying representation; alternatively, that child might be said to have stored /tʰæt/ as his/her lexical representation, which he/she is producing accurately. Both positions, and others, have been taken in the literature.

The overview offered below builds on previous surveys by Menn (1971), Ingram (1974), Ferguson and Farwell (1975), Ingram (1989), Smit (1993), Vihman (1996), Bernhardt and Stemberger (1998), and Inkelas (2003), as well as the contributions to Kager *et al.* (2004) and to McLeod (2007). It provides only a fragmentary view of the larger picture of child phonology, because the literature has thus far been able to focus on only a relatively small number of children, and predominantly on "western" languages (cf. McLeod 2007). The possibility is high that phenomena unattested as of yet will be uncovered as research continues.

2.1 *Segmental patterns*

At the segmental level, a plethora of systematic discrepancies between child and adult pronunciations have been documented. As we can see from the list in (1), virtually all articulatory dimensions of speech are affected. For the purposes of this chapter, transcriptions between vertical bars represent adult-like, target forms. Arrows indicate correspondence between adult target forms and child productions, but should not be taken to represent generative phonological derivations; as stated above, this survey is neutral with respect to the question of the child's lexical representations (see CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION, CHAPTER 13: THE STRICTURE FEATURES, and CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION for overviews of the behavior of these features in adult languages).

(1) *Segmental patterns (non-exhaustive list)*

		Pattern	Illustration[a]		
a.	Place features	Velar fronting (to coronal)	/k/	→	[t]
		Coronal backing (to velar)	/t/	→	[k]
		Labialization	/t/	→	[p/tʷ]
		De/palatalization	/ʧ/	→	[k]
			/k/	→	[ʧ]
		Dentalization	/s/	→	[θ]
		Debuccalization	/s/	→	[h]
b.	Manner features	Gliding/rhoticization	/r/	→	[w]
			/w/	→	[r]
		Stopping/spirantization	/s/	→	[t]
			/t/	→	[s]
		De/affrication	/ʧ/	→	[t]
			/t/	→	[ʧ]
		De/nasalization	/m/	→	[b]
			/b/	→	[m]
		De/vocalization	/j/	→	[l]
			/ɫ/	→	[u]
c.	Voicing features	De/voicing	/d/	→	[t]
			/t/	→	[d]
		De/aspiration	/tʰ/	→	[t]
			/t d/	→	[tʰ]
		De/glottalization	/ʔ/	→	[k]
			/k/	→	[ˀ]

[a] These patterns and those listed in later tables are not limited to the illustrations provided. For example, fronting and backing patterns in (1a) typically target voiced as well as voiceless consonants, even though this generalization is not without exceptions (e.g. McAllister 2009).

Segmental patterns such as those in (1) are often context sensitive, varying across segmental and prosodic contexts. For example, Rose (2000) shows that French target /ʁ/ corresponds to various labial, coronal, and velar consonants in the speech of Clara, a first language learner of Québec French. The place feature that target /ʁ/ assumes in Clara's productions is harmonic with the place of articulation of other consonants present in the target form (e.g. *robe* /ʁɔb/ → [wɔb] 'robe'; *rouge* /ʁuʒ/ → [juʃ] 'red'; *carotte* /kaʁɔt/ → [ka'gɜ]/[kə'jɔːt˺] 'carrot'). Syllable context matters as well. The substitutions affect /ʁ/ in singleton onsets but not in complex onsets (e.g. *citrouille* /sitʁuj/ → [θɔ'tʁuːj] 'pumpkin'). Also, when /ʁ/ occurs in word-medial coda or word-final position in the adult target word, it undergoes deletion altogether (e.g. *ourson* /uʁsɔ̃/ → [ʊ'sɔ̃] 'teddy bear'; *renard* /ʁənaʁ/ → [lena] 'fox').[1] Although substitution patterns are often referred to outside their larger contexts (as in the list in (1), which exemplifies patterns independently of any contextualization), complex conditioning of the kind illustrated by Clara's data is actually quite typical, as other examples discussed below also demonstrate.

[1] See Rose (2000) for further discussion and analysis of these data.

2.2 *Prosodic patterns*

Systematic discrepancies between adult and child productions are also observed in the prosodic domain, i.e. syllable structure, word shape, or the location of stress or tone. A representative selection is offered in (2) (CHAPTER 38: THE REPRESENTATION OF SC CLUSTERS, CHAPTER 67: VOWEL EPENTHESIS, and CHAPTER 36: FINAL CONSONANTS describe related phenomena in adult languages). These examples come from the general surveys cited in the introduction to §2 as well as from targeted studies of prosodic development by Spencer (1986), Fikkert (1994), Demuth (1995), Barlow (1997), Pater (1997), Ota (1999), Rose (2000), Gnanadesikan (2004), Goad and Rose (2004), and Vihman and Croft (2007), amongst others.

(2) *Common prosodic patterns (non-exhaustive list)*

		Pattern	*Illustration*	
a.	Affecting syllables	C + liquid cluster reduction	/pleɪ/	→ [peɪ]/[leɪ]
		s + C cluster reduction	/skiː/	→ [kiː]/[siː]
		Coda deletion	/kæt/	→ [kæ]
		Right-edge cluster reduction	/tɛnt/	→ [tɛn]/[tɛt]
		Consonant/glide epenthesis	/liːoʊ/	→ [liːjoʊ]
b.	Affecting word shapes	Vowel epenthesis/deletion	/bluː/	→ [bəluː]
			/əgɛn/	→ [gɛn]
		Syllable truncation	/ɛləfənt/	→ [ifənt]
		Syllable reduplication	/ɛləfənt/	→ [fæfæ]
		Stress shift	/dʒəˈɹæf/	→ [ˈdʒiːɹæf]

Similar to the segmental patterns illustrated in (1), the source of the prosodic phenomena in (2) is often open to a number of interpretations. Many of these patterns closely resemble phonological subsystems independently observed in adult languages (e.g. syllable reduplication; McCarthy and Prince 1995a), creating the temptation to analyze them as directly driven by the child's grammar. However, as we discuss further below, it is crucial that all available options be considered before any specific analysis is adopted. For example, evidence from infant speech perception shows that English-learning children tend initially to associate stressed syllables with word onsets (e.g. Jusczyk *et al.* 1999). This suggests that the apparent truncation of pretonic material in words such as *ga'zelle*, *a'bout*, and *gui'tar*, yielding productions like *zelle*, *bout*, and *tar*, may originate from a speech segmentation error that yields an incorrect lexical representation, as opposed to being the product of a grammatical rule restricting the prosodic shape of phonological productions. However, perceptual or speech segmentation errors certainly cannot be held responsible for all prosodic patterns. For example, Fikkert (1994) documents a pattern of stress shift displayed by Dutch learners who preserve both syllables of disyllabic words with final stress but systematically produce these forms with stress on the initial syllable, as exemplified in (3).

(3) *Stress overgeneralization* (data from Robin; Fikkert 1994)

gitaar	/χiːˈtaːr/	[ˈsiːtaː]	'guitar'
		[ˈsiːtau]	
giraf	/ʒiˈʀɑf/	[ˈʃiːɑf]	'giraffe'
ballon	/bɑˈlɔn/	[ˈbuːɔn]	'balloon'

Fikkert argues that these productions are conditioned by the learner's grammar, which – influenced by the predominant, trochaic, stress pattern that exists in the target language – regularizes away lexical exceptions. Fikkert's grammatical analysis is further motivated by the broad consensus that children are acutely sensitive to stress (e.g. Morgan 1996; Jusczyk 1997), a fact that argues against simple misperception of the target forms with final stress. (See Kehoe 1997, 1998 and Pater 2004 for related discussions.)

2.3 *Exotic patterns*

In addition to the two relatively clear types of patterns listed above, which have either parallels in adult language, direct phonetic motivations, or both, the literature on early phonological development contains an array of so-called "exotic" patterns, some of which are listed in (4), whose analysis defies both clean-cut classifications and, often, theoretical accounts. The term "exotic" is intended to convey that these patterns are not robustly attested in the literature on adult phonology.

(4) *Some exotic patterns*

	Pattern	*Illustration*
a.	Consonant harmony (affecting major place of articulation)[2]	/teɪbɫ/ → [beːbu]
b.	Vowel-to-consonant harmony (affecting major place of articulation)	/sxunə/ → [bunə]
c.	Long-distance consonant metathesis	/sak/ → [kas]
d.	Consonant fusion	/smouk/ → [fok]
e.	Velar fronting (across the board or in certain syllable positions)	/goʊ/ → [do]
f.	Chain shifts involving consonants	/sɪk/ → [θɪk]; /θɪk/ → [fɪk]
g.	Dummy segment or syllable insertion	/məskiːrou/ → [fi-giro]
h.	Spontaneous language games	/əlæskə/ → [ə'læskə-'bæskə][3]

As one example of an exotic type of pattern, consider the intricate yet systematic interaction between segmental and prosodic factors conditioning the productions of Marilyn, a learner of European French documented by dos Santos (2007). First, Marilyn displays syllable truncation driven by the segmental characteristics of the consonants found in the target forms: when two consonants have identical continuancy features in the target form, Marilyn produces both of these consonants, as illustrated in (5a). Conversely, in (5b), target forms that contain consonants with different continuancy features undergo truncation of the initial syllable.

[2] While consonant harmony is well attested in adult languages, it never affects major place of articulation as it does in child language (e.g. Hansson 2001; see also CHAPTER 72: CONSONANT HARMONY IN CHILD LANGUAGE; CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS).

[3] The pattern illustrated here consists of final foot reduplication combined with [b]-substitution in the foot-initial onset. Language games can take different forms across children; see Inkelas (2003) for some discussion.

(5) *Segmentally conditioned syllable truncation* (dos Santos 2007)

a. *Continuancy-harmonic onsets:*
Production of both onsets (and related syllables)

appétit	|ape'ti|	[pi'ti]	2;00.12[4]	'appetite'
biquet	|bi'kɛ|	[be'kɛ]	1;11.13	'goat'
escargot	|ɛskaʁ'go|	[ka'ko]	1;11.13	'snail'
chaussure	|ʃo'syʁ|	[ly'ly][5]	1;11.28	'shoe'

b. *Continuancy-disharmonic onsets:*
Truncation of the initial syllable

cassé	|ka'se|	['le]	1;11.13	'broken'
tennis	|te'nis|	['ni]	1;11.02	'tennis'
château	|ʃa'to|	['to]	1;11.02	'castle'
jumeaux	|ʒy'mo|	['mo]	1;11.28	'twins'

An understanding of these data immediately calls for an investigation of the child's attempts at other target word forms. In this regard, dos Santos (2007) also shows that during the same developmental period, all consonants are systematically produced when they appear in CVC forms, as we can see in (6). Interestingly, target words with continuancy-disharmonic consonants display harmony in this context, as opposed to deletion (cf. (5b)).

(6) CVC *targets: Consonant preservation and harmony* (dos Santos 2007)

bottes	|'bɔt|	['bɔt]	2;00.12	'boots'
coupe	|'kup|	['kup]	2;00.25	'cut'
passe	|'pas|	['pat]	1;11.13	'pass'
case	|'kaz|	['kak][6]	1;10.17	'box'

As we can see from these latter examples, the shapes of Marilyn's outputs are constrained by a combination of segmental and prosodic pressures, namely the types of consonants (stop *vs.* continuant) that are found in the target forms and the shape of the words in which the consonants appear (multisyllabic *vs.* CVC). Similar to Clara's data described in §2.1, the examples from Marilyn's productions illustrate how child phonological patterning often involves interactions between segmental and prosodic aspects of the developing phonological system.[7]

Partly because of their striking properties, but also because of the theoretical challenges they pose, exotic patterns like those in (4) have given rise to a rich and often contentious literature on child language development (e.g. Menn 1971; Smith 1973; Braine 1974, 1976; Ingram 1974; Ferguson and Farwell 1975; Priestly 1977; Chiat 1983, 1989; Levelt 1994; Velleman 1996; Pater 1997; Goad 2001;

[4] Here and elsewhere, ages are indicated using the Y;MM.DD format.

[5] While this example may suggest syllable reduplication, it is actually symmetrical with the other examples. Related to this is the fact that target fricatives that do not undergo deletion (or manner harmony; cf. examples in (6)) are pronounced as [l] at this stage. See dos Santos (2007) for additional discussion.

[6] The place harmony seen in this example arises through an independent process, as the child does not produce words that combine coronal and velar articulations at this stage. See dos Santos (2007) for additional discussion.

[7] For further discussion of segmental-prosodic interactions, see also contributions to Goad and Rose (2003) and to Kager *et al.* (2004), as well as Inkelas and Rose (2008) and McAllister (2009).

Gnanadesikan 2004; Menn 2004; dos Santos 2007; Vihman and Croft 2007; Fikkert and Levelt 2008; Inkelas and Rose 2008).[8] In the next section we discuss some of the fundamental questions relating to the interpretation of child phonological patterns, including exotic ones, and the formal apparatus needed to account for them. The discussion revolves around the degree of abstractness required to account for these phenomena.[9]

3 How abstract is child phonology?

The issue of abstractness in phonology has been – and continues to be – hotly debated in phonological theory, at least since Kiparsky (1968), and resides at the heart of theoretical controversy about child language development. Central questions, many of them pertaining to the degree of representational abstraction required to model children's grammars, emerge from a general lack of consensus about the phonetic or phonological level at which children operate from the onset of acquisition and throughout the developmental period. For example, phonologists often debate whether discrepancies between child and adult phonology are due to perception-based issues affecting representation, differences in grammar, or lack of motor control over production (see CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY).

While such questions are complex and cannot be resolved within the confines of this chapter, it seems clear that, just as adult phonological systems involve many different simultaneous levels of representation, children make generalizations about various kinds of representational units, depending on the types of evidence they are exposed to and able to interpret from the ambient language. Addressing this general observation, the field has gradually shifted from universalist claims about the unfolding of child phonology to a more evidence-based, individualist approach. Extreme examples of these two viewpoints can be found in Jakobson (1941) and Vihman and Croft (2007). Jakobson, on the one hand, famously argues for a maturational path of acquisition, which proceeds from one phonological feature to the next in universal fashion. On the other hand, Vihman and Croft argue that order of acquisition is determined by salient patterns in the ambient data, and can vary by individual as well as by language. A number of recent studies of infant speech perception, informed by theoretical models of categorization, have turned up support for the view that phonological structure and specificational detail emerge gradually within the lexicon. On this view, the child develops a phonetically detailed

[8] See Goad (forthcoming) for a recent summary of other theoretical issues.

[9] In this context, we must also mention the conspicuous absence of patterns affecting vowels from most discussions of phonological development (e.g. in (1)). Two main facts contribute to this asymmetry. First is a general impression, probably wrong, that children tend to produce vowels more accurately than consonants. This impression may be due in part to the lack, until very recently, of high-quality recordings, and in part to difficulties inherent to the acoustic analysis of child language vowel productions, especially with regard to place of articulation. Fortunately, the democratization of digital recording systems, the increased availability of databases documenting child language productions (especially through the Phon & PhonBank initiative; e.g. MacWhinney and Rose 2008), and the development of methods enabling less equivocal interpretations of child language speech articulations and related acoustics now offer new and exciting research possibilities. These include the consideration of potential covert contrasts and other articulatory artifacts in our characterization of phonological development (e.g. Buder 1996; Scobbie *et al.* 1996; van der Stelt *et al.* 2005; Vorperian and Kent 2007).

lexicon using inborn linguistic perceptual abilities, which are sharpened by language-specific exposure during the first year of life (see Werker and Fennell 2004 and Yoshida *et al.* 2009 for recent summaries and discussions). Pierrehumbert (2003) proposes that the phonetic detail contained in early lexical forms serves as the primary ingredient for building abstract segmental categories (see also Beckman and Pierrehumbert 2003). As Pierrehumbert (2003: 119) suggests: "the system of phonological categories includes not only segments, but also other types of discrete entities in the phonological grammar, such as tones, syllables, and metrical feet." Following similar reasoning, many acquisitionists have proposed a gradual emergence of phonological categories in the lexicon and established relationships between the emergence of these categories and phonological patterning. We return to these proposals in §5.

The issue of representational abstraction in phonological development also has implications for any theory assuming a set of universal primitives, including those based on strong interpretations of the Continuity Hypothesis, discussed in the next section.

3.1 *Child phonology and the Continuity Hypothesis*

According to the Continuity Hypothesis (e.g. Macnamara 1982; Pinker 1984), the formal properties of the grammar do not change over the course of development. The child's grammar starts with the same theoretical primitives that the adult grammar ends up with. Within the generative framework, this means that child phonology uses the same types of rules or constraints posited for adult systems. Each stage in phonological development is generally assumed to be compatible with the set of principles that regulate adult systems (e.g. Spencer 1986; Fikkert 1994; Levelt 1994; Demuth 1995; Freitas 1997; Goad 2000; Rose 2000; Goad and Rose 2004; Fikkert and Levelt 2008). The Continuity Hypothesis extends beyond grammar to lexical representations as well. For example, within feature-geometric frameworks (e.g. Sagey 1986), accounts of segmental development typically posit the emergence of feature hierarchies on the basis of phonological contrasts (e.g. Brown and Matthews 1993, 1997; Rice and Avery 1995; Dresher 2004; Fikkert 2005; see also Jakobson 1941 for an early discussion on the role of contrast in phonology). In most proposals, even the order of acquisition of the contrasts is fixed by the model, motivated by typological evidence and, at times, theory-centric learnability considerations, and thus falls under the scope of the Continuity Hypothesis. For example, the feature Coronal enjoys a special status in the literature on adult phonology, where it is often claimed to be inherently less complex than Labial or Dorsal (e.g. contributions to Paradis and Prunet 1991; cf. CHAPTER 12: CORONALS). Many acquisitionists have relied on this claim while describing phonological development (e.g. Levelt 1994) or accounting for consonant harmony patterns (e.g. Pater 1997; Fikkert and Levelt 2008; see Fikkert *et al.* 2008 for further discussion).

As the survey of exotic phonological phenomena reveals, however, child phonology sometimes requires phonological rules or constraints that are not independently motivated in adult phonological systems (e.g. Pater 1997 on child-specific constraints driving consonant harmony; see also Levelt and van Oostendorp 2007 in the context of segmental development). Similarly, lexical representations do not always conform to adult-based generalizations. For example, regarding the coronal asymmetry mentioned just above, Rose (2000: 175) points out that different children exhibit different featural behaviors in their productions, within and across target languages. Claims about the universality of phonological constructs and their emergence in acquisition are often at odds with the facts, especially when they are closely inspected.

In more functionalist approaches to phonological development, such as the constructivist approach entertained by Vihman and Croft (2007), the Continuity Hypothesis receives a different interpretation. Within Vihman and Croft's approach, there is no formal distinction between competence and performance. Language acquisition, just like its use, relies on general cognitive mechanisms operating on schematic templates, the nature of which is claimed not to change in any fundamental way over the course of one's life. (See §4.2.2 for additional discussion of Vihman and Croft's proposal.)

Whichever approach is favored, any theory of grammar is still obligated to explain why child phonology is different from adult phonology. We discuss a few of the toughest challenges in the next section.

3.2 *Some empirical challenges*

Exotic patterns or asymmetries in child phonological development, taken at face value, often suggest either rogue grammatical properties (e.g. Buckley 2003; Goad 2006) or, in some extreme cases, formal paradoxes. Such paradoxes are most evident in so-called chain shift patterns illustrated by the famous *puzzle–puddle–pickle* problem defined by Macken (1980), based on data from Amahl, an English-learning child whose productions were originally documented by Smith (1973). This chain shift takes the form of an A → B; B → C schema whereby a given phone arising from a substitution pattern may be the target of another substitution. For example, in Amahl's productions, the /z/ of *puzzle* is realized as [d] (/pʌzɫ̩/ → [pʌdɫ̩]), while /d/ itself surfaces as [g] in words like *puddle* (/pʌdɫ̩/ → [pʌgɫ̩]). Another chain shift, also found in Amahl's productions, involves the realization of word-initial /s/ as [θ] (e.g. *sick* /sɪk/ → [θɪk]), a consonant which, when present in a target form, surfaces as [f] (e.g. *thick* /θɪk/ → [fɪk]). We discuss the former chain shift in more depth in §4.[10] Such patterns are, at least in appearance, problematic for all theories of phonology in the sense that if a rule or constraint triggers the substitution of a target sound by another one, this target sound should logically be ruled out altogether from the child's output forms, at least within similar prosodic contexts.[11]

Other patterns are problematic because they contradict generalizations made from the study of adult languages. The case of positional velar fronting offers such an example. Velar fronting is a pattern whereby target velars /k g/ are fronted to coronals [t d]. Context-free (across-the-board) velar fronting can be attributed to the child's inability to either lexically represent or correctly articulate target velars. Neither of these potential analyses poses a difficult theoretical challenge beyond data interpretation per se. However, the positional version of velar fronting, whereby velars are neutralized to coronals in strong, but not in weak,

[10] Additional discussions of the latter chain shift can be found in Hale and Reiss (1998) and Rose (2009).

[11] Chain shifts do occur in adult language as well. However, they tend to involve lenition or vowel quality, not major consonantal place of articulation (see e.g. McCarthy 1999, 2008 and Mortensen 2006 for recent discussions).

prosodic environments (e.g. *go* /go/ → [do]; *bagel* /'beɪgl̩/ → ['beɪgu]), is problematic. As Inkelas and Rose (2008) point out, theories of positional asymmetries in phonology share the prediction that neutralization of segmental contrasts should occur in prosodically weak, as opposed to strong, positions, as generally happens in adult language.[12] Positional velar fronting thus contradicts the generalization that contrasts are preserved in prosodically strong positions, and challenges theories based upon that generalization.

The incidence of innovated chain shifts and apparently rogue patterns such as positional velar fronting in child phonology contradicts the view that human developing and end-state (adult) grammars share the same basic properties. Of course, we at times observe subsystems in adult grammars that exhibit phonological asymmetries or apparently odd properties. Many so-called "crazy" rules in adult language (e.g. Bach and Harms 1972) have their origins in diachronic re-analysis (e.g. rule inversion) or telescoping of sound changes. Arguably, a large proportion of these apparent rules are morphologically conditioned and can be handled via suppletive allomorphy. Opacity, which is not infrequent in adult phonology, has been analyzed synchronically by an appeal to lexical contrast preservation (Łubowicz 2003) or to abstract (underlying) properties of lexical forms, justified by semantically related morphological forms observed elsewhere in the lexicon (e.g. Kager 1999; McCarthy 1999, 2002, 2008; and Kiparsky 2000 for related discussions). Once put in their appropriate contexts, apparently rogue adult phonological subsystems can be reconciled with existing phonological theory and, in many ways, contribute to its elaboration.

In sum, morphological and lexical factors contributing to opacity and other phonological oddities in adult phonology are not directly applicable to most innovative patterns observed in child phonology. Recent literature, to which we turn next, suggests that, in addition to grammar itself, phonetic dimensions of speech perception and articulation play an important role in the exotic as well as in the ordinary, in children's phonological productions.

4 Theoretical approaches to child phonological productions

In this section, we survey a number of theoretical approaches to address empirical problems posed by child phonological productions. The sheer number of proposals available in the literature makes it impossible to provide a comprehensive survey of theoretical approaches to child language phonology. We thus restrict ourselves to contrasting representative examples of predominant approaches, and addressing related controversies. We then discuss how these approaches relate to the larger context of perceptual, lexical, and articulatory development.

[12] The extensive survey of consonantal place markedness in de Lacy (2002) upholds the generalization that contrast neutralization in prosodically weak environments – such as syllable codas or unstressed syllables – is frequent, while it is infrequent or unattested in strong positions, such as syllable onsets or stressed syllables. A noticeable counterexample to this generalization comes from Steriade (2001), who argues that "strong" must sometimes be perceptually defined; neutralization of apical and retroflex consonants, for example, can occur in word-initial onsets but not word-finally, because postvocalic position is perceptually optimal for these consonants. See also CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION.

4.1 Approaches assuming adult-like phonological representations

Since the influential study by Smith (1973), which set the tone for a vast program of research on phonological development within generative phonology and beyond, it has been largely assumed that the child's input is similar to the corresponding adult output form, as stated in (7). In analyses taking this assumption as their starting point, the child's pronounced form is generally derived from the adult form through ordered rules, parameter settings, or constraint rankings.

(7) *A "standard" assumption* (since Smith 1973)

Children's lexical representations are similar to corresponding adult phonetic forms.

This assumption is found in virtually all works couched in derivational versions of Generative Phonology, from standard *SPE* (Chomsky and Halle 1968), to autosegmental and prosodic approaches to phonology (e.g. Goldsmith 1976 and McCarthy and Prince 1995a, respectively). This assumption also holds in flavors of the non-derivational framework of Optimality Theory (henceforth OT; Prince and Smolensky 1993) that draw on formal relationships between inputs and outputs (e.g. Correspondence Theory; McCarthy and Prince 1995b). Common to these various models is the view that milestones in phonological development correspond to the relaxing of formal pressures on the child's grammar that militate against the production of complex phonological units contained in the input. For example, accounts of the acquisition of prosodic representations generally predict that children start with monosyllabic words displaying V or CV syllable structure and gradually expand the number and/or the complexity of the constituents allowed by their grammars (e.g. Spencer 1986; Fikkert 1994; Freitas 1997). The same general approach within OT is formally viewed as an initial ranking of markedness constraints over faithfulness constraints (see CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). The effect of the latter is typically evaluated against adult-like output forms, in accordance with the assumption in (7). The gradual demotion of the markedness constraints accounts for the concomitant appearance of more complex phonological structures in output forms (e.g. Demuth 1995; Barlow 1997; Pater 1997; Rose 2000; Gnanadesikan 2004).

Approaches based on the assumption in (7) raise a number of theoretical issues. Some of the most central questions can be illustrated through a consideration of Smith's (1973) and Dinnsen's (2008) analyses of opacity effects. Let us first consider a representative dataset, from Dinnsen's recent discussion of one of Smith's chain shift patterns we described in §3.2 (see CHAPTER 73: CHAIN SHIFTS).

(8) *Apparent chain shift* (as described by Dinnsen 2008; data from Smith 1973)

a. *Velarization in non-derived environments*

[pʌgl̩]	*puddle*	[æŋklaz]	*antlers*
[bɔkl̩]	*bottle*	[bʌklə]	*butlers*
[hæŋgl̩]	*handle*	[trɔglə]	*troddler*

b. *Velarization blocked in morphologically derived environments*

[kwæːtliː]	*quietly*	cf.	[kwæːt]	*quiet*
[sɔftliː]	*softly*		[sɔft]	*soft*
[haːdliː]	*hardly*		[haːd]	*hard*
[taitliː]	*tightly*		[tait]	*tight*

c. *Velarization blocked in phonologically derived environments (stopped fricatives)*

[pʌdl̩]	*puzzle*
[pɛntl̩]	*pencil*
[wɪtl̩]	*whistle*

In an analysis that reflects generative phonology of its time, Smith (1973) accounts for the set of data in (8) using a series of ordered transformational rules deriving the child outputs from their target counterparts. First, he posits two basic rules, one for velarization (R3) and one for stopping (R24).[13]

A constraint-based analysis of the same data is offered by Dinnsen (2008), who invokes Comparative Markedness (McCarthy 2002), a version of OT in which the original conception of markedness is redefined by splitting each markedness constraint into two distinct versions of itself, one that is violated by a property that is fully faithful to the input (similar to original markedness violations), the other violated by a property that is not present in the input. Dinnsen accounts for the data in (8) through a combination of input–output and output–output markedness constraints targeting sequences composed of alveolar stops and liquids. For example, the velarization pattern in (8a) and the absence thereof in (8b) are formally distinguished through an output-output constraint (OO-$_{O}$*dl), which cannot apply to monomorphemic forms that do not display alveolar–liquid sequences in the input (e.g. *quiet*).[14]

Assessing the full theoretical implications of McCarthy's (2002) theory of Comparative Markedness would go far beyond the scope of this chapter.[15] In the context of child language, the proposal relies on the OT tenets that (a) all constraints are universal and (b) phonological development can be accounted for as reranking of constraints from an initial ranking within which markedness constraints outrank faithfulness constraints (e.g. Smolensky 1996; cf. Hale and Reiss 1998). Dinnsen's (2008) account of the facts in (8) is entirely compatible with these assumptions. As he points out, the opaque chain shift in (8) is not "inferable from the primary linguistic data to which Amahl would have been exposed" (2008: 157). Amahl is nonetheless able to generate his chain shift by ranking the universal constraint penalizing "old" coronals higher than the constraint penalizing derived ones.

Both the rule-based and the constraint-based account of the chain shift in (8) are based on the key presupposition in (7) – shared by most generative approaches to acquisition – that the child's input is equivalent to the adult's output. In addition, Dinnsen's (2008) account must crucially assume that the child has access to morphological structure or, minimally, to grammatical awareness, enabling a distinction between morphologically derived *vs.* non-derived environments.

[13] In Smith's (1973) original formulation, each rule is labeled R*n*, where *n* indicates the relative order of the rule in Amahl's grammar.

[14] See Barlow (2007) for a similar approach to other cases of apparently opaque production patterns.

[15] See Hall (2006) for a discussion of Comparative Markedness in the context of adult phonology.

However, Dinnsen does not provide any independent evidence for the validity of this assumption. As discussed in the next section, alternative accounts can be formulated that do not require any such, potentially unwarranted, assumptions.

4.2 *Alternative views*

A number of challenges have been posed for the assumption in (7), which underlies the rule-based and constraint-based analyses we have just summarized. In this section we review three proposals, each of which offers its own perspective on children's inputs and, consequently, on the phonological system that governs their outputs.

4.2.1 Child phonological productions as theoretically irrelevant

One response to puzzles like the one in (8) is to deny their grammaticality, attributing them to performance factors. Focusing heavily on issues relating to speech articulation in child language, Hale and Reiss (1998) offer a rather provocative solution. Keeping with a strong version of the Continuity Hypothesis, they use a series of empirical observations to undermine the theoretical relevance of production-based studies for phonological theory. They argue that "[child language] deviations from targets are largely due to performance effects" (1998: 658), comparing child speech to "the intoxicated speech of the captain of the Exxon Valdez around the time of the accident at Prince William Sound, Alaska" (1998: 669).

Since its publication, Hale and Reiss's article has been the subject of much controversy in the literature, with the positive result of encouraging reflection on how phonological productions in child language should be approached. The general consensus that has emerged is that while there is no doubt that surface effects (intoxication, immature vocal apparatus) add noise that can obscure patterns in the data, this observation does not entail that we should throw the phonological baby out with the bathwater. The study of phonological development based on production data, if conducted in a careful manner, is worthy of empirical and theoretical investigation

4.2.2 Deriving the input from output considerations

Macken (1980), inspired by Braine's (1976) review of Smith (1973), sheds important light on differences between child and adult phonology. Macken argues that equating children's inputs to adults' outputs as per assumption (7) may give rise to misanalyses of some child-language data. She demonstrates that a reconsideration of this assumption can make child phonological systems much more transparent than they seem at first sight. In the case of the chain shift in (8), Macken argues that it is merely apparent, in that it can be decomposed into a series of simpler problems, each of which primarily involves either perception or production. According to Macken, Amahl's perception of coronal stops was influenced by the relative velarity of syllabic [ɫ] in the contexts in (8a).[16] Words like *puddle* were thus represented with a velar consonant (i.e. /g/) in Amahl's early lexicon. In (8b), the absence of a syllabic lateral after the coronal stop explains in a simple way the absence of velarization of the preceding stop, without any need to refer to

[16] This account also implies that laterals in words such as *antlers* also displayed a degree of velarity. Another potential influence in this context is the glottalization of the coronal preceding the lateral.

morphological complexity. Finally, in (8c), the consonant preceding the [ɫ] is a fricative, whose continuancy and/or stridency arguably prevented the type of velar influence that affected coronal stops in (8a). These fricatives, however, underwent stopping in onsets, a pattern widely and independently attested in the literature on child language (e.g. Ingram 1989; Bernhardt and Stemberger 1998).

In the face of this explanation of Amahl's apparently complicated production patterns, which appeals to relatively basic perception and production factors, one could be tempted to drop accounts based on grammatical processing altogether. This is in many ways the view espoused by Vihman and Croft (2007), who propose the constructivist, *Radical Templatic* model of phonology and phonological development. Within this model, the word is represented as a phonotactic template that directly encodes all melodic and prosodic characteristics of output (phonetic) forms. With regard to acquisition, Vihman and Croft's proposal can be divided into two main claims. First, phonological learning proceeds through implicit inferences based on memorized word forms:

> The child gradually develops first one or a small number of phonological templates, then a wider variety of them, while at the same time inducing a range of other phonological categories and structures from the known word shapes. (Vihman and Croft 2007: 686)

Second, template specification is constrained by the child's own productive abilities: children gradually specify their template as they discover how to reproduce the types of phonotactics that are present in their memorized target forms.

Vihman and Croft's (2007) view that the shape of children's lexical templates is governed by their own productive abilities presents, at the theoretical level, a circularity problem, in the sense that both the phenomena observed and their origins are the same. While this may be seen as a virtue in what-you-see-is-what-you-get constructivist approaches to grammar, it poses a challenge to approaches that formally separate grammatical processing from lexical representation.

Outside of theoretical considerations, it is undeniable that articulatory planning and execution are involved in speech production and, as such, may influence the learner's productions, especially at young ages. Vihman and Croft's (2007) proposal nicely highlights this fact. However, it is unclear whether all speech errors observed in child language should be related to the domain of production. For example, children are often not influenced by their own production mistakes during speech perception. This is demonstrated by Chiat (1983), who documents perceptual and productive abilities of English-learning Stephen. This child displays systematic velar fronting patterns, on the one hand, but shows, on the other, perfect discrimination abilities between alveolars and velars. Stephen's system thus illustrates the formal separation that must be made between perceptual and productive abilities. It also highlights, once again, the importance of considering the child's overall system in any investigation of developmental speech patterns.

4.2.3 Deriving outputs from statistical influences

Without addressing apparent phonological paradoxes such as the one illustrated in (8), but in a move away from intrinsic linguistic conditioning, Levelt *et al.* (2000) turn the focus to statistical pressures from the ambient language (see CHAPTER 90: FREQUENCY EFFECTS). They argue that the order of acquisition of syllable types (e.g. CV > CVC > CCV) in monolingual Dutch-learning children can be predicted through the relative frequency of occurrence of these syllable types in the ambient language. Levelt *et al.* further argue that the variability observed across groups of learners with regard to the acquisition of certain syllable types is also correlated with frequency figures from the adult language. Their argument thus highlights frequency as a potentially determining influence in child language development, an observation that matches many of the findings about infants' abilities to statistically process the ambient signal (e.g. Jusczyk 1997 for a summary).

However, further verifications of the hypothesis that frequency drives order of acquisition or the emergence of patterns of speech production in child language have yielded a series of criticisms from both empirical and theoretical perspectives (e.g. Kehoe and Lleó 2003; Demuth 2007; Edwards and Beckman 2008; Rose 2009; see also Brown 1973 for early refutations of statistical approaches). Critiques of the statistical approach generally contend that while input statistics can play a role in the emergence of phonological patterns in language development, and while it is important to look at properties of the ambient language as a whole, both linguistic and non-linguistic, frequency is only one of the several factors that can affect the developing phonological system and its outcomes in child spoken forms. Clearly, no frequency-based explanation can account for the chain shift exemplified in (8) or other patterns, such as consonant harmony or velar fronting.

4.3 *Interim discussion*

The approaches briefly addressed above far from exhaust the range of proposals available in the literature. Nonetheless, they express relatively clear views of child phonology, each of which highlights crucial areas of consideration about child phonological data. Each of these proposals, however, faces a number of relatively similar challenges, especially since they often neglect to situate the patterns observed in their larger context, that of an emerging system influenced by a variety of independent factors, whose combined effects may at times yield phonologically unexpected, yet entirely logical outcomes (Rose 2009). For example, Hale and Reiss's (1998) general argument against the validity of child language production data for theoretical investigation is based on a *prima facie* interpretation of these productions, without much consideration for the grammatical factors that may contribute to them.

While the most central object of study in both phonology and phonological development should evidently be the grammatical system, this system is connected to a series of perception- and production-related mechanisms, each of which has a potential influence on the shape of the developing lexicon and its manifestation in child speech. In light of this, we suggest that the ideal approach to phonological development should encompass all relevant considerations, a number of which are discussed in the next section.

5 Building a path between speech perception and phonological productions

During the first decade of the 21st century, some of the patterns observed in child phonological productions have been reconsidered in their larger context.

For example, thanks to recent advances in research on infant speech perception, we now know more about the shape of the emerging lexicon (e.g. Yoshida *et al.* 2009 for recent developments) and how pressures from the ambient language may impact productive abilities (e.g. Edwards *et al.* 2004). The variability observed within and across developmental paths has been assessed from learnability and grammatical perspectives (e.g. Dresher 1999 as well as contributions to Goad and Rose 2003 and to Kager *et al.* 2004). A positive result from this literature is the fact that grammatical explanations for some apparently puzzling patterns observed in children's productions now incorporate elements from both speech perception and production. In sum, the grammar no longer bears sole responsibility for differences between child and adult phonology, and child language productions can be better reconciled with phonological theories based on patterns in adult language.

5.1 *The development of lexical representations*

The question of the emergence of lexical representations and how they relate to the remainder of the phonological grammar is at the center of many current questions. A review of the recent literature on this topic reveals a promising convergence of interests between theoretical and experimental linguists and psychologists (e.g. Pater *et al.* 2004; Kager *et al.* 2007; Yoshida *et al.* 2009). Recent highlights from this literature strongly suggest that the child's early lexicon is phonetically detailed but lacks phonological sophistication. Related models of categorization also suggest that segmental and prosodic categories emerge from implicit computations of the types of phonetic distributions that appear in the early "phonetic" lexicon (e.g. Pierrehumbert 2003; Werker and Curtin 2005).

These proposals support Macken's (1980) contention that early input representations may contain artifacts of the child's misunderstanding of some of the phonotactics that exist in the adult language. Similar positions have been taken by Locke (1983), Menn (1983), Waterson (1987), Levelt (1994), and Vihman and Croft (2007), amongst others. Within the generative framework, Goad and Rose (2001, 2004) and Fikkert and Levelt (2008) establish explicit relationships between the child's developing phonological representations and the types of error patterns observed in early phonological productions. For example, Goad and Rose (2001, 2004) propose that prosodic representations gradually emerge in the lexicon through the child's implicit analysis of the distributional evidence available from the words stored within the (still developing) lexicon. While obstruent + sonorant onsets (e.g. [pr tr kr]) display clearly rising sonority profiles, this is not the case for s + consonant onsets, which show different sonority profiles, including flat (or, arguably, falling) sonority s + obstruent clusters (e.g. [sp st sk]; cf. [sl sr sw]). Goad and Rose propose that children may be temporarily misled by superficial aspects of the evidence, and thus develop early syllabic representations on the basis of sonority until the distributional facts are understood, specifically about the types of phonotactics that govern the appearance of s + consonant clusters in these languages.[17]

[17] From a broader perspective, these proposals also support the view that prosodic structure is specified in the developing lexicon (e.g. Colston 1996: 718ff. and references therein).

5.2 *The expression of phonological categories in child language*

Beyond grammatical development, the child's productions are also subject to factors pertaining to physiological growth and motor control, the effects of which must also be considered in any interpretation of child language production data.

Productive abilities figure prominently in works focusing on the transition between late babbling and early word productions. For example, Kern and Davis (2009) show that children learning different languages gradually attune their general productive abilities to the types of consonant and vowel combinations that are the most prominent in the ambient language. This literature thus emphasizes the child's progressive gain in control of his/her speech articulations. Also focusing on articulatory factors, Vihman and Croft (2007) highlight the fact that children's usage of early words and their pronounced forms are generally constrained by motor limitations that are gradually overcome during the course of development. These proposals, based on children's babbling and earliest word productions, suggest that basic biomechanical pressures that hinder early phonological production gradually give way to the types of articulations that are required to reproduce faithfully the full range of phonotactics found in the target language.

Data from slightly older learners, however, suggests that the link between lexical representations and related word productions is generally less direct. For example, in their explanation of positional velar fronting, Inkelas and Rose (2008) propose an interaction between grammatical and articulatory factors. Positional velar fronting consists of the pronunciation of velar consonants as coronal in prosodically strong positions (e.g. in word-initial or otherwise stressed onsets), but not in weak positions (e.g. non-initial onsets of unstressed syllables; codas), as exemplified in (9) (see also Chiat 1983; Stoel-Gammon 1996; McAllister 2009; see §4.2.2 above).

(9) *Positional velar fronting* (Inkelas and Rose 2008)

a. *Prosodically strong onsets*

['tʰʌp]	*cup*	1;09.23
['doː]	*go*	1;10.01
['hɛksəˌdɔn]	*hexagon*	2;02.22

b. *Prosodically weak onsets; codas*

['maɪŋki]	*monkey*	1;08.10
['bejgu]	*bagel*	1;09.23
['pædjɔk]	*padlock*	2;04.09

As already mentioned in §3.2, Inkelas and Rose (2008) point out that positional velar fronting is unexpected from a theoretical standpoint, since positional segmental neutralization in adult phonology generally occurs in prosodically weak, rather than strong, positions. In an analysis that reconciles the pattern with phonological theory, Inkelas and Rose argue that positional velar fronting derives from an interaction between the child's developing grammar and grammar-external, articulatory factors. They propose that children who display positional velar fronting are in fact attempting to produce stronger articulations in prosodically strong contexts. However, because of a combination of physiological and motor factors (proportionally larger tongue body, shorter palate, limited control of tongue articulations), the strengthening of target velars results in an articulation that extends too far forward, into the coronal area of the hard palate, yielding a fronted velar release.[18]

Inkelas and Rose's (2008) proposal potentially extends to the analysis of other apparently problematic positional substitutions. For example, in the pattern of positional stopping illustrated in (10), target fricatives undergo stopping in prosodically strong but not weak positions.

(10) *Positional stopping* (Chiat 1989; Rvachew and Andrews 2002; Marshall and Chiat 2003)

a. *Stopping in prosodically strong onsets*

[bɪˈp̱ɔ]	'before'
[ˈḏu]	'zoo'
[kəˈṯino]	'casino'

b. *No stopping in other (prosodically weak) positions*

[ˈpɜːs̱ən]	'person'
[ˈmɪs̱]	'miss'
[ˈkev̱]	'cave'

While positional stopping may suggest a fairly heavy articulatory component (strengthening of fricative articulations in prosodically strong positions), it also requires a grammatical reference to stress and word edges, in a way similar to that of positional velar fronting.

Grammatical influences, whether they relate to lexical representations or their expression in spoken forms, are also evidenced in long-distance patterns such as consonant harmony (Goad 1997, 2001; Pater 1997; Rose 2000, 2002; dos Santos 2007; Fikkert and Levelt 2008; CHAPTER 72: CONSONANT HARMONY IN CHILD LANGUAGE) and metathesis (e.g. Ingram 1974; Menn 1976; Macken 1996; Velleman 1996; Rose 2000, 2002; dos Santos 2007). For example, in cases of long-distance metathesis, consonantal place or manner features are preserved but swapped across vowels in systematic ways. This is illustrated in (11) by the productions of W, a learner of English, who produces every word-initial target fricative in word-final position (original data from Leonard and McGregor 1991, as reported by Velleman 1996).

(11) *Manner-conditioned metathesis patterns* (Leonard and McGregor 1991)

[uẕ]	'zoo'
[aɪnf̱]	'fine'
[ops̱]	'soap'
[nupis̱]	'Snoopy'
[taps̱]	'stop'

[18] Note that this analysis still applies whether the pattern is fully neutralizing or not (e.g. Edwards *et al.* 1997), as we are in both cases witnessing a difference in articulation that is prosodically driven. See also McAllister (2009) for additional arguments in favor of Inkelas and Rose's (2008) original proposal.

Systematic, word-level distributional patterns of this kind demand a grammatical analysis that transcends perceptual or articulatory factors.[19]

Finally, recent analyses of early word forms by Levelt and van Oostendorp (2007) and Fikkert and Levelt (2008) focus on the emergence of phonological features in the lexicon as well as their interaction in early word productions, which they claim to be responsible for other theoretically intriguing patterns such as consonant harmony. From a formal perspective, this research offers new and interesting ways to provide a bridge between the emergence of phonological representations and their expression in children's early productions.

6 Conclusion

Phonological patterns observed in child language offer a central source of evidence for our understanding of phonology as a grammatical system. However, careful analysis is required, as these patterns can be triggered by a number of potentially conspiring factors, be they perceptual, grammatical, or articulatory, the combination of which may elude limitations imposed by approaches that are theoretically or empirically too narrow. Recent experimental advances, combined with a growing body of scientific literature on child phonological production, show that phonological categories familiar from the literature on adult phonology cannot be taken for granted in the investigation of child language productions, especially at early stages in phonological development. Any analysis of child phonology must therefore question all properties of the child's target language that may affect development as well as consider all of the factors that might influence production throughout the relevant developmental period. This is a complicated challenge for practitioners as well as for the elaboration of theoretical models. However, theoretical simplicity cannot take precedence over the more central consideration of explanatory adequacy.

Several topics have been left aside in the above discussion, including disordered or protracted phonological systems. The characterization of these systems poses its own challenges, despite the fact that many of the patterns discussed above are also attested in clinical data (e.g. Bernhardt and Stemberger 1998 and contributions to Dinnsen and Gierut 2008). One central debate in this context pertains to whether the difference between these and typically developing child phonologies is qualitative or quantitative. While this debate lies outside the scope of this chapter, we contend that a method in which perception, grammar, and production are integrated into the explanation is the best one for all human phonological systems, no matter their characteristics or degree of development.

ACKNOWLEDGMENTS

Several aspects of our work have benefited from discussions with other scholars, especially Heather Goad, François Pellegrino, Sophie Kern, and Marilyn Vihman, as well as from

[19] One could object that [s] is hard to locate temporally in the signal and therefore is particularly prone to perceptual "migration." However, such an observation would not predict why [s] systematically goes to the end of the word, rather than to any other location within the word.

useful comments from two anonymous reviewers, to whom we are indebted. We also want to thank the *Companion*'s editors for their invitation to contribute, as well as for their feedback and support at all stages in the making of this chapter. Of course, all remaining errors or omissions are our own.

REFERENCES

Bach, Emmon & Robert T. Harms. 1972. How do languages get crazy rules? In Robert P. Stockwell & Ronald K. S. Macaulay (eds.) *Linguistic change and generative theory*, 1–21. Bloomington: Indiana University Press.

Barlow, Jessica. 1997. A constraint-based account of syllable onsets: Evidence from developing systems. Ph.D. dissertation, Indiana University.

Barlow, Jessica. 2007. Grandfather effects: A longitudinal case study of the phonological acquisition of intervocalic consonants in English. *Language Acquisition* 14. 121–164.

Beckman, Mary E. & Janet B. Pierrehumbert. 2003. Interpreting "phonetic interpretation" over the lexicon. In John Local, Richard Ogden & Rosalind Temple (eds.) *Phonetic interpretation: Papers in laboratory phonology VI*, 13–38. Cambridge: Cambridge University Press.

Bernhardt, Barbara H. & Joseph P. Stemberger. 1998. *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.

Braine, Martin D. S. 1974. On what might constitute learnable phonology. *Language* 50. 270–299.

Braine, Martin D. S. 1976. Review of Smith (1973). *Language* 52. 489–498.

Brown, Cynthia A. & John Matthews. 1993. The acquisition of segmental structure. *McGill Working Papers in Linguistics* 9. 46–76.

Brown, Cynthia A. & John Matthews. 1997. The role of feature geometry in the development of phonemic contrasts. In Hannahs & Young-Scholten (1997), 67–112.

Brown, Roger. 1973. *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Buckley, Eugene. 2003. Children's unnatural phonology. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 29. 523–534.

Buder, Eugene H. 1996. Experimental phonology with acoustic phonetic methods: Formant measures from child speech. In Barbara H. Bernhardt, John Gilbert & David Ingram (eds.) *Proceedings of the UBC International Conference on Phonological Acquisition*, 254–265. Somerville, MA: Cascadilla Press.

Chiat, Shulamuth. 1983. Why Mikey's right and my key's wrong: The significance of stress and word boundaries in a child's output system. *Cognition* 14. 275–300.

Chiat, Shulamuth. 1989. The relation between prosodic structure, syllabification and segmental realization: Evidence from a child with fricative stopping. *Clinical Linguistics and Phonetics* 3. 223–242.

Chitoran, Ioana, François Pellegrino & Egidio Marsico (eds.) 2009. *Approaches to phonological complexity*, 353–375. Berlin & New York: Mouton de Gruyter.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

de Lacy, Paul. 2002. The formal expression of markedness. Ph.D. dissertation, University of Massachusetts, Amherst.

Demuth, Katherine E. 1995. Markedness and the development of prosodic structure. *Papers from the Annual Meeting of the North East Linguistic Society* 25(2). 13–25.

Demuth, Katherine E. 2007. The role of frequency in language acquisition. In Insa Gülzow & Natalia Gagarina (eds.) *Frequency effects in language acquisition*, 528–538. Berlin & New York: Mouton de Gruyter.

Dinnsen, Daniel A. 2008. A typology of opacity effects in acquisition. In Dinnsen & Gierut (2008), 121–176.
Dinnsen, Daniel A. & Judith A. Gierut (eds.) 2008. *Optimality Theory: Phonological acquisition and disorders*. London: Equinox.
Dresher, B. Elan. 1999. Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30. 27–67.
Dresher, B. Elan. 2004. On the acquisition of phonological contrasts. In Jacqueline van Kampen & Sergio Baauw (eds.) *Proceedings of GALA 2003*, 27-46. Utrecht: LOT.
Edwards, Jan & Mary E. Beckman. 2008. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development* 4. 122–156.
Edwards, Jan, Mary E. Beckman & Benjamin Munson. 2004. The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research* 47. 421–436.
Edwards, Jan, Fiona Gibbon & Marios Fourakis. 1997. On discrete changes in the acquisition of the alveolar/stop consonant contrast. *Language and Speech* 40. 203–210.
Ferguson, Charles A. & Carol B. Farwell. 1975. Words and sounds in early language acquisition. *Language* 51. 419–439.
Fikkert, Paula. 1994. On the acquisition of prosodic structure. Ph.D. dissertation, University of Leiden.
Fikkert, Paula. 2005. From phonetic categories to phonological feature specification: Acquiring the European Portuguese vowel system. *Lingue e Linguaggio* 4. 263–280.
Fikkert, Paula & Clara C. Levelt. 2008. How does Place fall into place? The lexicon and emergent constraints in children's developing grammars. In Peter Avery, B. Elan Dresher & Keren Rice (eds.) *Contrast in phonology: Theory, perception, acquisition*, 231–268. Berlin & New York: Mouton de Gruyter.
Fikkert, Paula, Clara C. Levelt & Joost van de Weijer. 2008. Input, intake and phonological development: The case of consonant harmony. Unpublished ms., Radboud University Nijmegen, University of Leiden & University of Lund.
Freitas, Maria João. 1997. Aquisição da estrutura silábica do Português Europeu. Ph.D. dissertation, University of Lisbon.
Gnanadesikan, Amalia E. 2004. Markedness and faithfulness constraints in child phonology. In Kager *et al.* (2004), 73–108.
Goad, Heather. 1997. Consonant harmony in child language: An optimality-theoretic account. In Hannahs & Young-Scholten (1997), 113–142.
Goad, Heather. 2000. Phonological operations in early child phonology. Paper presented at SOAS Linguistics Colloquium Series, University of London.
Goad, Heather. 2001. Assimilation phenomena and initial constraint ranking in early grammars. In Anna H.-J. Do, Laura Dominguez & Aimee Johansen (eds.) *Proceedings of the 25th Annual Boston University Conference on Language Development*, 307–318. Somerville, MA: Cascadilla Press.
Goad, Heather. 2006. Are children's grammars rogue grammars? Glide substitution in branching onsets. *Recherches Linguistiques de Vincennes* 35, 103–132.
Goad, Heather. Forthcoming. Acquisition of phonology. In Patrick C. Hogan (ed.) *The Cambridge encyclopedia of the language sciences*. Cambridge: Cambridge University Press.
Goad, Heather & Yvan Rose. 2001. Headedness versus sonority in cluster reduction. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 36. 109–123.
Goad, Heather & Yvan Rose (eds.) 2003. *Segmental-prosodic interaction in phonological development: A comparative investigation*. Special issue, *Canadian Journal of Linguistics* 48. 139–452.
Goad, Heather & Yvan Rose. 2004. Input elaboration, head faithfulness, and evidence for representation in the acquisition of left-edge clusters in West Germanic. In Kager *et al.* (2004), 109–157.

Goldsmith, John A. 1976. An overview of autosegmental phonology. *Linguistic Analysis* 2. 23–68.

Golston, Chris. 1996. Direct Optimality Theory: Representation as pure markedness. *Language* 72. 713–748.

Hale, Mark & Charles Reiss. 1998. Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry* 29. 656–683.

Hall, T. A. 2006. Derived environment blocking effects in Optimality Theory. *Natural Language and Linguistic Theory* 24. 803–856.

Hannahs, S. J. & Martha Young-Scholten (eds.) 1997. *Focus on phonological acquisition.* Amsterdam & Philadelphia: John Benjamins.

Hansson, Gunnar Ólafur. 2001. Theoretical and typological issues in consonant harmony. Ph.D. dissertation, University of California, Berkeley.

Ingram, David. 1974. Phonological rules in young children. *Journal of Child Language* 1. 49–64.

Ingram, David. 1989. *First language acquisition: Method, description, and explanation.* Cambridge: Cambridge University Press.

Inkelas, Sharon. 2003. J's rhymes: A longitudinal case study of language play. *Journal of Child Language* 30. 557–581.

Inkelas, Sharon & Yvan Rose. 2008. Positional neutralization: A case study from child language. *Language* 83. 707–736.

Jakobson, Roman. 1941. *Kindersprache, Aphasie, und allgemeine Lautgesetze.* Uppsala: Almqvist & Wiksell. Translated (1968) by A. R. Keiler as *Child language, aphasia and phonological universals.* The Hague & Paris: Mouton.

Jusczyk, Peter W. 1997. *The discovery of spoken language.* Cambridge, MA: MIT Press.

Jusczyk, Peter W., Derek M. Houston & Mary Newsome. 1999. The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* 39. 159–207.

Kager, René. 1999. *Optimality Theory.* Cambridge: Cambridge University Press.

Kager, René, Suzanne van der Feest, Paula Fikkert, Annemarie Kerkhoff & Tania S. Zamuner. 2007. Representations of [voice]: Evidence from acquisition. In Jeroen van de Weijer & Erik Jan van der Torre (eds.) *Voicing in Dutch*, 41–80. Amsterdam & Philadelphia: John Benjamins.

Kager, René, Joe Pater & Wim Zonneveld (eds.) 2004. *Constraints in phonological acquisition.* Cambridge: Cambridge University Press.

Kehoe, Margaret. 1997. Stress error patterns in English-speaking children's word productions. *Clinical Linguistics and Phonetics* 11. 389–409.

Kehoe, Margaret. 1998. Support for metrical stress theory in stress acquisition. *Clinical Linguistics and Phonetics* 12. 1–23.

Kehoe, Margaret & Conxita Lleó. 2003. The acquisition of syllable types in monolingual and bilingual German and Spanish children. In Barbara Beachley, Amanda Brown & Francis Conlin (eds.) *Proceedings of the 27th Annual Boston University Conference on Language Development*, 402–413. Somerville, MA: Cascadilla Press.

Kern, Sophie & Barbara Davis. 2009. Emergent complexity in early vocal acquisition: Cross-linguistic comparisons of canonical babbling. In Chitoran *et al.* (2009), 353–375.

Kiparsky, Paul. 1968. How abstract is phonology? Distributed by Indiana University Linguistics Club. Reprinted in Paul Kiparsky (1982). *Explanation in phonology*, 119–163. Dordrecht: Foris.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–365.

Leonard, Laurence B. & Karla K. McGregor. 1991. Unusual phonological patterns and their underlying representations: A case study. *Journal of Child Language* 18. 261–271.

Levelt, Clara C. 1994. On the acquisition of place. Ph.D. dissertation, University of Leiden.

Levelt, Clara C. & Marc van Oostendorp. 2007. Feature co-occurrence constraints in L1 acquisition. In Bettelou Los & Marjo van Koppen (eds.) *Linguistics in the Netherlands 2007*, 162–172. Amsterdam & Philadelphia: John Benjamins.

Levelt, Clara C., Niels O. Schiller & Willem J. Levelt. 2000. The acquisition of syllable types. *Language Acquisition* 8. 237–264.

Locke, John L. 1983. *Phonological acquisition and change*. New York: Academic Press.

Łubowicz, Anna. 2003. Contrast preservation in phonological mappings. Ph.D. dissertation, University of Massachusetts, Amherst.

Macken, Marlys A. 1980. The child's lexical representation: The *"puzzle–puddle–pickle"* evidence. *Journal of Linguistics* 16. 1–17.

Macken, Marlys A. 1996. Prosodic constraints on features. In Bernhardt *et al.* (1996), 159–172.

Macnamara, John. 1982. *Names for things: A study of child language*. Cambridge, MA: Bradford Books/MIT Press.

MacWhinney, Brian & Yvan Rose. 2008. The Phon & PhonBank Initiative within CHILDES. Paper presented at the International Association for the Study of Child Language, Edinburgh.

Marshall, Chloe & Shulamuth Chiat. 2003. A foot domain account of prosodically-conditioned substitutions. *Clinical Linguistics and Phonetics* 17. 645–657.

McAllister, Tara. 2009. The articulatory basis of positional asymmetries in phonological acquisition. Ph.D. dissertation, MIT.

McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16. 331–399.

McCarthy, John J. 2002. Comparative markedness. In Angela C. Carpenter, Andries W. Coetzee & Paul de Lacy (eds.) *Papers in Optimality Theory II*, 171–246. Amherst: GLSA.

McCarthy, John J. 2008. Harmony in harmonic serialism. Unpublished ms., University of Massachusetts, Amherst.

McCarthy, John J. & Alan Prince. 1995a. Prosodic morphology. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 318–366. Cambridge, MA & Oxford: Blackwell.

McCarthy, John J. & Alan Prince. 1995b. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.

McLeod, Sharynne (ed.) 2007. *The international guide to speech acquisition*. Clifton Park, NY: Thomson Delmar Learning.

Menn, Lise. 1971. Phonotactic rules in beginning speech: A study in the development of English discourse. *Lingua* 26. 225–251.

Menn, Lise. 1976. Pattern, control, and contrast in beginning speech: A case study in the development of word form and word function. Ph.D. dissertation, University of Illinois.

Menn, Lise. 1983. Development of articulatory, phonetic, and phonological capabilities. In Brian Butterworth (ed.) *Language production*, vol. 2: *Development, writing, and other language processes*, 1–49. London: Academic Press.

Menn, Lise. 2004. Saving the baby: Making sure that old data survive new theories. In Kager *et al.* (2004), 54–72.

Morgan, James L. 1996. A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language* 35. 666–688.

Mortensen, David R. 2006. Logical and substantive scales in phonology. Ph.D. dissertation, University of California, Berkeley.

Ota, Mitsuhiko. 1999. Phonological theory and the acquisition of prosodic structure: Evidence from child Japanese. Ph.D. dissertation, Georgetown University.

Paradis, Carole & Jean-François Prunet (eds.) 1991. *The special status of coronals: Internal and external evidence*. San Diego: Academic Press.

Pater, Joe. 1997. Minimal violation and phonological development. *Language Acquisition* 6. 201–253.

Pater, Joe. 2004. Bridging the gap between receptive and productive development with minimally violable constraints. In Kager *et al.* (2004), 219–244.

Pater, Joe, Christine Stager & Janet F. Werker. 2004. The perceptual acquisition of phonological contrasts. *Language* 80. 384–402.

Pierrehumbert, Janet B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46. 115–154.

Pinker, Steven. 1984. *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Priestly, Tom M. S. 1977. One idiosyncratic strategy in the acquisition of phonology. *Journal of Child Language* 4. 45–65.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Rice, Keren & Peter Avery. 1995. Variability in a deterministic model of language acquisition: A theory of segmental elaboration. In John Archibald (ed.) *Phonological acquisition and phonological theory*, 23–42. Hillsdale, NJ: Lawrence Erlbaum.

Rose, Yvan. 2000. Headedness and prosodic licensing in the L1 acquisition of phonology. Ph.D. dissertation, McGill University.

Rose, Yvan. 2002. Relations between segmental and prosodic structure in first language acquisition. In Lynn Santelmann, Maaike Verrips & Frank Wijnen (eds.) *The annual review of language acquisition*, vol. 2, 117–155. Amsterdam & Philadelphia: John Benjamins.

Rose, Yvan. 2009. Internal and external influences on child language productions. In Chitoran *et al.* (2009), 329–351.

Rvachew, Susan & Ellen Andrews. 2002. The influence of syllable position in children's production of consonants. *Clinical Linguistics and Phonetics* 16. 183–198.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.

Santos, Christophe dos. 2007. Développement phonologique en français langue maternelle: Une étude de cas. Ph.D. dissertation, Université Lumière Lyon 2.

Scobbie, James M., Fiona Gibbon, W. J. Hardcastle & Paul Fletcher. 1996. Covert contrast as a stage in the acquisition of phonetics and phonology. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, 194–207. Cambridge: Cambridge University Press.

Smit, Ann Bosma. 1993. Phonologic error distribution in the Iowa–Nebraska articulation norms project: Consonant singletons. *Journal of Speech and Hearing Research* 36. 533–547.

Smith, Neil V. 1973. *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.

Smolensky, Paul. 1996. On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27. 720–731.

Spencer, Andrew. 1986. Towards a theory of phonological development. *Lingua* 68. 3–38.

Stampe, David. 1969. The acquisition of phonetic representation. In *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 5. 433–444.

Stelt, Jeannette van der, Krisztina Zajdó & Ton C. Wempe. 2005. Exploring the acoustic vowel space in two-year-old children: Results for Dutch and Hungarian. *Speech Communication* 47. 143–159.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 219–250. San Diego: Academic Press.

Stoel-Gammon, Carol. 1996. On the acquisition of velars in English. In Bernhardt *et al.* (1996), 201–214.

Velleman, Shelley L. 1996. Metathesis highlights feature-by-position constraints. In Bernhardt *et al.* (1996), 173–186.

Vihman, Marilyn M. 1996. *Phonological development: The origins of language in the child.* Cambridge, MA & Oxford: Blackwell.

Vihman, Marilyn M. & William Croft. 2007. Phonological development: Toward a "radical" templatic phonology. *Linguistics* 45. 683–725.

Vorperian, Houri K. & Ray D. Kent. 2007. Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research* 50. 1510–1545.

Waterson, Natalie. 1987. *Prosodic phonology: The theory and its application to language acquisition and speech processing*. Newcastle upon Tyne: Grevatt & Grevatt.

Werker, Janet F. & Suzanne Curtin. 2005. PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development* 1. 197–234.

Werker, Janet F. & Christopher T. Fennell. 2004. Listening to sounds versus listening to words: Early steps in word learning. In D. Geoffrey Hall & Sandra R. Waxman (eds.) *Weaving a lexicon*. Cambridge, MA: MIT Press.

Yoshida, Katherine A., Christopher T. Fennell, Daniel Swingley & Janet F. Werker. 2009. Fourteen-month-old infants learn similar-sounding words. *Developmental Science* 12. 412–418.

102 Category-specific Effects

Jennifer L. Smith

1 Introduction

The lexical categories *noun*, *verb*, and *adjective* are traditionally distinguished by means of distributional (syntactic) and morphological criteria. But in some languages, lexical categories also have distinct phonological behavior. This point has been discussed by Cohen (1964), Postal (1968), Kenstowicz and Kisseberth (1977), Smith (1997, 2001), Myers (2000), and Bobaljik (2008), among others. However, the cross-linguistic typology of phonological differences among lexical categories has not received much systematic investigation.

This chapter surveys category-specific phonological effects, identifying generalizations or patterns where possible. Noun/verb differences are discussed in §2, and the behavior of adjectives is examined in §3. Alternatives to allowing the phonological grammar to refer to category are discussed in §4, but none of them captures all category-specifiic effects.

2 Nouns *vs.* verbs

This section examines the basic distinction between nouns and verbs. (Adjectives, which seem to pattern as an intermediate category, are discussed in §3.)

The examples are organized according to which category shows greater phonological privilege: nouns (§2.1, §2.2), verbs (§2.3), or neither (§2.4). Here, *phonological privilege* is understood to mean the ability to support a greater array of phonological contrasts (CHAPTER 2: CONTRAST), whether this is manifested as a larger number of underlying distinctions, more variety in surface patterns, or a greater resistance to assimilation or other phonological processes (though see §2.2 for additional considerations). Within each section, examples are classified by type of phonological phenomenon. To preview the results (§2.5), noun privilege is the most common pattern, with a few cases each of verb privilege or distinct noun and verb requirements. Prosodic and suprasegmental phenomena are much more common than segmental or featural phenomena.

2.1 Phonological privilege in nouns

In the examples discussed here, nouns show greater phonological privilege than verbs. The phenomena range over suprasegmental and prosodic effects; no straightforward segmental or featural cases of noun privilege have been identified.

2.1.1 Stress, accent, and tone

In Spanish (Romance), stress location (CHAPTER 41: THE REPRESENTATION OF WORD STRESS) is contrastive for nouns but not for verbs (Harris 1983; Garrett 1996). Nouns have stress on the antepenultimate, penultimate, or final syllable; near-minimal pairs exemplifying antepenultimate and penultimate stress are shown in (1a). Verbs may appear with penultimate or final stress, but the stress location is determined by the verb's inflectional affix.

(1) *Stress location in Spanish*

a. *nouns: contrastive stress* (data from Castillo and Bond 1948; Solá 1981)

antepenultimate		*penultimate*	
['saβana]	'sheet'	[sa'βana]	'savanna'
['kaskara]	'husk'	[kas'kaða]	'waterfall'
['tortola]	'dove'	[tor'tuɣa]	'turtle'
['bispera]	'eve'	[es'pera]	'wait'

b. *verbs: stress determined by inflection* (data from Garrett 1996: 72–73)

['laβ-o]	'wash-1SG.PRES.INDIC'	[la'β-e]	'wash-1SG.PRET.INDIC'
['laβ-a]	'wash-3SG.PRES.INDIC'	[la'β-o]	'wash-3SG.PRET.INDIC'

Modern Hebrew (Semitic) has a similar pattern; nouns have stress contrasts, but verb stress is predictable (Becker 2003).[1]

A case resembling Spanish, but for pitch accent, is Tokyo Japanese (Japonic; McCawley 1968; CHAPTER 120: JAPANESE PITCH ACCENT). In nouns, accent location is contrastive. Accent, realized as a pitch fall, may appear on any syllable, and there are minimal sets among disyllabic nouns. For verbs, there is a contrast between accented and unaccented stems, but the accent location is determined by the affix category. (The principles governing verb accent location are complex; see McCawley 1968.)

(2) *Pitch-accent location in Tokyo Japanese* (data from Hirayama 1960)

a. *nouns: accent location and presence/absence both contrastive* (*-ga* marks nominative case; included to distinguish final accent from unaccented)

initial accent		*final accent*		*unaccented*	
[háʃi-ga]	'chopsticks'	[haʃí-ga]	'bridge'	[haʃi-ga]	'edge'
[káki-ga]	'oyster'	[kakí-ga]	'fence'	[kaki-ga]	'persimmon'
[kíbi-ga]	'millet'	[kibí-ga] 'sensation' ~ [kimí-ga]		[kimi-ga]	'you (INFORMAL)'

[1] See §3.2 below for further discussion of the Hebrew case, in which adjectives show a pattern intermediate between the noun and verb patterns.

b. *verbs: accent presence/absence contrastive, but location determined by affix*

accented stem		*unaccented stem*	
[kaké-ɾɯ]	'hoist-NON-PAST'	[kake-ɾɯ]	'be lacking-NON-PAST'
[káke-te]	'hoist-GERUND'	[kake-te]	'be lacking-GERUND'
[kake-nágaɾa]	'hoist-while'	[kake-nagaɾa]	'be lacking-while'

Similar patterns, in which nouns have more contrastive pitch accent choices than verbs, include other Japanese dialects (Haraguchi 1977), Proto-Korean (isolate?; Whitman 1994), Xibe (Tungusic; Kubo 2008), and Ancient Greek (Greek; Devine and Stephens 1994).

Finally, an analogous pattern can be found for tone (CHAPTER 45: THE REPRESENTATION OF TONE). In Mono (Niger-Congo, Banda; Olson 2005), nouns have lexically specified tone shapes. By contrast, verb surface tone patterns are predictable from their inflectional forms, although there is evidence from deverbal derivational forms that even verbs may have underlying tone contrasts.

(3) *Tone in Mono* (Olson 2005: 47–49, 51)

a. *nouns: any tone shape possible*

[gósá]	'type of green'	[kākó]	'leaf'	[bùdú]	'buttocks'
[zúwā]	'flour'	[lèngā]	'slit drum'	[zājā]	'anvil'
[lóbà]	'clothes'	[jāwò]	'firewood'	[gbādò]	'grub'

b. *verbs: tone determined by inflection*

Non-future:	H on first verb syllable; L on any other syllables
Future:	H on syllable preceding verb; L on all verb syllables
Imperative:	L on first verb syllable
Subjunctive:	M on first verb syllable
Stative:	Reduplicate first verb syllable; reduplicant bears HL; verb root bears L
Certainty:	Reduplicant bears HM; M on first verb-root syllable; L on any other syllables

Other languages in which nouns have more tone contrast possibilities than verbs (in complexity of underlying tone, in H tone location contrasts, or in resistance to tonal alternations) include Proto-Bantu and various modern Bantu languages (Kisseberth and Odden 2003) and Gã (Kwa; Paster 2000); see also CHAPTER 114: BANTU TONE.

2.1.2 *Prosodic shape*

In Hebrew (Semitic; Glinert 1988; Bat-El 1994), as well as in closely related Arabic (Semitic; McCarthy 2005; Ryding 2005), verbs are subject to a prosodic-shape restriction – they must fit into one of a number of disyllabic templates. Nouns may be templatic, but they need not be, particularly for loanwords (CHAPTER 95: LOANWORD PHONOLOGY). In (4), atemplatic nouns and verbs derived from those nouns are shown; regardless of the noun shape, the verbs are templatic, being disyllabic and (here) showing the /i e/ of the *piʔel* conjugation.

(4) *Prosodic shape in Hebrew* (Bat-El 1994: 577–578)

a. *nouns: not necessarily disyllabic*

[xantariʃ]	'nonsense'
[telegraf]	'telegraph'
[sinxroni]	'synchronic'
[ksilofon]	'xylophone'
[nostalgia]	'nostalgia'

b. *verbs: must fit disyllabic template*

[xintreʃ]	'talk nonsense'
[tilgref]	'telegraph'
[sinxren]	'synchronize'
[ksilfen]	'play the xylophone'
[nistelg]	'be nostalgic'

A different prosodic-shape effect is found in Mbabaram (Paman; Dixon 1991). Long vowels are relatively rare, but they are found only in nouns, never in verbs.

(5) *Long vowels in Mbabaram nouns* (Dixon 1991: 357)

a. *form with long vowel (noun)*

[guːɽ]	'nulla nulla'
[jaːɽ$^{(ə)}$]	'spear'
[ˈnɔːmbɨ]	'big red wallaroo'
[gaˈwiːr]	'tomahawk'
[namˈbuːɽ]	'big brown snake'

b. *minimally contrasting form, for comparison*

[guɽ]	'elbow'
[ja-ɽə]	'give-PAST'

2.1.3 *Absence of segmental patterns*

All the cases of noun privilege reviewed above involve a suprasegmental or prosodic contrast – tone, accent, or stress, or word shape or size. Even the vowel-length case in Mbabaram is prosodic rather than segmental, on the view that vowel length is not a segmental feature, but results from the association of one segmental melody to two timing units (Clements and Keyser 1983).

One apparent case of noun privilege that does involve segmental features is found in Nivkh (isolate). However, Shiraishi (2004) demonstrates that apparent noun/verb asymmetries in Nivkh can be analyzed in terms of base identity, since nouns can appear unaffixed in Nivkh but verbs cannot. (See §4.1 below for discussion of this point and, more generally, of the relationship between lexical category effects and the distinction between free and bound stems.)

2.2 *Phonological augmentation in nouns*

In several languages, only nouns are subject to word-minimality requirements. This might look like verb privilege, as requirements are imposed specifically on nouns. However, there is one circumstance in which phonological privilege correlates with special requirements: positional augmentation (Smith 2002), in which a privileged position is required to have some perceptually salient property.

Positional augmentation is technically a type of neutralization, because all instances of position *P* must have property *X* – but unlike other types of neutralization, it is a diagnostic for phonological strength.

Word minimality has been analyzed as a requirement for a (morphological) word to be coextensive with a prosodic word (Prince 1980; Broselow 1982; Crowhurst 1992; CHAPTER 51: THE PHONOLOGICAL WORD). Content morphemes are often subject to minimality effects in contrast to function morphemes, supporting the classification of minimality effects as positional augmentation (CHAPTER 104: ROOT–AFFIX ASYMMETRIES). Thus, the noun-specific minimality effects discussed here are compatible with the claim that nouns have greater phonological privilege than verbs.

Chuukese (Micronesian; Muller 1999) is one language in which nouns, but not verbs, are subject to a bimoraic word-size minimum. There is a general requirement, affecting both nouns and verbs, that the expected word-final mora not surface, so underlying final vowels are deleted if short and shortened if long. Crucially, when this truncation process would result in a monomoraic surface form, nouns undergo vowel lengthening. This results in a surface contrast between CVC and CV:C for verbs, but not for nouns, because potential *CVC nouns surface as CV:C.

(6) *Word-minimality in Chuukese* (final codas do not contribute weight; initial geminates do)

a. *nouns: minimally bimoraic* (Muller 1999: 395)

	UR	*Final mora loss*		
CCVC already bimoraic	/kkeji/	[kkej]	'laugh'	
	/ʧʧara/	[ʧʧar]	'starfish'	
*CVC undergoes lengthening	/fasa/	[fa:s]	'nest'	*[fas]
	/fæne/	[fæ:n]	'building'	*[fæn]

b. *verbs: no bimoraic minimum* (data from Goodenough and Sugita 1980: xiv–xv)

[fan]	'go aground'	[fa:n]	'break open'
[mær]	'move, be shifted'	[mæ:r]	'grow (plant)'

See also §4.1.2 for a discussion of Chuukese in the context of the morphological free/bound distinction and category-specific phonology.

Other languages in which nouns, but not verbs, have minimality requirements include Chukchee and Koryak (Chukotko-Kamchatkan; Krause 1979).

2.3 *Phonological privilege in verbs*

The languages in §2.1 are clearly classifiable as cases of noun privilege, since nouns allow more contrasts than verbs. Given that augmentation processes specifically target privileged positions, the cases in §2.2 are also compatible with the view that nouns are privileged compared to verbs. The languages considered in this section, however, present a different pattern; they seem to show greater phonological privilege for verbs than for nouns.

2.3.1 *Tone*

In Ewe (Kwa; Ansre 1961) nouns, the contrast between H and L tone is neutralized in syllables with voiced obstruent onsets; only L is possible in that context (CHAPTER 97: TONOGENESIS). However, verbs may have H or L tone with any onset type.

(7) *Consonant/tone co-occurrence restrictions in Ewe* (Ansre 1961: 27–28, 34–35, 39)

a. *nouns: voiced obstruent onset may not co-occur with H tone*[2]

voiceless obstruents	[ɸú]	'bone'	[ɸù]	'sea'
	[tú]	'gun'	[tè]	'yam'
sonorants	[jí]	'cutlass'	[à-jè]	'trick'
	[ŋɔ́]	'worm'	[à-ŋè]	'rubber'
voiced obstruents	(H unattested)		[βù]	'blood'
			[dà]	'snake'

b. *verbs: onsets and tones co-occur freely*

voiceless obstruents	[fá]	'is cold'	[fù]	'is white'
	[tú]	'to shut'	[tù]	'to grind'
sonorants	[jɔ́]	'to call'	[jɔ̀]	'to hurry'
	[ŋé]	'to break'	[ŋè]	'to groan'
voiced obstruents	[bú]	'to be lost'	[bù]	'to respect'
	[vó]	'to rot'	[vò]	'to be free'

Verbs thus have a greater number of surface tone contrasts than nouns.

2.3.2 *Segmental deletion?*

A small number of cases may involve verb privilege in resisting segmental deletion (CHAPTER 68: DELETION). These are unusual in two ways: noun privilege seems to be much more common than verb privilege, and category-specific phenomena tend to be suprasegmental or prosodic, rather than segmental or featural. In fact, the cases discussed here are not unambiguous examples of segmental deletion, and at least one might be re-analyzed as noun augmentation rather than verb privilege.

One apparent example of segmental deletion that affects nouns, but not verbs, is seen in Paamese (Oceanic; Crowley 1997: 243–244). Proto-Paamese */l/ was lost in northern Paamese in a variety of environments, including word-initially, but word-initial */l/ has been preserved in verbs specifically.

(8) *Loss of initial */l/ in northern Paamese does not apply to verbs* (Crowley 1997: 243–244)

a. *nouns show loss of initial */l/*

*leiai	→	[eiai]	'bush'
*la:la	→	[a:la]	'kind of bird'

b. *verbs preserve initial */l/*

*leheie	→	[lehei]	's/he pulled it'
*loho	→	[loh]	's/he ran'
*la:po	→	[la:po]	's/he fell'

Liquid onsets, being high in sonority, are marked, especially in word-initial position (Smith 2002; see also CHAPTER 49: SONORITY; CHAPTER 55: ONSETS; CHAPTER 31:

[2] Some noun examples are slightly modified from Ansre (1961). Ansre shows these L-tone examples for voiceless obstruents and sonorants with a specifier [lá], glossed 'the' ([ɸū lá] 'the sea'), in order to demonstrate that the L tone has a mid allotone when non-final. I have removed the specifier and adjusted the gloss and tone marking, because the distinction between allotones of L is not of concern here.

LATERAL CONSONANTS). Thus in Paamese we seem to have a case where a marked segment is tolerated in verbs even though it is not tolerated in nouns. If this is the correct interpretation, Paamese would be a case of greater phonological privilege for verbs. However, the avoidance of high-sonority onsets is arguably another kind of augmentation effect (Smith 2002; see also de Lacy 2001), providing a sharp sonority increase for syllables in prominent positions. Thus, a second interpretation is possible, according to which the avoidance of initial liquids in nouns is evidence for noun privilege after all. Moreover, viewing initial */l/ loss in nouns as driven by sonority (i.e. syllable structure) requirements would bring Paamese into line with the general observation that category-specific phenomena are prosodic rather than segmental.

Another example that might involve segmental deletion is Mohawk, in which, according to Postal (1968), word-final stops were lost in nouns (except reduplicating animal names), although they were retained in morphologically related verbs (CHAPTER 36: FINAL CONSONANTS). This does appear to be a phonological process affecting specifically nouns and not verbs – a case of verb privilege. However, it is unclear whether segmental deletion is the best characterization of the process. Postal's examples are compatible with the view that the driving force behind the noun deletion was word-final cluster simplification, which again is a matter of syllable structure, not segments per se.

2.4 Distinct restrictions on nouns and verbs

This section presents cases in which neither nouns nor verbs appear to have a greater array of phonological contrasts; both categories are subject to some phonotactic requirement. However, the requirements that hold of nouns and those that hold of verbs are distinct.

2.4.1 *Stress assignment*

In Lenakel (Oceanic; Lynch 1975, 1978), primary stress is always penultimate, but secondary stress assignment is different for nouns and verbs. In nouns, secondary stresses are assigned on alternating syllables *leftward* from the primary stress, and initial syllables might not bear stress. In verbs, the initial syllable always bears a secondary stress (unless the second syllable has primary stress), and additional secondary stresses are assigned on alternating syllables *rightward* from the initial syllable, subject to the condition that no secondary stress immediately precedes the primary stress.

(9) *Stress assignment in Lenakel* (Lynch 1978: 19)

a. *nouns: secondary stress assigned leftward from primary stress syllable*

/kamatoa/	[ˌka.ma.'do.a]	'kind of taro'
/nimʷakilakil/	[nɨ.ˌmʷɒ.gə.'la.gəl]	'beach'
/tupʷalukaluk/	[tu.ˌbʷɒ.lu.'ga.lukʰ]	'lungs'

b. *verbs: secondary stress assigned rightward from initial syllable*

/r-ɨm-olkeikei/	[ˌri.mɔl.'gɛj.gɛj]	'he liked it'
/n-ɨm-ar-olkeikei/	[ˌnɨ.ma.rɔl.'gɛj.gɛj]	'you-PL liked it'
/n-ɨm-am-ar-olkeikei/	[ˌnɨ.ma.ˌma.rɔl.'gɛj.gɛj]	'you-PL were liking it'
/t-n-ak-am-ar-olkeikei/	[ˌtɨ.na.ˌga.ma.rɔl.'gɛj.gɛj]	'you-PL will be liking it'

A similar, although less straightforward, case is English (Germanic; Kelly 1988; see also Chomsky and Halle 1968), which has a preference (not a requirement) for initial/trochaic stress in disyllabic nouns *vs.* final/iambic stress in disyllabic verbs. See §4.3 below.

2.4.2 *Tonal patterns*

In Lamang (Chadic; Wolff 1983), both nouns and verbs have predictable tone, but different factors determine the surface tones in each case (10). Noun tones are determined by the onset consonant, interacting with phrase-level and assimilatory effects. The basic pattern is that tone is low when the onset is a voiced obstruent (as for nouns in Ewe; see §2.3.1), and tone is high[3] when the onset is a sonorant, a voiceless obstruent, or one of [ɓ] or [ɗ], which Wolff (1983: 28) describes as "laryngealized" but "only . . . incidentally 'implosive' on some occasions." Verb tones, on the other hand, are entirely determined by inflectional category, except for two exceptional verb roots that pattern like nouns.

(10) *Predictable tone patterns in Lamang* (Wolff 1983: 67–68, 77)

a. *nouns: L after voiced obstruent, H otherwise*

voiced obstruents (L)		*voiceless obstruents (H)*		*sonorants (H)*	
[ɣ̱ʷà]	'mountain'	[ɬ̱á]	'cow'	[éw̱é]	'mouth'
[ḏẕə̀v̱ò]	'hand'	[úṯáḵá]	'country'	[ɪhkíḻí]	'rat (sp.)'

b. *verbs: tone determined by inflectional category*

[kə̀lì]	'I take (CONT)'	[kə́lí]	'I take (DUR CONT)'
[kàkə̀lì]	'I have begun to take (PERF II)'	[kákə́lí]	'I have taken (PERF I)'
[kə̀làjò]	'that I take (SUBJ I)'	[kə̀láјó]	'I took (AORIST)'

Lamang is typologically unusual, in that, unlike nouns and verbs, function morphemes do have contrastive tones (Wolff 1983: 74). It is more common typologically for lexical morphemes to have greater freedom in phonological contrast than function morphemes have (McCarthy and Prince 1995: §6.2; Beckman 1999; CHAPTER 104: ROOT–AFFIX ASYMMETRIES). Lamang ideophones, which Wolff calls "expressives," also have contrastive tones.

2.4.3 *Prosodic shape*

Finally, there are some languages in which prosodic-shape restrictions are found for both nouns and verbs, but the restrictions are different for the two categories.

One example, discussed by McCarthy (2005), is Classical Arabic (Semitic; CHAPTER 108: SEMITIC TEMPLATES). As mentioned in §2.1.2, all verbs in Arabic are templatic; most nouns are templatic as well. McCarthy (2005) shows that different restrictions on template shape hold for nouns and for verbs. Noun templates may only begin with one consonant, but verb templates may begin with either one or two consonants. On the other hand, verb templates must end with CVC, but noun templates may end with CVC, CV:C, or CVCC.

[3] In some cases a preceding low tone causes this potential high tone to assimilate and become low (Wolff 1983: 67). In addition, there are certain phrase-level dissimilatory effects, as well as an "accent" that may boost tone on the antepenultimate syllable of the phrase. For a summary of predictable tone effects in nouns, see Wolff (1983: 72).

(11) *Template shape in Arabic* (McCarthy 2005: 178, 209)

a. *noun templates: no initial* CC; *may end with* CVC, CV:C *or* CVCC

Triliteral	*Quadriliteral*
CV CVC	CVC CVC
CV: CVC	
CV CV:C	CVC CV:C
CV: CV:C	
CVCC	

b. *verb templates: initial* CC *allowed; must end with* CVC

CV CVC	CVC CVC	CV: CVC
CCV CVC	CCVC CVC	CCV: CVC

Another language in which prosodic-shape requirements affect nouns and verbs differently is Itelmen (Bobaljik 1998, 2008). In nouns, a "resonant" consonant (sonorants and [z]) must be adjacent to a vowel; otherwise, a preceding schwa is epenthesized. Nouns consequently show schwa–zero alternations, since the environment for epenthesis is met in some morphological forms but not in others (CHAPTER 26: SCHWA; CHAPTER 67: VOWEL EPENTHESIS). In verbs, resonants likewise never violate this restriction, but there are no schwa–zero alternations. According to Bobaljik's interpretation, schwa epenthesis overapplies in verbs, in that it applies to *all* forms of a verb if its environment is met in *some* form.

(12) *Schwa epenthesis in Itelmen* (Bobaljik 2008: 44)

a. *nouns: epenthesis and alternations*

[ɬxəm]	'sable-SG'	[ɬxm-en']	'sable-PL'
[spəl]	'wind-DIRECT'	[spl-ank]	'wind-LOCATIVE'
[ʷtχəz-xʔal]	'road-ABLATIVE'	[ʷtχz-enk]	'road-LOCATIVE'

b. *verbs: no alternations; overapplication of epenthesis*

[t-zəl-ʧen]	'I gave it.'	[zəl-en]	'You gave it.'	*[zlen]
[t-ɬəm-ʧen']	'I killed them.'	[q-ɬəm-in]	'Kill it!'	*[qɬmin]
[spəl-qzu-in]	'It was windy.'	[spəl-in]	'It was windy.'	*[splin]

The Itelmen pattern involves a prosodic-shape requirement that affects both nouns and verbs: a syllabification restriction on resonants. However, nouns may show morphological alternations with respect to the schwa epenthesis, whereas verbs may not. Bobaljik's (1998, 2008) formal analysis of this pattern has epenthesis apply cyclically in verbs but non-cyclically in nouns; it is therefore only verbs that undergo epenthesis in cases where subsequent affixation would potentially bleed that process. On this approach, it is not immediately clear whether nouns or verbs should be seen to have greater phonological privilege, since both are subject to an epenthesis process. (On the other hand, if the avoidance of alternation is seen as an additional requirement that holds of verbs only, then Itelmen could be a case of noun privilege, as in §2.1.)

Both Classical Arabic and Itelmen have featured in discussions of base identity as an alternative to category-specific phonology; see §4.1.2 below.

2.4.4 Absence of patterns involving segmental contrasts

As with the examples of noun privilege discussed in §2.1 and §2.2, the languages with distinct predictable patterns for nouns and verbs seem to involve exclusively suprasegmental and prosodic properties such as stress, tone, and syllable structure. One language that has been said to have different segmental inventories in nouns and verbs is Michif (Bakker 1997). However, this difference would probably not strictly speaking be a lexical category effect, as Michif is a mixed language in which nouns and verbs tend to derive from distinct source languages, Canadian French (Romance) and Plains Cree (Algonquian) respectively. Furthermore, Rosen (2007) argues that French/Cree stratification is not synchronically relevant for Michif phonology.

2.5 *Summary: Survey of category-specific effects*

The category-specific effects reviewed in §2, involving differences in phonological behavior between nouns and verbs, are summarized in Table 102.1.

Noun privilege (plus noun augmentation, which is arguably related to noun privilege) appears to be the most common pattern, with fewer cases of verb privilege and distinct predictable patterns for nouns and verbs. Nearly all of these examples of category-specific phonological behavior involve either suprasegmental properties like stress, accent, and tone, or else prosodic shape (word-minimality, word or syllable shape, or vowel length). The only cases possibly involving segmental phenomena that have emerged in this survey are the two examples of diachronic segment deletion specific to nouns, and, as noted in §2.3.2, these may be reinterpretable as prosodic effects as well.

Table 102.1 Noun/verb differences in phonological behavior

Language	*Phenomenon*	*N/V pattern*
Spanish Hebrew Japanese Proto-Korean Xibe Ancient Greek Mono Proto-Bantu Hebrew Mbabaram	stress stress accent accent accent accent tone tone prosodic shape prosodic shape	N privilege
Chuukese Chukchee, Koryak	prosodic shape prosodic shape	N augmentation
Paamese	diachronic segment deletion(?)	V privilege?/N augmentation?
Ewe Mohawk	tone diachronic segment deletion(?)	V privilege
Lenakel Lamang Arabic Itelmen	stress tone prosodic shape prosodic shape	distinct but predictable

3 Adjectives

Adjectives are in a sense intermediate between nouns and verbs morphosyntactically, and their phonological behavior reflects this. In some languages, adjectives fall together either with nouns or with verbs in terms of their phonology, and the grouping they form often correlates with the nature of adjectival inflectional morphology in the language (§3.1). In other languages, adjectives have "partial privilege" – they lie between nouns and verbs on a continuum of phonological behavior (§3.2).

3.1 *Adjectives as a subcase of nouns or verbs*

In a language with category-specific phonological patterns, adjectives often pattern with either nouns or verbs in a way that mirrors their categories of inflection (see Table 102.2).

When adjectives inflect for N-type morphological categories such as person, gender, number, or case – as in Spanish, Mono, Mbabaram, and Hebrew, Table 102.2(a) – they tend to pattern phonologically with nouns. Analogously, when adjectives inflect for V-type categories such as tense, mood, or aspect – as in Japanese and Ewe, Table 102.2(b) – they tend to pattern phonologically with verbs.

The languages in Table 102.2(c) present additional complications, however. In Hebrew, adjectives inflect for nominal categories, but their behavior with respect to stress is actually intermediate between that of nouns and verbs. Mandarin (Sino-Tibetan) is a language that does not have much of an inflectional system at all, but here again, adjectives show a pattern that is distinct from both nouns and verbs. These two cases are discussed in §3.2, along with an additional case, Finnish (Finnic), which shows distinct behavior between nouns and adjectives even when they bear the *same* inflectional morphemes. Finally, Lenakel (see §2.4.1 above) appears to show a tight correlation between inflectional morphology and category-specific phonology; this language is discussed further in §4.2 below.

Table 102.2 Adjective behavior and inflection type

Language	*Phenomenon*	*N/V pattern*	*Adjective behavior*	*Adjective inflection*
a. Adjectives pattern with nouns				
Spanish Mono Mbabaram Hebrew	stress tone prosodic shape prosodic shape	N privilege	same as N	N-type
b. Adjectives pattern with verbs				
Japanese Ewe	accent tone	N privilege V privilege	same as V	V-type
c. Pattern is more complicated				
Hebrew Mandarin Lenakel	stress reduplication stress	N privilege distinct distinct	N > A > V distinct depends on role?	N-type isolating language depends on role?

3.2 *Adjectives as an intermediate category*

Many languages show adjectives patterning together with either nouns or verbs, but Hebrew stress, Mandarin reduplication, and Finnish mutation and deletion are phonological phenomena in which adjectives have their own specific pattern. These cases nevertheless suggest that, even when adjectives show distinct behavior, they fall at a point intermediate between nouns and verbs with respect to phonological privilege.

Becker (2003) discusses stress in Hebrew, in which nouns, adjectives, and verbs all show distinct behavior. The default is "mobile" stress, in which stress is attracted to the right edge of the word: mobile stress falls on the final syllable of an unaffixed form, or on the rightmost suffix. All verbs have mobile stress. Adjectives and nouns differ from verbs; they have a phonological contrast between mobile stress and "fixed" stress, a pattern in which stress remains on a particular syllable of the base. But there is a further difference between adjectives and nouns. When adjectives have fixed stress, it always falls on the root-final syllable, whereas the location of fixed stress is contrastive for nouns.

(13) *Stress contrasts in Hebrew* (Becker 2003: 1–2)

a. *nouns: location of fixed stress is contrastive*

mobile stress			
[dik'duk]	'grammar-SG'	[dikduk-'im]	'grammar-PL'
fixed stress			
['kopirajter]	'copywriter-SG'	['kopirajter-im]	'copywriter-PL'
[dik'tator]	'dictator-SG'	[dik'tator-im]	'dictator-PL'
['tut]	'strawberry-SG'	['tut-im]	'strawberry-PL'

b. *adjectives: fixed stress is always root-final*

mobile stress			
['tov]	'good-SG'	[tov-'im]	'good-PL'
fixed stress			
[mal'jan]	'rich-SG'	[mal'jan-im]	'rich-PL'
[fono'log-i]	'phonological-SG'	[fono'log-i-im]	'phonological-PL'

c. *verbs: always mobile stress*

[ʃa'mar]	'keep-SG'	[ʃamr-'u]	'keep-PL'
[biz'bez]	'spend-SG'	[bizbez-'u]	'spend-PL'

As Becker (2003) observes, this means that adjectives have greater phonological freedom than verbs, but not as much as nouns.

In Mandarin, nouns, adjectives, and verbs show distinct behavior in reduplication (Feng 2003; see also CHAPTER 100: REDUPLICATION). Disyllabic (AB) adjectives reduplicate as AABB, while disyllabic verbs reduplicate as ABAB. (Disyllabic nouns do not reduplicate, although monosyllabic nouns do.) Some adjective or verb bases reduplicate both ways, in which case the AABB form is an adjective, and the ABAB form is a verb.

(14) *Reduplication in Mandarin* (Feng 2003: 2)[4]

a. *adjectives:* AB → AABB

[kán.tɕîŋ]	'clean'	[kán.kán.tɕîŋ.tɕîŋ]	'clean (intensified)'
[mʲǐŋ.pǎi]	'clear'	[mʲǐŋ.mʲǐŋ.pǎi.pǎi]	'clear (intensified)'

b. *verbs:* AB → ABAB

[tɕʰîŋ.ʈʂʷû]	'celebrate'	[tɕʰîŋ.ʈʂʷû.tɕʰîŋ.ʈʂʷû]	'celebrate a little'
[tǎ.sàw]	'clean up'	[tǎ.sàw.tǎ.sàw]	'clean up a little'

c. *shape of reduplicated form determines category*

adjective base	[kāw.ɕîŋ]	'happy'
AABB = adjective	[káw.kǎw.ɕîŋ.ɕîŋ]	'happy (intensified)'
ABAB = verb	[káw.ɕîŋ.káw.ɕîŋ]	'have some fun'
verb base	[ʈʂǐ.tʲɛ̀n]	'suggest'
AABB = adjective	[ʈʂǐ.ʈʂǐ.tʲɛ̀n.tʲɛ̀n]	'critical, bossy'
ABAB = verb	[ʈʂǐ.tʲɛ̀n.ʈʂǐ.tʲɛ̀n]	'comment here and there'

Feng (2003: 7) presents evidence from third-tone sandhi alternations (CHAPTER 107: CHINESE TONE SANDHI) that the morphosyntactic bracketing in the two cases is distinct; namely [A[AB]B] (adjectives) but [AB][AB] (verbs). She sees this as driving the difference between adjective and verb reduplication; for verbs, it is more important for the edges of morphosyntactic constituents to align with the edges of prosodic constituents (on the assumption that all four-syllable reduplicated forms have the prosodic constituency (σσ)(σσ)), whereas, for adjectives, it is more important that the linear sequence AB from the base form not be disrupted. The different structures and their differing priorities are shown in (15).

(15) *Morphosyntactic and prosodic constituents in Mandarin reduplication*

verbs

		A	B	A	B
Morphosyntactic	[		][		]
Prosodic	(		)(		)

Edges match; linear sequence of base not respected

adjectives

		A		A	B		B
Morphosyntactic	[		[			]	]
Prosodic	(			)(			)

Edges do not match; linear sequence of base respected

Mandarin differs from Hebrew because it is not entirely clear in Mandarin whether it is adjectives or verbs that should be seen as having greater phonological privilege; each category has a predictable reduplication type, even though they differ. However, adjectives do set a higher priority for maintaining the linear sequence of the base. From the viewpoint of Optimality Theory (Prince and Smolensky 2004), under Correspondence Theory (McCarthy and Prince 1995), this

[4] Mandarin tone marks are given in accordance with IPA usage, rather than the pinyin transcriptions used by Feng (2003).

suggests higher-ranking faithfulness for adjectives than for verbs, and therefore greater phonological privilege for adjectives.[5]

Finally, Anttila (2002) discusses a category-specific effect in Finnish involving two different phonological alternations: under certain morphological and phonological conditions, a stem-final /a/ either deletes or changes to /o/ (mutates), when the plural morpheme /-i-/ is added. An examination of a corpus of Finnish shows that, in the absence of phonological factors leading to a preference for one strategy or the other (e.g. a tendency to avoid mutation after labial consonants), the mutation option is preferred by nouns and the deletion option is preferred by adjectives. This difference is seen even when nouns and adjectives are inflected with identical affixes.

(16) *Deletion* vs. *mutation in Finnish* (Anttila 2002: 13)

a. *nouns: prefer mutation*

/kihara-i-ssa/	'curl-PL-INESS'	→	[kiharo̱-i-ssa]
/korea-i-ssa/	'Korea-PL-INESS'	→	[koreo̱-i-ssa]

b. *adjectives: prefer deletion*

/kihara-i-ssa/	'curly-PL-INESS'	→	[kiharØ-i-ssa]
/korea-i-ssa/	'beautiful-PL-INESS'	→	[koreØ-i-ssa]

Again, it is somewhat difficult to interpret this pattern in terms of relative phonological privilege, because both mutation and deletion involve a phonological process. However, mutation, the pattern favored by nouns, does preserve all input segments (even though certain feature values are changed), so it is not out of the question to view the correlation between mutation for nouns and deletion for adjectives as a consequence of greater noun privilege.

3.3 Adjective patterns as scale conflation

The examples in §3.1 and §3.2 all appear to be compatible with one of the following scales:[6] adjectives fall together with nouns, fall together with verbs, or exhibit a degree of privilege intermediate between that for nouns and verbs.

(17) *Scales of phonological privilege by lexical category*

a. {N, A} > V
b. N > {A, V}
c. N > A > V

[5] Another interesting point related to lexical categories in Mandarin reduplication is that AB verbs actually reduplicate as ABAB only if both the A and the B morphemes are verbal; V + object (N) forms reduplicate as AAB (Feng 2003: 3).

[6] The scales in (17) assume a more basic scale N > V, with greater phonological privilege for nouns than for verbs. As noted in §2.5 above, cases that seem to show greater privilege for verbs than for nouns appear to be few in number and exceptional in pattern; in particular, two out of three of the cases may involve segmental phonology, which category-specific patterns generally do not. Further investigation is needed to determine whether the scale of privilege N > V is truly a typological (near-)universal, but the intermediate behavior of adjectives reviewed in this section could be seen as additional support for this view.

This pattern resembles *markedness conflation* (de Lacy 2004), in which there is a universal markedness scale X > Y > Z, but on a language-specific basis adjacent levels of the scale can be conflated and pattern as a single class with respect to that markedness dimension (CHAPTER 4: MARKEDNESS; CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). So, positing a universal scale of phonological privilege for lexical categories, as in (17c), is compatible with the existence of languages that conflate the middle category, A, with N (17a) or V (17b).

4 Attempting non-phonological explanations

The discussion so far has focused on demonstrating that there exist languages with category-specific phonological effects. However, important questions remain. Does the phonological grammar actually need to refer to lexical categories? Or can all category-specific effects be reduced to epiphenomena, simple outcomes of the interaction between phonology and other modules of the grammar such as morphology or syntax? Finally, to the extent that there are category-specific effects in phonology, does this require reference to the same lexical category labels used by the morphosyntax, or should it instead be handled in the same way as cases of exceptional phonological behavior by arbitrary classes of morphemes?

This section considers the distinction between morphologically free and bound forms (§4.1), the relationship between nominal or verbal inflection and category-specific behavior (§4.2), prosodic structure (§4.3), and morpheme-specific effects (§4.4). While some of these factors are relevant some of the time, not all cases of category-specific phonology can be re-analyzed in these terms.

4.1 *Free vs. bound*

As seen in Table 102.3, there is often overlap between the lexical categories in a language with phonological privilege, and the lexical categories in that language that occur as free forms (without obligatory overt inflection). Indeed, some of the apparent category-specific effects discussed in the literature have been shown either to reduce directly to the free/bound distinction, or to exhibit base-identity effects that can be formally modeled with reference to the free/bound distinction. Examples of each type are reviewed in §4.1.1.

As shown in Table 102.4, however, some cases of category-specific phonological patterns cannot be directly equated with the distinction between free and bound forms. These are discussed in §4.1.2.

Table 102.3 Phonological privilege matches free/bound distinction

Language	*Phenomenon*	*N/A/V privilege*	*Free/bound*
Japanese	accent	N > A, V	N free | A, V bound
Mono	tone	N, A > V	N, A free | V bound
Mbabaram	prosodic shape	N, A > V	N, A free (absolutive case) | V bound
Lenakel	stress	N, (A) ≠ (A), V	N free | A by function | V bound

Table 102.4 No exact match between phonological privilege and free/bound distinction

Language	*Phenomenon*	*N/A/V privilege*	*Free/bound*
Spanish	stress	N, A > V	N, A free or bound | V bound
Chuukese	prosodic shape	N > V	both N, V free (modulo truncation)
Mandarin	reduplication	N ≠ A ≠ V	N, A, V can all be free
Hebrew	prosodic shape stress	N, A > V N > A > V	N, A free or bound | V bound

4.1.1 *Category-specific effects that reduce to free/bound effects*

It has been argued that some apparent category-specific effects reduce to the distinction between free and bound forms, either directly, as for word-minimality requirements, or indirectly, as for base-identity effects.

In Chuukese (§2.2; see also §4.1.2), minimality requirements truly differentiate between nouns and verbs. However, some cases of apparent category-specific minimality are actually due to the free/bound distinction: if the minimality requirement holds at the level of the word, and (by definition) only free forms can constitute words on their own, then only free forms show minimality effects.

For example, consider Mono (Niger-Congo, Banda). Olson (2005: 75, 79) observes that there are no monosyllabic surface forms of nouns (or adjectives), and proposes a process of Subminimal Root Augmentation (SRA), which epenthesizes a copy vowel into an underlyingly monosyllabic noun: /CV_1/ → [V_1CV_1]. However, Olson (2005: 82) explicitly notes that SRA does not apply to verbs, because verbs must appear with some inflectional affix, so they never happen to be monosyllabic on the surface even if they have a subminimal root. Moreover, Olson (2005: 89, 94) shows that even /CV/ nouns fail to undergo SRA if they bear a plural affix or form part of a compound. In short, there is no need for a category-specific analysis of minimality in Mono. It is simply the case that *words* must be disyllabic; for unrelated reasons, only nouns and adjectives may surface unaffixed, so only they are ever in danger of violating the category-free requirement on word size.

There is another, more abstract way in which the free/bound distinction potentially has implications for privilege in maintaining phonological contrast. In some languages, morphologically complex forms show base-identity effects – phonological similarity to some aspect of their morphological base forms. This has been modeled as phonological cyclicity, or in terms of constraints that enforce faithfulness to morphologically related base forms (e.g. Kiparsky 1982, 2000; Kenstowicz 1996; Benua 2000; CHAPTER 83: PARADIGMS; CHAPTER 85: CYCLICITY). Schematically, suppose that a language has a base form /X/ and a complex form /X + Y/. In the absence of base-identity effects, the phonological grammar simply applies to the segments in /X + Y/ as they appear there. But, if there is a base-identity effect, then some similarity requirement holds between the surface form of [X + Y] and the surface form of its base [X], giving rise to a property of [X + Y] that would not be expected if this form were simply subject to the phonological grammar of the language on its own.

Base-identity effects are relevant in the context of the free/bound distinction because if a root /X/ never appears unaffixed, as *[X], then the non-existent surface form *[X] will never influence the phonology of the morphologically complex /X + Y/. If nouns and verbs differ precisely in this way, such that [N] is a possible surface form but [V] is not, this could potentially lead to differences in the phonology between [N + affix] and [V + affix] forms: [N + affix] forms might show base-identity effects that [V + affix] forms do not show. Such a pattern would not require crucial use of category-specific phonology, however, because the fundamental distinction would instead be that between free and bound forms.

Precisely this argument has been made by Shiraishi (2004) for Nivkh (isolate), in which noun phonology and verb phonology differ in several ways, involving segmental effects (which, as noted above, are rare among category-specific phonological phenomena). Nivkh has a four-way contrast among obstruents: stops and fricatives contrast with each other, and, furthermore, aspirated stops and voiceless fricatives (the "fortis" obstruents) contrast with plain stops and voiced fricatives (the "lenis" obstruents).

(18) *Nivkh obstruent inventory* (Shiraishi 2004: §2.1)[7]

	fortis					lenis				
stops	pʰ	tʰ	cʰ	kʰ	qʰ	p	t	c	k	q
fricatives	f	ř	s	x	χ	v	r	z	ɣ	ʁ

Although these sounds are all contrastive, there are contexts in which neutralization processes override these contrasts (CHAPTER 80: MERGERS AND NEUTRALIZATION). One such case involves the neutralization of the stop–fricative contrast in non-phrase-initial position in a morphologically derived environment (CHAPTER 88: DERIVED ENVIRONMENT EFFECTS): following a vowel, glide, or stop, obstruents surface as fricatives, but following a nasal or fricative, obstruents surface as stops (Shiraishi 2004: §2.1).

An apparent difference between categories appears in a systematic exception to this stop/fricative pattern. Nouns resist changing stem-initial fricatives into stops when a pre-stem morpheme is added. The requirements of morphologically derived environment and non-phrase-initial position are met, but the underlying fricatives still surface as fricatives.

(19) *Hardening in Nivkh* (Shiraishi 2004: §2.1–2)

a. *nouns: resist hardening*

[tʰulv vo]	*[tʰulv bo]	'winter + village'
[cʰŋər vox]	*[cʰŋər box]	'grass + hill'
[təf řə]	*[təf tʰə]	'house + door'
[tʰeŋ vaqi]	*[tʰeŋ baqi]	'coal + box'

b. *verbs: undergo hardening*

[cʰxəf qʰa-]	(< /χa-/)	'bear + shoot'
[cus tʰa-]	(< /řa-/)	'meat + bake'
[tux ke-]	(< /xe-/)	'axe + take'
[pʰnənx təu-]	(< /rəu-/)	'one's sister + teach'

[7] Shiraishi (2004: §2.1) describes /c cʰ/ as (pre-)palatals that are inconsistently characterized in the literature as plosives or affricates. He notes that /r/ and its (partially) devoiced counterpart /ř/ pattern phonologically with fricatives.

Shiraishi argues that the important difference here is that noun stems can appear in isolation, but verb stems require affixation. Thus, only nouns are potentially subject to base-identity effects. In Shiraishi's (2004: §2.4) analysis, the fricative status of the initial consonant in a derived form must match that of its underived base, and this identity requirement takes priority over the usual process of hardening. For verbs, which have no base to enforce identity, hardening applies unimpeded. No reference to lexical category is needed to account for the noun/verb asymmetry. Shiraishi also presents a similar analysis for a second segmental phonology difference between lexical categories in Nivkh, a process of stem-final fricative voicing under suffixation from which nouns are, again, exceptionally exempt.

Base-identity accounts of category-specific patterns have also been developed by Kenstowicz (1996) for cluster simplification in Korean and by Cable (2005) for schwa epenthesis in Itelmen. However, Bobaljik (2008) argues that base identity is not the right way to approach Itelmen; see the discussion in §4.1.2 below. (See also Albright 2008 and Albright and Kang 2008 for a different view of "base" for Korean nouns and verbs.)

Thus, for languages where a phonological difference between nouns and verbs aligns with the distinction between free and bound roots, it is possible that base identity could be invoked instead of category-specific phonological processes. This is a particularly attractive approach to Nivkh, where the phenomenon involved (segmental alternation) is not one that typically participates in category-specific effects. For a case like Mono, where lexical category is empirically less successful than the free/bound distinction for characterizing the environment where the phonological process applies, it is even more clear that appealing to lexical categories is undesirable. However, for the other languages listed in Table 102.3 above, either a base-identity account or a category-specific account appears to be feasible; the choice may come down to theory-internal considerations.

In any case, the free/bound distinction cannot be the source of all category-specific effects. Some languages show category-specific phonology that cannot be handled in terms of differences between morphologically free and bound forms. Examples are discussed in the following section.

4.1.2 *Mismatches with the free/bound distinction*

The difference between free and bound roots is not always consistent with phonological differences between nouns and verbs. This is true when both nouns and verbs are bound, and when both nouns and verbs are free – or, more generally, when some free forms have more phonological privilege than others.

Spanish (§2.1.1) provides evidence that lexical category differences in phonology are possible even among bound roots. It is true that verbs are always bound, while nouns and adjectives need not be. Crucially, however, the lexically contrastive antepenultimate stress pattern occurs even on noun and adjective stems that consist of a bound root and a (productive) gender suffix.

(20) *Spanish bound roots with antepenultimate stress* (data from Castillo and Bond 1948; Solá 1981)

		masculine	*feminine*	
a.	*nouns*	[ˈnawfraɣ-o]	[ˈnawfraɣ-a]	'shipwrecked person'
		[ˈbiɣam-o]	[ˈbiɣam-a]	'bigamist'

b.	*adjectives*	['loβreɣ-o]	['loβreɣ-a]	'murky, dismal'
		['prosper-o]	['prosper-a]	'prosperous'
		[me'toðik-o]	[me'toðik-a]	'methodic'
		['benet-o]	['benet-a]	'Venetian'
		[su'perflu-o]	[su'perflu-a]	'superfluous'
		[pur'pure-o]	[pur'pure-a]	'purple'
		[simul'tane-o]	[simul'tane-a]	'simultaneous'

If noun- or adjective-specific contrast in Spanish were dependent on nouns and adjectives being able to appear as unaffixed forms, these antepenultimate examples would not be possible.

Chuukese (§2.2) provides further evidence that category-specific effects are not always due to the free/bound distinction. In this language, both nouns and verbs may appear unaffixed,[8] and yet only nouns are subject to a minimality requirement. Similarly, Mandarin (§3.2) shows a three-way difference between nouns, adjectives, and verbs in reduplication patterns, but this is a language with essentially no inflectional morphology at all.

An interesting case is the category-specific nature of stress in Hebrew (discussed in §3.2). Both nouns and adjectives may be atemplatic, i.e. morphologically free, but verbs may not. Arguably, status as an atemplatic form is precisely what correlates with the capability for a noun or adjective to take the fixed stress pattern (Becker 2003), because even nouns and adjectives predictably have mobile stress if they are templatic. However, the free/bound distinction cannot account for the further difference between free nouns, in which the location of fixed stress is phonologically contrastive, and free adjectives, in which fixed stress always falls on the root-final syllable.

As noted in §2.4.3, McCarthy (2005) describes a noun/verb difference in template shape in Classical Arabic. McCarthy analyzes this pattern using Optimal Paradigms Theory, a variation on the base-identity approach (§4.1.1), in which similarity is enforced among members of a paradigm even in the absence of a free-standing base form. Bobaljik (2008) calls into question whether McCarthy's approach fully accounts for the Arabic pattern, noting, for example, that it predicts contrasts between noun and verb stem shapes that should be able to emerge under derivation, even if not within inflectional paradigms (see Bobaljik 2008: §3.2.2 for detailed discussion). Regardless of the success of a (quasi-)base-identity approach to template shape, however, it is important to note that McCarthy's analysis, which replaces reference to lexical category with reference to facts about affix shape and template shape, does not address *all* category-specific effects in Arabic. As seen in §2.1.2, the fact remains that nouns may be templatic or atemplatic, but verbs must be templatic. In other words, Arabic nouns still allow a greater degree of phonological contrast in their prosodic shape than verbs do, in a way that stands outside the template system and therefore cannot be derived from the differences between affix inventories for templatic nouns and verbs.

Finally, Bobaljik (2008), discussing Itelmen (described in §2.4.3 above), argues specifically against the attempt to recast all category-specific effects in terms of

[8] As seen in §2.2, both categories are also subject to a final-mora truncation process, so roots do not in fact surface unaltered. However, this process does not distinguish between nouns and verbs, so it cannot be the source of the category-specific phonological difference.

base identity. He shows that some verbs are derived from "category-neutral" roots, which can also be nouns and therefore have a free base form. Nevertheless, these verbs still follow the general verb pattern of schwa epenthesis. This is similar to Chuukese, where verbs are no less free than nouns, but nouns are phonologically privileged. Conversely, some Itelmen nouns have obligatory singular inflection and are therefore not free, but they nevertheless follow the noun pattern of schwa epenthesis. This is similar to Spanish, where nouns and adjectives have greater phonological freedom even when bound. As Bobaljik argues, these two types of mismatch between noun/verb and free/bound show that, in Itelmen, lexical category predicts phonological patterning more accurately than the free/bound distinction does.

In summary, the fact that nouns often appear phonologically privileged as compared to verbs may well be related at some fundamental level to the fact that nouns are cross-linguistically more likely to be free forms. However, attempts to relate these two asymmetries directly, in frameworks that make crucial use of the bound/free distinction to invoke the existence of a morphological base form that accounts for special aspects of noun phonology – or even frameworks that simply make use of differences in the inventories of inflectional morphemes for the two categories to account for phonological differences – are unable to capture the full range of category-specific phonological effects.

4.2 *Inflectional morphology*

As noted in §3, whether adjectives pattern phonologically with nouns or verbs shows a striking correlation with whether adjectives inflect for nominal or verbal categories. A particularly interesting case is Lenakel (see §2.4.1), where adjectives take verbal inflection when they are predicates, but not when they modify nouns (Lynch 1975, 1978). The presence or absence of verbal morphology on adjectives probably determines whether they take on the stress pattern of nouns or that of verbs (John Lynch, personal communication).

Nonetheless, this *correlation* between inflection type and category-specific phonology is not an *explanation*. For example, in Spanish or Mono, it is true that verb stress or tone is entirely determined by the inflectional paradigm (§2.1.1). However, the fact that verbs take inflectional suffixes does not preclude the logical possibility that verb roots might have underlying stress or tone contrasts (which might emerge in some particular inflectional form). That this is often not the case requires explanation; apparently, the phonological grammar does need to enforce the lack of contrast in verbs as a property separate from the fact that individual verbal inflectional morphemes happen to assign stress or tone.

Furthermore, languages like Mandarin, Hebrew, and Finnish (see §3.2) show that adjectives sometimes behave differently from both nouns and verbs – even if they have N-type inflection, as in Hebrew and Finnish.

4.3 *Prosodic structure*

Some apparent category-specific effects can be attributed to prosodic structure. For example, in Digo (Narrow Bantu; Kisseberth 1984), tones originating with the verb may end up on a following noun. However, Kisseberth shows that this is caused by phrase-level tonal phonology; verb tones surface on syllables within the noun because tone-assignment rules refer to the right edges of phonological phrases (CHAPTER 50: TONAL ALIGNMENT).

Kelly (1988) argues that the different stress preferences for disyllabic nouns and verbs in English (§2.4.1) originate in their syntactic and prosodic contexts; nouns prefer initial stress because they are typically preceded by an unstressed determiner, and sequential alternation between stressed and unstressed syllables is desirable. Verbs occur in a distinct syntactic and prosodic context, so they prefer final stress.

While explanations based on prosodic structure may cover some apparent cases of category-specific behavior, however, this approach cannot handle all the diverse cases discussed in §2 and §3.

4.4 *Morpheme-specific effects*

Whether or not they have category-specific phonology, languages generally have morphemes or morpheme classes that exhibit exceptional behavior (Saciuk 1969; see also the discussion in CHAPTER 103: PHONOLOGICAL SENSITIVITY TO MORPHOLOGICAL STRUCTURE). So, are category-specific effects meaningfully distinct from morpheme-specific effects? Is there a qualitative difference between assigning certain morphemes a phonology-relevant feature named "[−native]" and assigning certain morphemes a phonology-relevant feature named "[+noun]"? While "[−native]" and "[+noun]" might participate formally in the phonological grammar in much the same way, there are arguably important differences between the two. First, lexical categories do, obviously, have significance outside the phonology. If items of the same morphosyntactic category also pattern together phonologically, then allowing the phonology to use the morphosyntactic label captures a generalization that would be missed if an arbitrary, phonology-specific feature were invoked instead. Second, to the extent that the privilege scale N > A > V (§2, §3) is a linguistic (near-)universal, using the morphosyntactic category labels to demarcate phonologically relevant classes predicts their relative ability to support phonological contrast in a way that using arbitrary phonological labels does not.

4.5 *Summary: Alternatives to category-specific phonology*

While there are instances of category-specific behavior that may be accounted for by morphological, prosodic, or other factors, there remains a core of cases that do appear to require reference to lexical category within the phonology.

5 Conclusions

In this examination of category-specific phonological phenomena, a number of patterns have emerged. Many, although perhaps not all, cases are consistent with a universal scale of phonological privilege, N > A > V. Furthermore, the overwhelming majority of cases involve prosodic and suprasegmental phenomena rather than segmental or featural phenomena. Finally, there appear to be correlations between phonological behavior and type of inflection, seen especially in the case of adjectives. However, purely morphological or prosodic factors do not provide adequate accounts for all instances of category-specific phenomena, indicating that the phonological grammar must be able to refer to lexical categories.

In addition to phonology, there are other ways in which lexical categories show differences beyond morphosyntax, especially in child language acquisition and in psycholinguistics; for recent reviews, see e.g. Ogura *et al.* (2006) or D'Odorico and Fasolo (2007) for acquisition, and Rapp and Caramazza (2002) or Mätzig *et al.* (2009) for psycholinguistic evidence from aphasic speakers. Perhaps future research will uncover ways in which category-specific phonology is related to other sources of differentiated behavior among words of different categories.

The full array of facts about category-specific phonology – including the intermediate status of adjectives; the fact that nouns are more likely to be phonologically privileged than verbs, but in some languages both nouns and verbs are subject to equally predictable but nevertheless distinct requirements; and the fact that category-specific phonological differences often parallel, but do not fully match, other differences between categories, such as free-morpheme status – has not yet been captured by any single theoretical approach, including noun faithfulness (Smith 1997, 2001) and the various implementations of base identity or paradigm uniformity (e.g. Shiraishi 2004; McCarthy 2005; Cable 2005; Bobaljik 2008). An intriguing challenge remains for phonological theory.

ACKNOWLEDGMENTS

Many thanks for discussion and suggestions to Adam Albright, Randall Hendrick, F. Roger Higgins, John Kingston, John Lynch, John McCarthy, Elliott Moreton, Keren Rice, Elisabeth Selkirk, Ellen Woolford, and two anonymous reviewers. Language classifications are from Lewis (2009).

REFERENCES

Albright, Adam. 2008. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.) *Linguistics universals and language change*, 144–181. Oxford: Oxford University Press.

Albright, Adam & Yoonjung Kang. 2008. Predicting innovative alternations in Korean verb paradigms. Paper presented at the 18th International Congress of Linguists, Seoul.

Ansre, Gilbert. 1961. *The tonal structure of Ewe*. Hartford: Hartford Studies in Linguistics.

Anttila, Arto. 2002. Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20. 1–42.

Bakker, Peter. 1997. *A language of our own*. Oxford: Oxford University Press.

Bat-El, Outi. 1994. Stem modification and cluster transfer in Modern Hebrew. *Natural Language and Linguistic Theory* 12. 571–596.

Becker, Michael. 2003. Lexical stratification of Hebrew: The disyllabic maximum. In Yehuda N. Falk (ed.) *Proceedings of the Israeli Association of Theoretical Linguistics 19*. Available at http://linguistics.huji.ac.il/IATL/19/TOC.html.

Beckman, Jill N. 1999. *Positional faithfulness: An optimality theoretic treatment of phonological asymmetries*. New York: Garland.

Benua, Laura. 2000. *Transderivational identity: Phonological relations between words*. New York: Garland.

Bobaljik, Jonathan D. 1998. Mostly predictable: Cyclicity and the distribution of schwa in Itelmen. In Vida Samiian (ed.) *Proceedings of the 26th Western Conference on Linguistics*, 14–28. Fresno: Department of Linguistics, California State University.

Bobaljik, Jonathan D. 2008. Paradigms (Optimal and otherwise): A case for skepticism. In Asaf Bachrach & Andrew Nevins (eds.) *Inflectional identity*, 29–54. Oxford: Oxford University Press.

Broselow, Ellen. 1982. On predicting the interaction of stress and epenthesis. *Glossa* 16. 115–131.

Cable, Seth. 2005. Phonological noun–verb dissimilarities in Optimal Paradigms. Unpublished ms., MIT. Revised version of paper presented at the Workshop on (Non)identity within a Paradigm, MIT.

Castillo, Carlos & Otto F. Bond (eds.) 1948. *The University of Chicago Spanish dictionary*. New York: Pocket Books.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Cohen, Marcel. 1964. Catégories de mots et phonologie. In Carroll E. Reed (ed.) *Etudes phonologiques dédiées à la mémoire de M. le Prince N. S. Trubetzkoy*, 36–42. Alabama: University of Alabama Press. Reprint with new preface of *Travaux du Cercle Linguistique de Prague* 8 (1939).

Crowhurst, Megan J. 1992. Minimality and foot structure in metrical phonology and prosodic morphology. Ph.D. dissertation, University of Arizona.

Crowley, Terry. 1997. *An introduction to historical linguistics*, 3rd edn. Oxford: Oxford University Press.

de Lacy, Paul. 2001. Markedness in prominent positions. *MIT Working Papers in Linguistics* 40. 53–66.

de Lacy, Paul. 2004. Markedness conflation in Optimality Theory. *Phonology* 21. 145–199.

Devine, A. M. & Laurence D. Stephens. 1994. *The prosody of Greek speech*. New York: Oxford University Press.

Dixon, R. M. W. 1991. Mbabaram. In R. M. W. Dixon & Barry J. Blake (eds.) *Handbook of Australian languages*, vol. 4, 348–402. Oxford: Oxford University Press.

D'Odorico, Laura & Mirco Fasolo. 2007. Nouns and verbs in the vocabulary acquisition of Italian children. *Journal of Child Language* 34. 891–907.

Feng, Guanjun Bella. 2003. Lexical category specific constraints: Mandarin verb versus adjective reduplication. *USC Working Papers in Linguistics* 1. 1–12.

Garrett, Susan. 1996. Another look at Spanish stress and syllable structure. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 32. 61–75.

Glinert, Lewis. 1988. *The grammar of Modern Hebrew*. Cambridge: Cambridge University Press.

Goodenough, Ward & Hiroshi Sugita. 1980. *Trukese–English dictionary*. Philadelphia: American Philosophical Society.

Haraguchi, Shōsuke. 1977. *The tone pattern of Japanese: An autosegmental theory of tonology*. Tokyo: Kaitakusha.

Harris, James W. 1983. *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: MIT Press.

Hirayama, Teruo. 1960. *Zenkoku akusento ziten* [*All Japan accent dictionary*]. Tokyo: Tôkyôdô.

Kelly, Michael H. 1988. Phonological biases in grammatical category shifts. *Journal of Memory and Language* 27. 343–358.

Kenstowicz, Michael. 1996. Base identity and uniform exponence: Alternatives to cyclicity. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, 365–394. Salford: ESRI.

Kenstowicz, Michael & Charles W. Kisseberth. 1977. *Topics in phonological theory*. New York: Academic Press.

Kiparsky, Paul. 1982. From Cyclic Phonology to Lexical Phonology. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part I, 131–175. Dordrecht: Foris.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–365.

Kisseberth, Charles W. 1984. Digo tonology. In G. N. Clements & John A. Goldsmith (eds.) *Autosegmental studies in Bantu tone*, 105–182. Dordrecht: Foris.

Kisseberth, Charles W. & David Odden. 2003. Tone. In Derek Nurse & Gérard Phillipson (eds.) *The Bantu languages*, 59–70. London: Routledge.

Krause, Scott. 1979. Topics in Chukchee phonology and morphology. Ph.D. dissertation, University of Illinois, Urbana-Champaign.

Kubo, Tomoyuki. 2008. A sketch of Sibe phonology. *Gogaku Kenkyuu Fooramu* 16. 127–142.

Lewis, M. Paul (ed.) 2009. *Ethnologue: Languages of the world*, 16th edn. Dallas: SIL International. Available at www.ethnologue.com/.

Lynch, John D. 1975. Lenakel phonology. *University of Hawaii Working Papers in Linguistics* 7. i–vii, 1–244.

Lynch, John D. 1978. *A grammar of Lenakel.* Canberra: Australian National University.

Mätzig, Simone, Judit Druks, Jackie Masterson & Gabriella Vigliocco. 2009. Noun and verb differences in picture naming: Past studies and new evidence. *Cortex* 45. 738–758.

McCarthy, John J. 2005. Optimal paradigms. In Laura J. Downing, T. A. Hall & Renate Raffelsiefen (eds.) *Paradigms in phonological theory*, 170–210. Oxford: Oxford University Press.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese.* The Hague & Paris: Mouton.

Muller, Jennifer. 1999. A unified mora account of Chuukese. *Proceedings of the West Coast Conference on Formal Linguistics* 18. 393–405.

Myers, Scott. 2000. Boundary disputes: The distinction between phonetic and phonological sound patterns. In Noel Burton-Roberts, Philip Carr & Gerard J. Docherty (eds.) *Phonological knowledge: Conceptual and empirical issues*, 245–272. Oxford: Oxford University Press.

Ogura, Tamiko, Philip S. Dale, Yukie Yamashita, Toshiki Murase & Aki Mahieu. 2006. The use of nouns and verbs by Japanese children and their caregivers in book-reading and toy-playing contexts. *Journal of Child Language* 33. 1–29.

Olson, Kenneth S. 2005. *The phonology of Mono.* Dallas: SIL International & University of Texas at Arlington.

Paster, Mary. 2000. Issues in the tonology of Gã. Honors thesis, Ohio State University.

Postal, Paul. 1968. *Aspects of phonological theory.* New York: Harper & Row.

Prince, Alan. 1980. A metrical theory for Estonian quantity. *Linguistic Inquiry* 11. 511–562.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar.* Malden, MA & Oxford: Blackwell.

Rapp, Brenda & Alfonso Caramazza. 2002. Selective difficulties with spoken nouns and written verbs: A single case study. *Journal of Neurolinguistics* 15. 373–402.

Rosen, Nicole. 2007. Domains in Michif phonology. Ph.D. dissertation, University of Toronto.

Ryding, Karin C. 2005. *A reference grammar of Modern Standard Arabic.* Cambridge: Cambridge University Press.

Saciuk, Bohdan. 1969. The stratal division of the lexicon. *Papers in Linguistics* 1. 464–532.

Shiraishi, Hidetoshi. 2004. Base-identity and the noun–verb asymmetry in Nivkh. In Dicky Gilbers, Maartje Schreuder & Nienke Knevel (eds.) *On the boundaries of phonology and phonetics*, 159–182. Groningen: University of Groningen.

Smith, Jennifer L. 1997. Noun faithfulness: On the privileged behavior of nouns in phonology. Unpublished ms., University of Massachusetts, Amherst. Available (July 2010) at www.unc.edu/~jlsmith/home/pdf/nfaith97.pdf.

Smith, Jennifer L. 2001. Lexical category and phonological contrast. In Robert Kirchner, Joe Pater & Wolf Wikely (eds.) *Papers in experimental and theoretical linguistics 6: Workshop on the Lexicon in Phonetics and Phonology*, 61–72. Edmonton: University of Alberta.

Smith, Jennifer L. 2002. Phonological augmentation in prominent positions. Ph.D. dissertation, University of Massachusetts, Amherst. Published 2005, New York: Routledge.

Solá, Donald F. (ed.) 1981. *The Random House basic dictionary: Spanish–English, English–Spanish*. New York: Ballantine Books.

Whitman, John. 1994. The accentuation of nominal stems in Proto-Korean. In Young-Key Kim-Renaud (ed.) *Theoretical issues in Korean linguistics*, 425–439. Stanford: CSLI.

Wolff, Ekkehard. 1983. *A grammar of the Lamang language*. Glückstadt: J. J. Augustin.

103 Phonological Sensitivity to Morphological Structure

Jochen Trommer

1 Introduction

Virtually all areas of phonology show extensive sensitivity to morphological structure. Many segmental and tonal processes are restricted to apply only across morpheme boundaries (or only morpheme-internally). Computation of stress typically reflects morphological constituency, and phonotactic generalizations often seem to be restricted to a subset of the morphological material of a given language (e.g. affixes). Thus one of the central questions research on the morphology–phonology interface addresses is what types and amount of morphological structure phonology can access and what consequences this has for phonological processes.

The chapter is structured as follows: in §2, I will introduce the different types of sensitivity which phonology shows to morphological structure. §3 summarizes different ways of reflecting these types of sensitivity in the architecture of the morphology–phonology interface, and §4 discusses possible restrictions on phonological sensitivity to morphology.

2 Types of phonological sensitivity to morphological structure

2.1 *Sensitivity to the presence of a morpheme or morpheme type*

Probably the simplest type of sensitivity phonology can show to morphological structure is the capacity to detect the presence of a morpheme, i.e. to determine for a specific piece of phonological structure which specific morpheme (if any) it belongs to, and which morphological type the morpheme is (noun or verb, root or affix, etc.)

A number of phonological processes are simply sensitive to whether segmental material is part of a morpheme or not. For example, Mohawk (Michelson 1989; Piggott 1998) and Selayarese (Mithun and Basri 1986; Piggott 2001) stress the penultimate syllable of a word if the last syllable contains a vowel with morphological affiliation (1a), whereas stress appears on the antepenultimate if the vowel heading the final or penultimate syllable is not part of a morpheme, i.e. it is epenthetic; see CHAPTER 67: VOWEL EPENTHESIS and (1b) (Selayarese examples from Mithun and Basri 1986: 240; Mohawk examples from Michelson 1989: 41, 44, epenthetic vowels are underlined):[1]

(1)	a.	*Mohawk*	/k-atir-ut-haʔ/ 1SG-ACT-pull-HABIT	[kati'ruthaʔ]	'I pull it'
		Selayarese	/sahala/	[sa'ha:la]	'sea cucumber'
	b.	*Mohawk*	/ʌ-k-r-ʌʔ/ FUT-1SG AGT-fill into-PUNCT	['ʌkerʌʔ]	'I will put it into a container'
		Selayarese	/sahal/	['sa:hala]	'profit'

A typical locus of phonological sensitivity to morpheme type is segmental inventories which are different for different types of morphemes. Thus German inflectional affixes only contain coronal or nasal consonants and the reduced vowel /ə/, whereas lexical roots and derivational affixes also employ labial and dorsal consonants and 23 different vowels. Under the assumption that this asymmetry follows from phonological rules or constraints (e.g. morpheme structure constraints or OT-markedness constraints, cf. CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS), the phonology must minimally "know" which segments belong to inflectional affixes and which do not, to impose the appropriate restrictions on specific morphemes.

Sensitivity to morpheme type is also pervasively attested in effects on prosody and word-minimality. For example, German inflectional affixes can never be stressed, whereas lexical roots have obligatory stress. In the same vein, lexical roots in German have the minimal size of a bimoraic syllable (Golston and Wiese 1998), whereas inflectional affixes consist of light syllables or single consonants/consonant clusters.

A typical case where sensitivity to morpheme type shows up in an alternation context are harmony processes based on an asymmetry between roots and affixes (CHAPTER 104: ROOT–AFFIX ASYMMETRIES). Thus in Wolof, a Niger-Congo language of the Atlantic sub-branch, root and affix vowels agree in the feature [ATR], where the underlying [ATR] feature of roots is systematically retained whereas affix vowels adjust to the root (Pulleyblank 1996: 314–315):

(2)	+ATR		−ATR	
	gən-e	'be better in'	xam-ɛ	'know in'
	re:r-e	'be lost in'	dɛm-ɛ	'go with'
	do:r-e	'hit with'	xɔl-ɛ	'look with'

A standard optimality-theoretic interpretation of these data is that they exhibit positional faithfulness, i.e. faithfulness constraints referring to segments which are part of a root (or root-initial) are ranked higher than faithfulness constraints

[1] Under a derivational account, it could be argued for Mohawk and Selayarese that stress assignment is simply ordered before vowel epenthesis. See Alderete (1999) for detailed arguments that rule ordering accounts for stress–epenthesis interactions of this type are generally problematic.

for affix vowels or general faithfulness constraints (Beckman 1998; McCarthy and Prince 1995). But, whatever the specific implementation is, if the intuition is right that root vowels act differently from phonologically identical affix vowels, then the phonology must be able to access the information as to which vowel belongs to a root and which one to an affix.

An obvious question is to which properties of morphemes phonology can be sensitive. It is well established that phonology tends to reflect the affix/root distinction. A related relevant distinction is that between lexical words and function words (Selkirk 1995). Languages often also treat prefixes and suffixes differently for prosodic restrictions and segmental processes (see Hyman 2008, and references cited there). CHAPTER 102: CATEGORY-SPECIFIC EFFECTS shows that many languages also differentiate phonologically between nouns and verbs. Phonological distinctions often also refer to etymological strata (Itô and Mester 1999; CHAPTER 95: LOANWORD PHONOLOGY) or to arbitrary lexical or exceptional classes (CHAPTER 106: EXCEPTIONALITY). Thus in Finnish, root-final /a/ before suffixal /i/ is raised to [o] in specific roots (3a), while it is deleted with other roots (3b), even though the roots do not differ in any crucial phonological aspect; (3c) is an example of a third class of roots, which sometimes show deletion and sometimes raising (Pater 2009: 9, based on Anttila 2002):

(3)	a.	/tavara-i-ssa/	[tavaroissa]	'thing (PL INESS)'
	b.	/jumala-i-ssa/	[jumalissa]	'God (PL INESS)'
	c.	/itara-i-ssa/	[itaroissa] ~ [itarissa]	'stingy (PL INESS)'

2.2 Sensitivity to morpheme boundaries

Sensitivity to morpheme boundaries can be illustrated with *s*-dissimilation in Swabian German. In Swabian, the contrast between /s/ and /ʃ/ (4a) is neutralized to [ʃ] before all instances of /t/ and /p/ (4b) (Hall and Scott 2007: 154):[2]

(4)		*Standard German*	*Swabian*	
a.	*Saal*	[za:l]	[sa:l]	'hall'
	Schale	[ʃa:lə]	[ʃa:l]	'bowl'
	Bus	[bʊs]	[bʊs]	'bus'
	Rausch	[rauʃ]	[rɔuʃ]	'intoxication'
b.	*Speize*	[ʃpaizə]	[ʃpais]	'food'
	Haspel	[haspə]	[haʃpəl]	'hasp'
	Post	[pɔst]	[pɔʃt]	'mail'
	Fenster	[fɛnstɐ]	[fɛnʃtɐ]	'window'

However, if there is a morpheme boundary between underlying /s/ and /t p/ (or /b/) the process is blocked (Hall and Scott 2007: 158):

[2] As shown by the first example in (4a), the /s/~/z/ contrast of Standard German is neutralized to /s/ in Swabian. Standard German also exhibits s-dissimilation, but only word-initially (cf. the first example in (4b)).

(5)

	Swabian	
Kinder-Brei	[kendlesbrəi]	'pap'
miss-trauen	[mistrəuə]	'mistrust'
pass-t	[past]	'fit-3SG'
ge-wuss-t	[gvist]	'KNOW-PAST PART'

S-dissimilation is an instance of what Wolf (2008) calls a *non-derived environment* effect: a specific process only applies if the trigger and the target of the process are not separated by a morpheme boundary. The opposite type of restriction also occurs for specific processes in so-called *derived-environment effects* (or *non-derived environment blocking*: Kiparsky 1993; CHAPTER 88: DERIVED ENVIRONMENT EFFECTS). Thus in Finnish assibilation, /t/ is fricativized to [s] before /i/ (6a, 6c), but only if there is a morpheme boundary between /t/ and /i/ (6b, 6c) (Kiparsky 1973, 1993):[3]

(6)

a.	/halut-i/	[halusi]	'want-PAST'
b.	/koti/	[koti]	'home'
c.	/tilat-i/	[tilasi]	'order-PAST'

Morpheme boundaries also play a role in syllabification and processes which are sensitive to syllabification (cf. e.g. Nespor and Vogel 1986; Booij 1995; Raffelsiefen 2005; and the contributions in Hall and Kleinhenz 1999). Thus Kager (1999: 136) argues that Diola-Fogny shows a preference for aligning the left edge of a morpheme with the left edge of a syllable, which accounts for the fact that the language deletes morpheme-final, not morpheme-initial consonants to avoid a syllable-final stop (underlying /let-ku-jaw/ gives [le.ku.jaw], not *[let.ku.jaw] or *[le.tu.jaw]).[4]

It is important to distinguish sensitivity to boundaries from sensitivity to boundary symbols, a theoretical device which plays a crucial role in early rule-based phonology. Thus in Chomsky and Halle (1968; *SPE*) it is assumed that the agentive suffix *-er* in *singer* (/sıng#ər/ → [sıŋər]) is separated from its base by a different boundary symbol from the one separating comparative *-er* from its base in *longer* (/lɔng+ər/ → [lɔŋgər]), accounting for the fact that deletion of stem-final /g/ is obligatory before agentive, but not before comparative *-er*. Different types of morphemes are separated by different types of boundary symbols even though the morphological boundary is in both cases a morpheme boundary. However, this type of boundary sensitivity can in most cases be reconstructed as a combination of sensitivity to morpheme type and sensitivity to morpheme boundaries. Thus we might say that *g*-deletion in English is blocked by a morpheme boundary involving the lexical class of morphemes instantiated by comparative *-er* (so-called level 1 affixes), but not if the morpheme boundary is adjacent to a level 2 affix.[5] More generally, sensitivity to morpheme boundaries might be understood as a special case of sensitivity to

[3] See Wolf (2008) for a derivational view on non-derived environment blocking which does not explicitly refer to boundaries, and a discussion of cases of non-derived environment blocking which might not be reduced to sensitivity for morpheme boundaries.

[4] See McCarthy (2008) for a recent analysis of Diola which does not refer to morphological structure, and for references to other analyses of the same data.

[5] Boundary symbols will be discussed in more detail in §2.2.

morphemes. Thus with respect to Finnish assibilation, the statements in (7a) and (7b) are equivalent:[6]

(7) a. Assibilation happens only if /t/ and /i/ are separated by a morpheme boundary.
b. Assibilation happens only if /t/ and /i/ are part of different morphemes.

2.3 *Sensitivity to the morphological function of a morpheme*

2.3.1 *Heads* vs. *non-heads*

Revithiadou (1999: 5) argues that the selection of lexical stress in lexical accent systems such as Russian (8) is determined by the morphological head of the construction. Thus in (8a) the head is the lexical root, not the inflectional affix, therefore the underlying stress of the root survives; in (8b) the head is the derivational suffix, therefore affixal stress survives:

(8)	a.	/ʧ'eʧ'e'vits-'a/	ʧeʧe'vits-a	'lentil-NOM SG'
	b.	/'gorl-'ast-'a/	gorl-'ast-a	'loud-mouthed-NOM SG-FEM'

2.3.2 *Adjuncts* vs. *complements*

Bachrach and Wagner (2007) argue that phonological processes might be sensitive to the distinction between affixes which are morphosyntactically heads and those which are morphosyntactically adjuncts (see also Newell 2005 for a similar analysis of bracketing paradoxes). Their crucial empirical observation is that diminutive affixes behave phonologically differently from other affixes in Brazilian Portuguese (see also Vigário 2003 for the same observation). For example, whereas all derivational suffixes attract stress, nasalization of stressed vowels before nasal consonants is lost with other suffixes (9a), but not with diminutive suffixes (9b).

(9)	a.	'*f*[ã]*ma*	'fame'	*f*[a]'*moso*	'famous'
	b.	'*k*[ã]*ma*	'bed'	*k*[ã]'*miña*	'small bed'
				ˌ*k*[ã]*ma*'*ziña*	'small bed'

Bachrach and Wagner (2007) relate the different behavior of diminutive affixes to the fact that they behave morphologically as adjuncts – evidenced, for example, by the fact that they don't change part of speech or gender of the base to which they attach – while other derivational affixes are heads taking their bases as complements. On the assumption that adjuncts are assigned stress independently from the constituents they select, they argue that bases maintain independent main stress with diminutives, which is only changed by late phonological processes. Hence base vowels are stressed at the relevant phonological level and are not denasalized.

[6] The idea that morphological affiliation replaces morpheme boundaries is originally due to McCarthy (1979). See van Oostendorp (2007) for a theory of derived environment effects which is based on this intuition.

2.4 *Sensitivity to stems*

Any discussion of stems faces the problem that the literature invokes very different notions of "stem." One usage of this term is that every morphologically complex structure containing a lexical root morpheme is a stem (Downing 2006). I will call this the *morphological stem*. Thus in the word form *naturalists* we might isolate the morphological stems *nature, natural, and naturalist*. Under a second reading, "stem" is intended as a constituent which contains a root and all derivational affixes of a word form, hence in *naturalists* only *naturalist* would count as a stem. I will call this type of stem *derivational stem*. A third reading of "stem" is often found in phonological work where stem is a designated morphological stem which is relevant to the phonology. Thus in Stratal OT (Bermúdez-Otero, forthcoming), it is assumed that every stem headed by a level 1 affix and not selected by a level 1 affix forms a phonological stem, which is an arbitrary choice from the point of view of the morphology. I will call this type of stem the *phonological stem* (see §3.2 for more discussion).

For all three types of stems it has been argued that they are relevant for phonology. This is most obvious for the phonological stem. Thus, in English, many processes, such as nasal place assimilation, stress assignment, and trisyllabic laxing, take place in the phonological stem as defined above. Downing (2006) argues that in many languages morphological stems have to obey prosodic minimality requirements not applying to non-complex words. Thus German infinitives (e.g. /geː-ə-n/ 'to go', with epenthetic schwa) must be at least bisyllabic, whereas bare verb roots and other roots (e.g. the imperative /geː/ 'go!' and the numeral /tseːn/ 'ten') can be monosyllabic.[7] In Albanian, the relevant domain of word stress is the morphological stem (Trommer 2008b; see §2.5 for examples), and Hyman (2008) shows that in many Bantu languages processes of vowel, consonant, and tone harmony apply to the morphological stem as well as prosodic restrictions on minimal or maximal size.

However for most cases, diagnosing sensitivity to stems is conceptually problematic, because the phonological sensitivity to stems is by no means unambiguous. Thus most of the effects discussed in Downing (2006) as evidence for the morphological stem are construction-specific. For example, the bisyllabicity requirement for German infinitives does not extend to other inflected forms such as the 3rd singular (e.g. /geː-t/ '(s/he) goes'), which also has an overt affix.

Once this is taken into account, cases like this might be reduced to affixation of prosodic material. Thus the restriction on the German infinitive could be derived by a bisyllabic foot template which is part of the infinitive suffix. Reference to the morphological or the phonological stem can often be replaced by reference to morpheme type. Thus, in a canonical Bantu language, it is tantamount to claiming that a specific phonological process applies only to derivational suffixes and to stating that it is restricted to a stem constituent.

Data which exhibit relatively clear-cut evidence for the phonological sensitivity to stem-like constituents seem to be restricted to the boundaries of morphology, to cases of non-concatenative morphology, and to specific cases of opacity

[7] There are only two high-frequency German verbs which have a monosyllabic infinitive, /tuːn/ 'to do' and /zaɪn/ 'to be' (Neef 1996: 135). The claim that schwa in German is in general epenthetic is defended in detail in Wiese (1996).

discussed in detail in §2.5. An example of the first type is stress assignment in English compounds. Consider for example the compounds in (10) (see also CHAPTER 116: SENTENTIAL PROMINENCE IN ENGLISH):

(10) a. [[*blackboard*]$_N$ *eraser*]$_N$
b. [*linguistics* [*language requirement*]$_N$]$_N$

In (10a), the first two nouns form the simple compound *blackboard*, which is in turn part of the complex compound *blackboard eraser*. Main stress is on the first word; secondary stress on the third. In (10b), the second and third nouns form an internal compound, and as a consequence main stress is on the second noun and secondary stress on the first one. To compute English compound stress clearly requires access to the bracketing of stem constituents, but compounding is notorious for being a boundary case between morphology and syntax.

A case of non-concatenative morphology/phonology in which bracketing seems to be crucially visible in the phonology is the tonal behavior of affixes in Hausa (Inkelas 1998). Hausa has two types of affixes: recessive (or tone-integrating) affixes add their underlying tone melody to the tone melody of the base to which they attach (11a), whereas dominant affixes replace the tone of the base by their own melody (11b) (all data originally from Newman 1986, 2000):

(11) a. *Recessive affix* (Inkelas 1998: 127)

ba-goobir	+	-ii	→	bagoobirii
L L H		H		L L HH
'from-Gobir'		ethonym		'a Gobir man'
ba-zamfara	+	-ii	→	bazamfarii
L L L L		H		L L L H
'from-Zamfara'		ethonym		'a Zamfara man'

b. *Dominant affix* (Inkelas 1998: 127)

ba-katsina	+	-ee	→	bakatsinee
L L HL		HL		HH H L
'from-Katsina'		ethonym		'a Katsina man'
ba-zamfara	+	-ee	→	bazamfaree
L L L L		HL		HH H L
'from-Zamfara'		ethonym		'a Zamfara man'

Crucially, the domain of tone integration or replacement is delimited by morphological bracketing, as is obvious from the examples in (12). In both forms, the dominant suffix /-ii/ imposes its tone pattern LH on the constituent [karanta + -ii], and the recessive prefix /ma-/ "adds" an additional initial H tone, resulting in the output form shown in (12a). In (12b), this form is embedded inside a further dominant affix, /-ijaa/, whose tone melody HLH overwrites everything in its scope. Thus both forms contain a final dominant affix, but in (12b) this is the outermost affix of the construction, whereas it is embedded inside of a (recessive) prefix in (12a). As a consequence, the final affix in (12b) replaces the tone patterns of the entire word, whereas it doesn't affect the prefix in (12a), which shows that it is not the phonologically linear position of a dominant affix that determines the extension domain of its tonal melody, but its hierarchical position in morphological structure (Inkelas 1998: 130):

(12)	a.	[ma- [karanta + -ii]]	→	makarancii	
		H HL H LH		H L L H	
		NML 'read' -AGENT		'reader (MASC)'	
	b.	[[ma- [karanta + -ii]] -ijaa]	→	makaranciijaa	
		H HL H LH HLH		H H H L H	
		NML 'read' -AGENT		'reader (FEM)'	

The degree to which phonological sensitivity to stems is a necessary assumption is thus a delicate question.

2.5 *Sensitivity to paradigmatic structure*

All types of sensitivity we have discussed so far refer to the syntagmatic structure of morphological expressions. An alternative way to approach morphology (and phonological sensitivity to morphology) is by paradigmatic relations between words and word forms (CHAPTER 83: PARADIGMS). Thus sensitivity to paradigmatic structure gives us a different view on the type of data already discussed rather than a new range of phenomena. For example, Hall and Scott (2007) assume that the restriction of Swabian s-dissimilation to monomorphemic contexts is due to the faithfulness of inflected forms to their paradigmatic bases: in an inflected form such as /griəs-t/ 's/he greets', /s/ is not dissimilated to /ʃ/, because this would disrupt featural identity to the base form /griəs/ 'I greet', which doesn't have /ʃ/, due to the absence of a following /t/. For words such as /poʃt/ 'post', s-dissimilation is also triggered in the base and therefore found in the whole inflectional paradigm, and for an inflectional suffix such as /-st/ there is no base form, so it freely undergoes dissimilation in a form as /doːf-ʃt/ 'most stupid'.

Like s-dissimilation, most phenomena which can be described as morpheme boundary effects are cases of opacity where processes are simply blocked in the absence or presence of boundaries ("underapplication"). In contrast, the paradigmatic view also extends to cases of opacity where a process applies even though the appropriate context is not present in the word form itself, but only in a paradigmatically related form ("overapplication"). For example, in Albanian, uninflected words ending in a mid-vowel have stress on the penultimate syllable (e.g. *'ba.bo* 'midwife'), whereas words ending in a closed syllable or non-mid vowels have final stress (e.g. *pa.'tok* 'gander' and *ash'tu* 'thus'). Crucially, inflected forms maintain the stress of their base word. Thus the definite form of *pa.'tok*, *pa.'to.ku*, has stress on the final syllable of the stem even though this syllable is open in the actual form, and the word ends in a high vowel. A paradigmatic account can capture facts like this by assuming that inflected forms must be faithful in their stress position to their bases (Trommer 2008b).

A type of sensitivity to morphology which seems to be systematically beyond the reach of a paradigmatic approach is that of derived environment effects such as Finnish assibilation discussed in §2.2. The restriction of a process to occurring across a morpheme boundary obviously doesn't serve the purpose of making the derived form more similar to its base. Even under the assumption that the derived environment effect is made in some way possible by the fact that the derived form gets more dissimilar from its base (see Kenstowicz 2005 for cases of effects which are triggered by the requirement to make paradigmatically related form less similar), this would not account for the fact that this effect does only

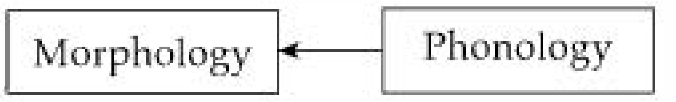


Figure 103.1 Sensitivity to morphology in a modular architecture

occur locally, at the morpheme boundary (recall that underlying /tilat-i/ becomes [tilas-i], not [silas-i] or [silat-i]).[8]

3 Sensitivity to morphology and the structure of the morphology–phonology interface

Considering the data discussed in §2, it is clear that many phonological processes reflect morphological structure. However, the idea that morphology as a whole communicates information to phonology as a whole is only one type of approach to morphologically conditioned phonology presupposing a modular architecture which cleanly separates morphology and phonology, as is schematically shown in Figure 103.1.

Under this approach, the nature and extent of sensitivity to morphology is fully determined by the single interface that the two modules phonology and morphology use to communicate. The nature of this interface will be discussed further in §3.3. Alternatives to this architecture are discussed in §3.2.

3.1 Fine-grained sensitivity to morphology in Optimality Theory

A much less clear-cut picture of the morphology–phonology interface is standard in most versions of Optimality Theory, where different constraint types have different amounts of access to morphological structure. Thus it is often implicitly assumed that standard markedness constraints are incapable of accessing any morphological information, while alignment constraints can access the position of morphological boundaries, and positional faithfulness constraints are sensitive to information on boundaries and affiliation of segments to morphological constituents, but only if these are prominent in some well-defined way (e.g. roots but not affixes). This amounts to an architecture where phonological constraint types have different access paths to morphological structure, as illustrated in Figure 103.2.

In fact, most proponents of OT go even further, denying that phonology and morphology are different modules, such that phonological and morphological constraints directly interact, and phonological constraints can enforce violations of morphological constraints (see e.g. Golston 1995; Wolf 2008). The question of whether abandoning morphophonological modularity in this way is empirically and conceptually justified goes beyond the topic of this chapter; but, also under a non-modular and fully parallel architecture, the access of phonological

[8] A paradigmatic account might be invoked to account for cases of derived environment effects which are non-local (Burzio 2000); see §4.2 for discussion.

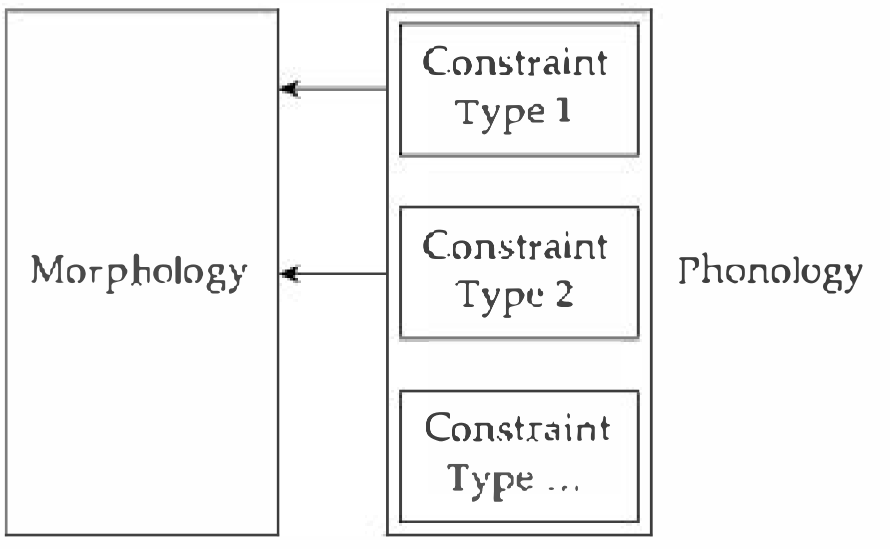


Figure 103.2 Sensitivity to morphology in OT

constraints to morphological structure follows the same principles as depicted in Figure 103.2. Note also that a strictly modular architecture of morphology–phonology interaction might be emulated by restrictions on specific constraint types. Thus Bermúdez-Otero (forthcoming) argues that the only phonological constraints which can access morphological information are alignment constraints for prosodic categories, whereas other constraints can only refer to prosodic categories. This reconstructs a derivational modular system, where prosodic categories serve as the interface between morphology and phonology.

3.2 *Sensitivity to morphology vs. external control: Co-phonologies and cyclicity*

An even more radical alternative approach to phonological processes which are sensitive to morphological structure is to assume that morphology governs phonology in a way which mirrors its own structure-building processes. In effect, phonology reflects morphological structure, but, without phonological rules or constraints having access to morphological information, phonology undergoes external control. Hence no morphological structure is accessed by the phonology. This approach is sketched schematically in Figure 103.3.

It is not the goal of this chapter to discuss arguments for this or for the sensitivity-based architecture, since they are partially equivalent, and many of the differences discussed in the literature depend on orthogonal assumptions which

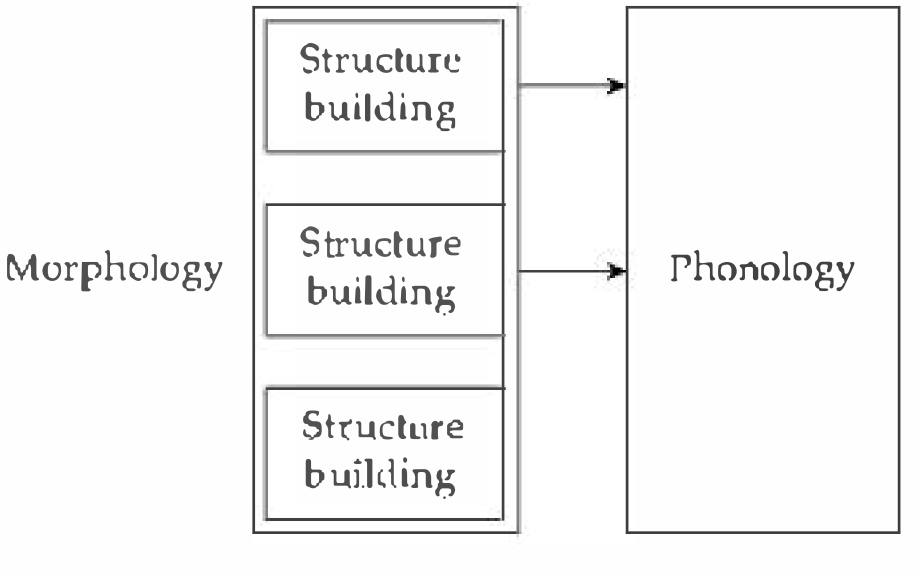


Figure 103.3 External control of phonology by morphology

advocates of these architectures make.[9] Moreover, the external-control approach and the sensitivity approach can be combined in a fruitful way. Here I will discuss some of the most important features of the external-control approach to morphophonology, but take a mostly sensitivity-based perspective in the rest of the chapter, keeping an eye on the alternative or complementary role external control might play.

A theory which consequently applies the idea of external morphological control to morphologically sensitive phonology is sign-based phonology and morphology (Orgun 1996; Inkelas and Zoll 2005). As an illustration, consider a case of exceptional phonology in Turkish, where velar stops are deleted intervocalically in the affixation of the dative suffix to a nominal stem, but not in the affixation of an aorist suffix to a verbal stem:

(13) a. *Dative suffix: Velar deletion*

nominative	*dative*	
bebek	bebe-e	'baby'
inek	ine-e	'cow'

b. *Aorist suffix: No velar deletion*

past	*aorist*	
gerek-ti	gerek-ir	'be necessary'
bırak-tı	bırak-ır	'leave'

A possible interpretation in terms of sensitivity to morphology would be that phonology is sensitive to the presence of the exceptional dative suffix. In the system of Pater (2009), this might take the form of a lexically indexed constraint, i.e. a phonological constraint which is by definition restricted to an arbitrary set of morphemes ***M***, or, put slightly differently, which only induces constraint violations for phonological structure that overlaps with phonological material of a morpheme in ***M***. For the case at hand we might assume a constraint *VKV$_{Dat}$, which doesn't allow /k/ in the context of a preceding and a following vowel, and is morphologically indexed for (restricted to) the dative morpheme. Since the phonological configuration inducing a constraint violation of *VKV$_{Dat}$ overlaps with an exponent of dative in (14), /k/ is deleted, whereas no constraint violation obtains in the absence of the dative, therefore the /k/ in (15) is protected by MAX-C:

(14) *Dative: By indexed constraints*

inek-e	*VKV$_{Dat}$	MAX-C
☞ a. ine-e		*
b. inek-e	*!	

[9] For example, it is often claimed that an important difference between the co-phonology approach (an external control approach) and the indexed constraint approach is that the former, but not the latter, can capture "markedness" reversals. However, as shown in detail in Inkelas and Zoll (2007), both frameworks can incorporate the disputed assumption (or refrain from doing so) with equal ease.

(15) *Aorist: By indexed constraints*

gerek-ir	$*VKV_{Dat}$	MAX-C
a. gere-ir		*!
☞ b. gerek-ir		

The principal idea of the sign-based approach is that different constructions such as dative and aorist forms simply invoke different phonologies. Thus the dative construction has *VKV ranked above MAX-C, resulting in intervocalic /k/-deletion (16), while the ranking of the aorist construction is the opposite, leading to maintenance of /k/ (17):

(16) *Dative: Co-phonology* Φ_1

inek-e	*VKV	MAX-C
☞ a. ine-e		*
b. inek-e	*!	

(17) *Aorist: Co-phonology* Φ_2

gerek-ir	MAX-C	*VKV
a. gere-ir	*!	
☞ b. gerek-ir		*

Crucially, the phonology in this system doesn't have direct access to morphological information. The constraints in (16) and (17) do not "know" by themselves that specific structures involve a dative or aorist affix. Instead, the morphology chooses the right phonology for specific contexts. This is achieved by an architecture of morphology where every construction is a hierarchical structure which fixes the phonology of the entire structure as a function of its substructures: thus the dative construction (18a) has the phonology which results from applying the ranking Φ_1 to the concatenation of the base noun and the dative suffix, while the aorist construction has the phonology resulting from applying Φ_2 to the concatenation of the base verb and the aorist suffix (18b):

(18) a. *Dative construction*

[Syntax = Dative N
  Phonology = $\Phi_1(P_x, /\text{-E}/)$]

[Syntax = N
  Phonology = P_x]   [/-E/]

b. *Aorist construction*

[Syntax = Aorist V
  Phonology = $\Phi_2(P_x, /\text{-r}/)$]

[Syntax = V
  Phonology = P_x]   [/-r/]

Co-phonologies and indexed constraints seem to be largely equivalent, but there are important differences in the specific implementations of locality restrictions on morphological accessibility which will be discussed in §4.2. Moreover, co-phonologies allow an alternative approach to opacity which largely obviates the use of paradigmatic constraints. Thus the fact that Albanian inflected word forms inherit stress from their base forms (e.g. the definite form *pa.'to.ku* from the indefinite form *pa.'tok* 'gander'; see §2.5) can be derived by assigning weight-sensitive co-phonologies to bare roots and derivational affixes, and co-phonologies which are faithful to their input stress to inflectional affixes (Trommer 2008b). Since the phonology of a complex expression is computed as a function of the phonology of the base, this effectively preserves base stress.

A different approach, where morphological structure is also reflected by different co-phonologies, is found in Stratal Optimality Theory (Kiparsky 2000; Bermúdez-Otero, forthcoming) and its predecessor Lexical Phonology (Kiparsky 1982; Mohanan 1986). In Stratal Optimality Theory, not every morphological construction has a specific co-phonology, but co-phonologies are attached to stems, words (and, going beyond morphology, to phrases). For example, in a word with a stem-level affix (*-ous*) and two word-level affixes (*-ness* and *-less*) such as *glori-ous-ness-less*, the stem *glori-ous* is evaluated by the stem-level phonology, and *glori-ous-ness-less* (the concatenation of the resulting stem with *-less* and *-ness*) by the word-level phonology (see also CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME).

Triggering of phonology by morphology and sensitivity to morphological information by phonology are not necessarily mutually exclusive options Thus proponents of Stratal Optimality Theory usually assume that there is some form of sensitivity to morphology in the form of the prosodic hierarchy (cf. §3.3). Stratal Optimality Theory might therefore be called a hybrid model of the morphology–phonology interface, because it integrates sensitivity and triggering. This hybridity is reminiscent of the architecture assumed in *SPE*, where phonology reflects syntax not only by the virtue of boundary symbols, but also through the cycle, a mechanism which subjects morphosyntactic chunks of increasing size to the same battery of phonological rules. Scheer (2009) also adopts a very similar architecture, where the use of boundary symbols (cf. §3.3) is supported by a version of the cycle in the form of recent phase theory (Chomsky 2001; see also CHAPTER 85: CYCLICITY).

3.3 *Dialects of phonology–morphology communication*

As pointed out by Scheer (2008), the situation of phonology–morphosyntax interactions is similar to the communication of speakers speaking different languages: morphology might communicate to phonology via morphological vocabulary, via phonological vocabulary, or via an interlanguage, i.e. a vocabulary which is neither proper to phonology nor to morphology. In this subsection I will discuss these different possibilities.

3.3.1 *Phonology and morphology speaking an interlanguage*

Probably the most influential approach to phonological sensitivity to morphology in the last decades is the idea that communication between phonology and morphology is based on an interlanguage which conveys morphological information to the phonology module. The classical version of this approach is proposed in *SPE*, where the interlanguage consists of boundary symbols such as the weak boundary "+" and the strong boundary "#." These symbols are inserted after the completion of morphosyntax, mirroring the hierarchical structure of syntax. Thus + is roughly speaking inserted at morpheme boundaries, and # at word boundaries, subject to further readjustment rules, and then interpreted in the phonology by phonological rules. Word-final devoicing in German can be captured by the rule [–son] → [–voice] / __ #. It is important to note that something like # is not really a syntactic symbol (although it is derived by reference to syntactic structure), since no syntactic rules refer to it or build structures containing it. While the positions of # assumed for English correspond roughly to pre-theoretic intuitions about English word boundaries, it doesn't correspond to morphosyntactic word boundaries in *SPE*, since *SPE* is a purely syntactic theory of morphosyntax, where words do not have any theoretical status at all. Word-like boundaries are only introduced to communicate with phonology; # is not a phonological symbol either, because it doesn't undergo phonological rules and is not present in the output of phonology to phonetics. Its only function is the communication between morphosyntax and phonology.

A slightly different interlanguage approach has been developed in Prosodic Phonology, in the form of prosodic constituents such as the prosodic word and the clitic group (Nespor and Vogel 1986; CHAPTER 51: THE PHONOLOGICAL WORD; CHAPTER 84: CLITICS).[10] Whereas the lower-level units (e.g. the syllable, the mora, and the foot) of Prosodic Phonology can be seen as purely phonological entities which aren't necessarily linked to morphosyntax, the prosodic word and clitic group seem to serve similar duties to the boundary symbols of *SPE*. Here is a classic example for the working of prosodic words. In Hungarian, backness vowel harmony applies to suffixes attached to lexical roots (hence across a root–suffix boundary) (19a), but not to prefixes (19b) or across compounds (19c) (CHAPTER 123: HUNGARIAN VOWEL HARMONY):

(19)	a.	*mos-at*	wash-CAUS	'make wash'
		küld-et	send-CAUS	'make send'
	b.	*oda-menni*	there-go	'to go there'
		be-utazni	in-commute	'to commute in'
	c.	*Buda-Pest*		'Budapest'
		könyv-tár	book-collection	'library'

The prosodic word-based analysis of Nespor and Vogel (1986: 121–122) assumes that prosodic words in Hungarian are built as follows:

(20) a. The domain of the prosodic word consists of a stem and any linearly adjacent string of suffixes.
b. Any unattached element forms a prosodic word on its own.

As a consequence, suffixes form a prosodic word together with the preceding root ($[mos\text{-}at]_{PW}$), two roots in a compound form different prosodic words ($[Buda]_{PW}$ $[Pest]_{PW}$), and prefixes form separate prosodic words, since they are stranded ($[oda]_{PW}$ $[menni]_{PW}$). On the assumption that vowel harmony in Hungarian cannot cross prosodic word boundaries, the distribution of vowel harmony follows. The major argument for the prosodic word is that the phonology instantiates domains which do not correspond to any morphosyntactic domain. Thus *Budapest-ben*

[10] The discussion of the Prosodic Hierarchy in this section follows closely the insights of Scheer (2008).

'in Budapest' has the morphological bracketing [[*Buda Pest*] *ben*] and the prosodic bracketing $[Buda]_{PW}$ $[Pest\text{-}ben]_{PW}$, where $[Pest\text{-}ben]_{PW}$ is a prosodic word which is not isomorphic with a morphological constituent.[11] However, the overall function of prosodic words is still to make morphological boundaries visible to phonology. Thus, in the analysis of Nespor and Vogel, (21a) and (21b) are equivalent:

(21) a. Vowel harmony cannot cross a prosodic word boundary.
b. Vowel harmony cannot cross left edges of words or root morphemes.

In general in the prosodic word literature, prosodic word boundaries correspond to morpheme boundaries, in contrast to syllables and feet (cf. the monomorphemic name of the Hungarian town *Debrecen*, which is parsed as $[De.bre]_{Ft}$*.cen*). Just like the boundary symbols of *SPE*, prosodic words are created by the morpho-syntax to communicate with phonology, and just like boundary symbols, the only motivation of prosodic words in the phonology is to trigger or block phonological processes at specific boundaries.[12] Note also that although prosodic words are formally quite distinct from boundary symbols – prosodic words are constituents in a hierarchical tree, not symbols which are linearized in the same way as segments – phonological processes refer to them in much the same way as to boundary symbols. Thus Nespor and Vogel suggest (following Selkirk 1980) that prosodic word-related phonological rules all belong to a small set of formats which refer only to the boundaries of prosodic words. In fact Neeleman and van de Koot (2006) show that the tree representations of Nespor and Vogel (1986) can be translated without any loss of information in a system where all parts of the Prosodic Hierarchy are boundary symbols.

The interlanguage approach is also used to encode information on single morphemes. Thus *SPE* uses boundary symbols to mark exceptional morphemes. For Nespor and Vogel (1986: 140), specific affixes are marked by a diacritic, which has the effect that they form independent prosodic words. On the assumption that prosodic words can be recursive, they can also directly encode more complex bracketing patterns (see Peperkamp 1997).

3.3.2 *Morphology speaking phonology*

A system where morphology communicates with phonology in the vocabulary of phonology, called "Direct Interface," is developed in Scheer (2008). Scheer assumes that the only information morphosyntax can transmit to phonology consists of phonological symbols such as the phonological timing elements C and V. More specifically, morphosyntactic boundaries can be encoded at the output of morphosyntax by inserting specific phonological material between morphemes

11 As rightly pointed out by one of the reviewers, there are no detailed arguments for the assumed morphological structure for many cases where non-isomorphism between morphology and phonology is claimed. Thus it is quite possible that the morphological structure of *Budapestben* is actually [*Buda*-[*Pest-ben*]], since the semantic scope of *-ben* 'in' is in any case larger than the morphological base to which it attaches (its scope is over the entire DP, including determiners and modifiers).

12 This is not strictly true in theories where the prosodic word is an inextricable part of the representation for word stress, but Nespor and Vogel assume that word stress is implemented by the metrical grid, not the prosodic word, and is in principle independent of the prosodic word. Note also that the domain of stress in Hungarian is bigger than the domain for vowel harmony since it comprises compounds and clitics (Vogel 1988).

or by restricted types of manipulation of phonological segments at morpheme edges. For example, it is assumed that many languages insert the sequence CV (an unassociated C timing slot and an unassociated V timing slot) to mark the left edge of morphosyntactic words. In the version of Government Phonology assumed by Scheer, an unassociated V slot is only possible if the next V slot to its right is associated to a phonetically visible vowel: C-t V-a and C-t V-Ø C-r V-a are fine, but C-t V-Ø C-k V-Ø C-r V-a is ruled out because the initial V-Ø is followed by another V-Ø, which accounts for the absence of triconsonantal word-initial clusters.

Obligatory addition of unassociated CV at the left word edge now leads to an additional empty nucleus, which cannot be licensed. Hence something like #tr is systematically impossible in a language of this type, because it would be represented as C-Ø V-Ø C-t V-Ø C-r V-a. Crucially, Direct Interface works very much like *SPE*: CV elements inserted at boundaries are boundary symbols in the sense that they correspond to the edges of morphosyntactic constituents, but are not present themselves in the syntax. The crucial difference is that CV slots (and other boundary markers assumed in Direct Interface) are part of the independently motivated vocabulary of the assumed phonological formalism. They occur both in lexical representations of single morphemes and in the output of phonology, with well-defined consequences.

In a Direct Interface approach, information on specific morpheme types must also be encoded by phonological vocabulary. Thus apparent exceptionality can be implemented by assuming different phonological representations for exceptional and regular items. For example, Inkelas and Cho (1993) argue that exceptions to final devoicing in Turkish (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION) are due to the fact that "regularly" devoicing obstruents are underlyingly underspecified for voicing, whereas exceptionally non-devoicing obstruents are underlyingly voiced (and consistently voiceless obstruents underlyingly voiceless). On the other hand it is hard to imagine how a Direct Interface approach would handle sensitivity to complex bracketing as in the cyclic computation of compound stress (cf. §2.4), and in fact Scheer (2009) assumes that effects of this type are due to external control, much as in *SPE*.

3.3.3 *Phonology speaking morphology*

In most varieties of Optimality Theory, phonological constraints have direct access to morphological structure. Hence phonological constraints can access morphological structure without an interlanguage being invoked or morphosyntax bothering about translating boundaries into phonological vocabulary.[13]

4 Restrictions on phonological sensitivity to morphology

Approaches to phonological sensitivity for morphology are inherently theories of non-sensitivity. In other words, they try to impose interesting and general restrictions on the amount of morphological structure which can be accessed by

[13] Approaches where phonology has direct access to morphosyntactic boundaries have also been advocated by Odden (1990) and Kaisse (1990).

phonological mechanisms. These restrictions are generally of two different types: restrictions on the types of accessible information (e.g. specific boundaries, but not others), and restrictions on the locality of accessibility (e.g. the identity of a specific morpheme might be accessible for a process which applies to sounds which are exponents of this morphemes, but not to other sounds). I will discuss the first type of restriction in §4.1, and the second one in §4.2. In §4.3, I will discuss the most prominent restriction on sensitivity to morphology which has been proposed in the literature, "indirect reference," which basically claims that the communication between morphology and phonology is achieved by an interlanguage or by phonological vocabulary in a modular architecture. I will argue that indirect reference is not by itself an interesting restriction, but only in as far as it predicts that sensitivity to morphology works in a consistent way for all types of phonological processes.

4.1 *Restrictions on the type of accessible morphology: Impoverishment*

Most approaches to the morphology–phonology interface propose explicitly or implicitly that phonological sensitivity to morphology is restricted as to which types of morphological properties are accessible to phonology. For example, specific types of morphemes or morpheme boundaries might be visible to phonology, whereas other types are not. Procedurally, this type of restriction can be interpreted as a kind of impoverishment process: a part of the structure built by morphology is deleted before the phonology has the chance to interpret it.[14]

A simple example is the version of Positional Faithfulness Theory in Beckman (1998) applied to morphological categories, where there are special faithfulness constraints for prominent categories, but not for non-prominent categories, e.g. there is a constraint Ident-σ_1 protecting sounds in root-initial syllables, but no corresponding constraint for affixes. Affix segments are only protected by a general version of Ident, which extends to all segments. Under this approach, phonology (at least for faithfulness constraints) cannot directly access whether a segment belongs to an affix. A related restriction which has been proposed in the OT literature is the claim that only faithfulness constraints can be indexed for morphological categories, which predicts that morphology cannot trigger markedness reversals (cf. Itô and Mester 1999 and Alderete 2001, but see Inkelas and Zoll 2007, Pater 2009, and Bermúdez-Otero, forthcoming, for critical discussion).

Under a modular architecture, a substantive restriction on the types of brackets which can be accessed is introduced by Kaye (1995). Kaye argues that only analytic morpheme boundaries (level 2 boundaries) are visible to phonology, while synthetic boundaries (level 1 boundaries) are systematically invisible. A similar claim is implicit in Prosodic Phonology. Thus Nespor and Vogel's analysis for Hungarian makes root–suffix boundaries virtually invisible to the phonology, since they do not correspond to prosodic word boundaries (nor to any other prosodic boundaries, except by accidental correspondence to foot and syllable boundaries).

Finally, impoverishment also plays an important role in external-control models of sensitivity to morphology. Thus, in Stratal OT, constraint evaluation takes place – apart from applying to words – in phonological stems. For example,

14 The label "impoverishment" in this context is due to Bermúdez-Otero (forthcoming).

Bermúdez-Otero (forthcoming) shows that in Spanish the (partially lexically controlled) process of stressed stem vowel diphthongization applies to stems, since it is bled by stem-level affixation of (stress-shifting) *-(r)al*, but not at the word level, since the affixation of the word-level affix *-on* doesn't affect diphthongization, even though it also leads to stress shift:

(22)	a.	*cuerpón*	[kwer.'pon]	'big body (N); of large and robust build (A/N)'
		cf. *cuerpo*	['kwer.po]	'body (N)'
		corporal	[kor.po.'ɾal]	'bodily, corporal (A)'
	b.	*huevón*	[we.'βon]	'big egg (N); stupid, plucky, lazy (A/N)'
		cf. *huevo*	['we.βo]	'egg (N); (colloq.) testicle (N)'
		oval	[o'.βal]	'oval (A)'

A second process found in some varieties of Spanish illustrates phonological evaluation at the word level. *n* becomes a placeless nasal ([N]) in coda position, which is maintained under resyllabification across word boundaries, but is bled by resyllabification triggered by the augmentative word-level affix *-azo*:

(23)	*huevón*	[we.'βoN]	'big egg (N); stupid, plucky, lazy (A/N)'
	huevon-azo	[we.βo.'na.θo]	'big *huevón*'
	huevón imbécil	[we.'βo.Nim.'be.θil]	'stupid *huevón*'

The crucial argument of Bermúdez-Otero is now that, in the morphological structure [[[*huev*]-*on*]-*azo*], the constituent [[*huev*]-*on*] does not trigger evaluation by the phonology, which is evident from the fact that the *n* does not undergo debuccalization. Crucially, "stem" in Stratal Optimality Theory is a phonological stem (cf. §2.2), not a genuinely morphological constituent (thus stem is not a label in the morphological representation), but a specific sub-word constituent selected by the morphology–phonology interface as visible for phonological evaluation, which effectively makes all other sub-word constituents invisible to the phonological evaluation algorithm.

4.2 Restrictions on the locality of sensitivity to morphology

Since most types of sensitivity to morphology can at least partially be reduced to sensitivity to morphemes (cf. §2), the discussion here will be restricted to the locality for phonological sensitivity to specific morpheme types. Sensitivity to morpheme type is usually quite local, which can be best illustrated by positional faithfulness effects. Recall the vowel harmony data from (2), where the special protection to underlying [ATR] values "imported" by root morphemes into the phonological representations of words is radically local in the sense that it only applies to segments which are part of the very same root. If it were otherwise (if the high-ranked faithfulness attributed to the root extended to the whole word containing a root), the relevant asymmetry between roots and affixes would disappear, since affix segments should be just as faithful to their underlying ATR values as root segments.

In the literature one finds three major proposals for relevant locality domains in phonological sensitivity to morphology.[15] First, locality of morphological sensitivity might be subject to phonological locality. This type of locality follows directly from the architecture of the grammar in approaches where all communication between phonology and morphology is carried out by boundary symbols of the interlanguage (*SPE*) or of the phonological (Direct Interface) type. For example, since boundary symbols are treated as phonological units, they should only be able to have an impact on dissimilation or assimilation processes if they intervene between trigger and target, simply because this is the only position where other phonological units could influence (dis-)harmony.

A second proposal implicit in much of the OT literature is that the locality of morphological sensitivity is restricted to exponence of morphemes. I will call this type of locality *morpheme-based locality*. Thus the schema of positional faithfulness constraints is defined in such a way that only segments which are in a prominent position enjoy special protection. For positional root faithfulness this means root consonants, hence sensitivity to morphology in this case is restricted to the phonemes which are exponents of a given morpheme. For markedness constraints which are sensitive to morpheme type we might assume a similar locality convention as in (24):

(24) $*X_L$

Assign a violation mark to any instance of X that contains only phonological exponents of morphemes specified as L.

Thus, if *NÇ (Pater 1999), a constraint which penalizes sequences of a nasal and a following voiceless obstruent (NÇ), were restricted to root segments it would be violated by a word form where both the nasal and the voiceless stops are part of a root (or of different roots), but not if at least one of these segments would be affiliated to an affix (thus it would be violated by *mp* in *imperial*, but not in *im-polite*).

Pater (2009) argues that the convention in (24) is likely to be too restrictive. For example, in the case of velar deletion discussed in §3.2, $*X_L$ corresponds to $*VKV_{Dat}$. However, this constraint is obviously relevant as soon as one of the segments (the final vowel) is part of the dative affix. Pater (2009) himself proposes that morphologically indexed phonological constraints are applied in such a way that they count constraint violations if the locus of the constraint violation just overlaps with some exponent of the morpheme:

(25) $*X_L$

Assign a violation mark to any instance of X that contains a phonological exponent of a morpheme specified as L.

[15] One prominent proposal on locality of morphological sensitivity which I do not discuss here is Bracket Erasure, a key assumption of Lexical Phonology (Kiparsky 1982; Mohanan 1986), which requires that at the end of a phonological cycle all hierarchical structure internal to the cycle is erased. This derives, for example, the important generalization that phrase-level processes do not have access to the internal morphological structure of words (Bermúdez-Otero, forthcoming). However, Inkelas and Zoll (2007) show that bracket erasure which requires locality with respect to phonological stems follows a fortiori in a system which requires locality with respect to morphological stems (see also Bermúdez-Otero, forthcoming). The latter is discussed in detail below.

A third proposal for the locality domain of morphological sensitivity is made in sign-based morphology and phonology (Inkelas and Zoll 2005), where all sensitivity to morphological structure is restricted to "constructions," which roughly translates to the notion of morphological stem defined in §2.4. I will call this type of locality stem-based locality. Although morpheme-based and stem-based locality are advocated as alternative approaches to the same problem, they are conceptually not mutually exclusive, and hence an argument for one of these locality domains does not necessarily mean that the other one is wrong.

Evidence for morpheme-based locality comes from cases of sensitivity to exceptional morphemes or to boundaries. Thus Assamese has regressive iterative [ATR] vowel harmony (26a) (Mahanta 2007), where [ɑ] is generally opaque to harmony (26b) (Mahanta 2008: 166):

(26)	a.	[b^{h}ɛkola]	'frog'	+	[i]	[b^{h}ekuli]	'frog (DIM)'
		[k^{h}ɔrɔs]	'spend'	+	[i]	[k^{h}orosi]	'prodigal'
	b.	[kɔpɑh]	'cotton'			[kɔpɑhi]	'made of cotton'
		[zɔkɑr]	'shale'			[zɔkɑri]	'shake (INF)'

However, before the affixes [-ijɑ] and [-uwɑ], [ɑ] exceptionally undergoes [ATR] harmony ([-ijɑ] derives adjectives; [-uwɑ] is a causative suffix):

(27)	[sɑl]	'roof'	[solijɑ]	'equipped with aroof'
	[mɑr]	'beat (VB)'	[moruwɑ]	'beat (CAUS)'

The crucial observation is now that exceptionality only affects an [ɑ] if it immediately precedes the [+ATR] vowel of [-ijɑ/-uwɑ], but not if it is separated from the suffix by other vowels, as shown in (28). This follows from the formulation of morphemic locality if exceptional vowel harmony is triggered by the constraint *[−ATR][+ATR] indexed to [-ijɑ/-uwɑ]:

(28)	[pɑtɔl]	'light'	[pɑtolijɑ]	'lightly'
	[ɑdhɑ]	'half'	[ɑdhoruwɑ]	'halved'

Sensitivity to boundaries is implicitly local in a way which is very similar to sensitivity to exceptional morphemes. Thus in a derived environment effect such as Finnish assibilation (cf. §2.2), a boundary triggers (or allows an otherwise not found) process, but only if the trigger and the target of the process are adjacent to the boundary, not by virtue of being part of the same word as the relevant boundary. Thus underlying /tilat-i/ becomes [tilas-i] in the output, not [si-las-i]. This follows under the assumption that boundary sensitivity is actually a subtype of morpheme sensitivity (cf. §2.2). Thus, if Finnish assibilation is interpreted as spreading of a [+continuant] feature from /i/ to a preceding /t/, its restriction to derived environments could be captured in the theory of van Oostendorp (2007) by the constraint ALTERNATION, which bans derived association of features to tautomorphemic segments (Wolf 2008: 329), as would be the case for the two initial segments of /tilat-i/ (see also Pater 2006 for a related implementation of locality in derived environment effects).

Evidence for stem-based locality comes from cases like the Hausa data discussed in §2.4, where exceptional deletion of tone affects the complement of dominant affixes (i.e. the base to which the affix attaches), but not of affixes attached outside of dominant affixes. Hausa also provides a problem for morpheme-based locality, because it affects not only tones phonologically adjacent to the dominant affixes, but all tones in the complement of the affix. This is structurally similar to cases of long-distance derived-environment effects. For example, Catalan does generally not allow the dental fricative [θ]. However, in Spanish loanwords [θ] is possible, indicating that loanwords are exceptional phonologically and introduce sensitivity to morpheme type (CHAPTER 95: LOANWORD PHONOLOGY). If a loanword containing [θ] undergoes derivation by a Catalan affix, the whole word becomes subject to the ban on [θ], and instances of [θ] are turned into [s], which is a licit sound of Catalan. This effect of affixation is phonologically non-local, since it also applies to cases of [θ] which are not adjacent to the triggering affix (Mascaró 2003):

(29)	a.	*Zamora*	(place name)	/θəmorə/	→	[θə'morə]
	b.	*Zamorá*	'pertaining to Zamora'	/θəmorə-'a/	→	[səmu'ra]

Cases like the one in (29) seem to be fatal for morpheme-based locality, since the initial /θ/ in /θəmorə-'a/ does not even overlap with the affix *-á*. For example the application of a constraint *θ indexed for *-á* to stem-initial /θ/ in (29b) satisfies neither the locality convention in (24) nor the relaxed version in (25). There is a simple solution, which has been proposed in different places in the literature (Kiparsky 1993; Bonet 2004; Wolf 2008). If the phonology is not sensitive to the exceptionality of roots, but of stems (or words), exceptionality of a morpheme must be inherited by the stem/root which contains it by feature percolation (CHAPTER 106: EXCEPTIONALITY). In a derived form, where the derivational affix is the head, not the exceptional root, percolation of exceptionality fails, the stem or word is not marked as exceptional, and the entire form is subject to the native phonology. A problem for this escape hatch for morpheme-based locality is that it automatically provides a loophole which allows violations of stem-based locality: if exceptionality features can be percolated upwards, they can spread far beyond the stem which contains a specific affix. A potential problem for stem-based locality is described in detail in Pater (2009) for two homophonous affixes *-wa* in Yine. Suffixes in Yine generally trigger syncope in vowel-final bases (e.g. /heta+ja/ → [hetja] 'see there'). Whereas $-wa_1$ exceptionally fails both to trigger syncope for material it selects and to undergo it if selected by another affix (e.g. /heta+wa+lu/ → [hetawalu] 'going to see him yet'), $-wa_2$ fails to trigger syncope, but undergoes it (/meji+wa+lu/ → [mejiwlu] 'celebration'). As Pater points out, neither the local tree headed by $-wa_{1/2}$ nor the local tree headed by an affix outside them can provide the adequate locality domain for capturing the difference between $-wa_1$ and $-wa_2$. But, as admitted by Pater, the Yine data could also be captured by an analysis which grounds the different behavior of affixes in different underlying phonological representations, and which is thus in accordance with the Direct Interface approach and phonological locality. Note, however, that the data showing non-local effects of the Hausa and Catalan type are also crucially problematic for phonological locality, i.e. the assumption that the only access phonology has to morphology is by the way of phonological material.

4.3 *Indirect reference and consistency of sensitivity*

Indirect reference is a grammatical meta-principle which requires that phonological rules or constraints are not allowed direct access to morphological or syntactic information (Inkelas 1990: 47; Bermúdez-Otero, forthcoming). However, indirect reference does not by itself imply substantial empirical restrictions on the amount of sensitivity to morphology. Thus a great number of morphological categories might be translated into a rich system of prosodic categories (see e.g. Inkelas 1990 and Downing 2006 for well-articulated inventories of morphologically motivated prosodic categories), satisfying indirect reference, but allowing detailed access to morphological information. On the other hand, we might imagine a system where phonology has direct access to morphological structures, but only to a very restricted subset, since access is restricted by general interpretability conditions or a kind of impoverishment filter which removes specific aspects of morphological structure from the information communicated to phonology. In principle the relevant restrictions are the substantial restrictions on phonology–morphology interaction, not the language which phonology and morphology use to achieve communication.

However, indirect reference – at least in a modular architecture – predicts an interesting generalization, which I will call "Consistency of Sensitivity": different phonological processes are expected to have in principle the same access to morphological structure as other processes. Take again as an illustration Hungarian vowel harmony to which root-suffix boundaries are invisible but prefix–root boundaries are visible. Under Consistency of Sensitivity we might expect that other phonological processes in the language work in the same way, i.e. they respect prefix–root boundaries in some way, but disrespect root-suffix boundaries. Nespor and Vogel argue that this holds true for Hungarian, as illustrated by processes of palatalization (Nespor and Vogel 1986: 123–124) and ellipsis (Trommer 2008a). This follows under a restrictive version of the interlanguage approach. On the assumption that no other prosodic constituent of Hungarian has edges which systematically correspond to root–suffix boundaries, it is predicted that these boundaries are generally invisible throughout the language.[16] Consistency of Sensitivity follows directly not only in interlanguage approaches, but also in theories where morphosyntax translates boundaries into phonological structure – since phonology does not have direct access to morphosyntax, but only to the phonological elements generated by morphosyntax, all processes can access morphological structure to the same extent, and more generally in a modular architecture where phonology and morphology communicate through a unique interface.[17]

Hence, if Consistency of Sensitivity is an empirically valid prediction cross-linguistically, it provides an important argument against the direct access of

[16] Of course there might be a language where prosodic words are constructed in a different way, such that the claim is that Consistency of Sensitivity holds for single languages. However, depending on the details of the theory, it might also make predictions about morphological boundaries which are universally invisible.

[17] A modular architecture where phonology has access to a restricted subset of morphological vocabulary might work in such a way that specific types of morphological structure are systematically deleted before morphological information is transmitted to phonology. For example, specific features of morphemes or boundaries might be deleted.

phonology to morphological structure. Unfortunately, there are few empirical studies which attempt to test this prediction. Nespor and Vogel (1986) provide a number of nice examples for languages which show Consistency of Sensitivity, but many of their examples are empirically flawed, since crucial data are not taken into account.[18] Moreover, citing cases which confirm a given hypothesis is not instructive as long as it is not shown in a systematic way (e.g. by the evaluation of a typologically balanced sample of languages) that there are no counterexamples – or at least significantly fewer languages with counterexamples. In fact, Bickel *et al.* (2010) discuss a number of languages which have different word-like prosodic domains which differ from each other for different phonological processes, providing explicit counterevidence to Consistency of Sensitivity.[19] Thus it is fair to say that it is unclear at the moment whether Consistency of Sensitivity is a valid empirical generalization.

5 Summary

We have identified three basic types of sensitivity of phonology to morphological structure (sensitivity to the presence and type of morphemes, sensitivity to morpheme boundaries, and sensitivity to stems), and three basic approaches to model it (external control, paradigmatic relations, and intermodular communication), but the central question which formal and substantive restrictions constrain possible cases of such sensitivity of remains a major topic for future research.

REFERENCES

Alderete, John. 1999. Head dependence in stress–epenthesis interaction. In Hermans & van Oostendorp (1999), 29–50.

Alderete, John. 2001. Dominance effects as transderivational anti-faithfulness. *Phonology* 18. 201–253.

Anttila, Arto. 2002. Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20. 1–42.

Bachrach, Asaf & Michael Wagner. 2007. Syntactically driven cyclicity vs. output–output correspondence: The case of adjunction in diminutive morphology. Unpublished ms., MIT & Cornell University.

Beckman, Jill N. 1998. Positional faithfulness. Ph.D. dissertation, University of Massachusetts, Amherst.

Beckman, Jill N., Laura Walsh Dickey & Suzanne Urbanczyk (eds.) 1995. *Papers in Optimality Theory*. Amherst: GLSA.

Bermúdez-Otero, Ricardo. Forthcoming. *Stratal Optimality Theory*. Oxford: Oxford University Press.

Bickel, Balthasar, René Schiering & Kristine Hildebrandt. 2010. The prosodic word is not universal. *Journal of Linguistics* 46. 657–709.

[18] To cite just two examples: Turkish vowel harmony has been re-analyzed by Vogel herself in a way which is completely independent of prosodic constituency due to the fact that the patterns are highly exceptional (Kabak and Vogel 2001). Northern Italian *s*-voicing is shown to be in principle independent of the prosodic word by Krämer (2005).

[19] Seidl (2001) makes similar observations for phrasal phonology.

Bonet, Eulàlia. 2004. Morph insertion and allomorphy in Optimality Theory. *Journal of English Studies* 4. 73–104.

Booij, Geert. 1995. *The phonology of Dutch.* Oxford: Clarendon Press.

Burzio, Luigi. 2000. Cycles, non-derived-environment-blocking, and correspondence. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: Phonology, syntax, and acquisition*, 47–87. Oxford: Oxford University Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.) *Ken Hale: A life in language*, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Downing, Laura J. 2006. *Canonical forms in prosodic morphology.* Oxford: Oxford University Press.

Downing, Laura J., T. A. Hall & Renate Raffelsiefen (eds.) (2005). *Paradigms in phonological theory.* Oxford: Oxford University Press.

Golston, Chris. 1995. Syntax outranks phonology: Evidence from Ancient Greek. *Phonology* 12. 343–368.

Golston, Chris & Richard Wiese. 1998. The structure of the German root. In Wolfgang Kehrein & Richard Wiese (eds.) *Phonology and morphology of the Germanic languages*, 165–185. Tübingen: Niemeyer.

Hall, T. A. & Ursula Kleinhenz (eds.) 1999. *Studies on the phonological word.* Amsterdam & Philadelphia: John Benjamins.

Hall, T. A. & John H. G. Scott. 2007. Inflectional paradigms have a base: Evidence from s-dissimilation in Southern German dialects. *Morphology* 17. 151–178.

Hermans, Ben & Marc van Oostendorp (eds.) 1999. *The derivational residue in phonological Optimality Theory.* Amsterdam & Philadelphia: John Benjamins.

Hyman, Larry M. 2008. Directional asymmetries in the morphology and phonology of words, with special reference to Bantu. *Linguistics* 46. 309–350.

Inkelas, Sharon. 1990. *Prosodic constituency in the lexicon.* New York: Garland.

Inkelas, Sharon. 1998. The theoretical status of morphologically conditioned phonology: A case study of dominance effects. *Yearbook of Morphology 1997.* 121–155.

Inkelas, Sharon & Young-mee Yu Cho. 1993. Inalterability as prespecification. *Language* 69. 529–574.

Inkelas, Sharon & Draga Zec (eds.) (1990). *The phonology–syntax connection.* Chicago: University of Chicago Press.

Inkelas, Sharon & Cheryl Zoll. 2005. *Reduplication: Doubling in morphology.* Cambridge: Cambridge University Press.

Inkelas, Sharon & Cheryl Zoll. 2007. Is grammar dependence real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics* 45. 133–171.

Itô, Junko & Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.) *The handbook of Japanese linguistics*, 62–100. Malden, MA & Oxford: Blackwell.

Kabak, Barış & Irene Vogel. 2001. The phonological word and stress assignment in Turkish. *Phonology* 18. 315–360.

Kager, René. 1999. Surface opacity of metrical structure in Optimality Theory. In Hermans & van Oostendorp (1999), 207–245.

Kaisse, Ellen M. 1990. Toward a typology of postlexical rules. In Inkelas & Zec (1990), 127–143.

Kaye, Jonathan. 1995. Derivations and interfaces. In Jacques Durand & Francis Katamba (eds.) *Frontiers of phonology: Atoms, structures, derivations*, 289–332. London & New York: Longman.

Kenstowicz, Michael. 2005. Paradigmatic uniformity and contrast. In Downing *et al.* (2005), 145–169.

Kiparsky, Paul. 1973. Phonological representations: Abstractness, opacity, and global rules. In Osamu Fujimura (ed.) *Three dimensions of linguistic theory*, 57–86. Tokyo: Taikusha.

Kiparsky, Paul. 1982. From Cyclic Phonology to Lexical Phonology. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part 1, 131–175. Dordrecht: Foris.

Kiparsky, Paul. 1993. Blocking in nonderived environments. In Sharon Hargus & Ellen M. Kaisse (eds.) *Studies in Lexical Phonology*, 277–313. San Diego: Academic Press.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–365.

Krämer, Martin. 2005. Contiguity and non-derived environment blocking of s-voicing in Lombardian and Tuscan Italian. *Probus* 17. 227–251.

Mahanta, Shakuntala. 2007. Directionality and locality in vowel harmony. Ph.D. dissertation, University of Utrecht.

Mahanta, Shakuntala. 2008. Local vs. non-local consonantal intervention in vowel harmony. In Sylvia Blaho, Camelia Constantinescu & Erik Schoorlemmer (eds.) *Proceedings of ConSOLE XV*, 165–188. Leiden: University of Leiden.

Mascaró, Joan. 2003. Comparative markedness and derived environments. *Theoretical Linguistics* 29. 113–122.

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.

McCarthy, John J. 2008. The gradual path to cluster simplification. *Phonology* 25. 271–319.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Beckman *et al.* (1995), 249–384.

Michelson, Karin. 1989. Invisibility: Vowels without a timing slot in Mohawk. In Donna Gerdts & Karin Michelson (eds.) *Theoretical perspectives on Native American languages*, 38–69. New York: SUNY Press.

Mithun, Marianne & Hasan Basri. 1986. The phonology of Selayarese. *Oceanic Linguistics* 25. 210–254.

Mohanan, K. P. 1986. *The theory of Lexical Phonology*. Dordrecht: Reidel.

Neef, Martin. 1996. *Wortdesign: Eine deklarative Analyse der deutschen Verbflexion*. Tübingen: Stauffenburg Verlag.

Neeleman, Ad & J. van de Koot. 2006. On syntactic and phonological representations. *Lingua* 116. 1524–1552.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Newell, Heather. 2005. Bracketing paradoxes and particle verbs: A late adjunction analysis. In Sylvia Blaho, Luis Vicente & Erik Schoorlemmer (ed.) *Proceedings of ConSOLE XIII*, 249–272. Leiden: University of Leiden.

Newman, Paul. 1986. Tone and affixation in Hausa. *Studies in African Linguistics* 17. 249–267.

Newman, Paul. 2000. *The Hausa language: An encyclopedic reference grammar*. New Haven & London: Yale University Press.

Odden, David. 1990. Syntax, lexical rules and postlexical rules in Kimatuumbi. In Inkelas & Zec (1990), 259–277.

Oostendorp, Marc van. 2007. Derived environment effects and Consistency of Exponence. In Sylvia Blaho, Patrik Bye & Martin Krämer (eds.) *Freedom of analysis?*, 123–148. Berlin & New York: Mouton de Gruyter.

Orgun, Cemil Orhan. 1996. Sign-based morphology and phonology with special attention to Optimality Theory. Ph.D. dissertation, University of California, Berkeley.

Pater, Joe. 1999. Austronesian nasal substitution and other NC̦ effects. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*, 310–343. Cambridge University Press.

Pater, Joe. 2006. The locus of exceptionality: Morpheme-specific phonology as constraint indexation. In Leah Bateman, Michael O'Keefe, Ehren Reilly & Adam Werle (eds.) *Papers in Optimality Theory III*, 259–296. Amherst: GLSA.

Pater, Joe. 2009. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In Steve Parker (ed.) *Phonological argumentation: Essays on evidence and motivation*, 123–154. London: Equinox.

Peperkamp, Sharon. 1997. Prosodic words. Ph.D. dissertation, University of Amsterdam.
Piggott, Glyne L. 1998. Foot form and the parsing of weightless syllables. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 34. 315–332.
Piggott, Glyne L. 2001. The phonotactics of a "Prince" language: A case study. Unpublished ms., McGill University.
Pulleyblank, Douglas. 1996. Neutral vowels in Optimality Theory: A comparison of Yoruba and Wolof. *Canadian Journal of Linguistics* 41. 295–347.
Raffelsiefen, Renate. 2005. Paradigm uniformity effects versus boundary effects. In Downing *et al.* (2005), 211–262.
Revithiadou, Anthi. 1999. Headmost accent wins: Head dominance and ideal prosodic form in lexical accent systems. Ph.D. dissertation, University of Leiden.
Scheer, Tobias. 2008. Why the prosodic hierarchy is a diacritic and why the interface must be direct. In Jutta M. Hartman, Veronika Hegedűs & Henk van Riemsdijk (eds.) *Sounds of silence: Empty elements in syntax and phonology*, 145–192. Amsterdam: Elsevier.
Scheer, Tobias. 2009. Intermodular argumentation and the word-spell-out-mystery. In Kleanthes Grohmann (ed.) *Explorations of phase theory: Interpretation at the interfaces*, 23–65. Berlin & New York: Mouton de Gruyter.
Seidl, Amanda. 2001. *Minimal indirect reference: A theory of the syntax–phonology interface*. New York: Routledge.
Selkirk, Elisabeth. 1980. Prosodic domains in phonology: Sanskrit revisited. In Mark Aronoff & Mary-Louise Kean (eds.) *Juncture*, 107–129. Saratoga: Anma Libri.
Selkirk, Elisabeth. 1995. The prosodic structure of function words. In Beckman *et al.* (1995), 439–469.
Trommer, Jochen. 2008a. "Case suffixes," postpositions, and the phonological word in Hungarian. *Linguistics* 46. 403–437.
Trommer, Jochen. 2008b. Stress uniformity in Albanian: Morphological arguments for cyclicity. Unpublished ms., University of Leipzig.
Vigário, Marina. 2003. *The prosodic word in European Portuguese*. Berlin & New York: Mouton de Gruyter.
Vogel, Irene. 1988. Prosodic constituents in Hungarian. In Pier Marco Bertinetto & Michele Loporcaro (eds.) *Certamen Phonologicum*, 231–250. Turin: Rosenberg & Sellier.
Wiese, Richard. 1996. *The phonology of German*. Oxford: Clarendon Press.
Wolf, Matthew. 2008. Optimal interleaving: Serial phonology–morphology interaction in a constraint-based model. Ph.D. dissertation, University of Massachusetts, Amherst.

104 Root–Affix Asymmetries

SUZANNE URBANCZYK

1 Introduction

When taking stock of the phonological patterns found in the languages of the world, some patterns can be allocated to roots, some to affixes, and even more to when roots combine with affixes. As Jakobson (1965: 29) observed, "... roots as lexical and affixes as inflectional morphemes ..." show asymmetric patterning, such that affixes "... illustrate a selected use of phonemes and their combinations." For example, in Semitic languages pharyngeal consonants (CHAPTER 25: PHARYNGEALS) are found in roots, but not in affixes (McCarthy and Prince 1995), and in many languages ejectives are only found in roots (Rose and Walker 2004; Bybee 2005). Examples from Classical Arabic (Semitic; McCarthy 2005) and Lushootseed (Salish; Bates *et al.* 1994) illustrate these patterns in (1) below.

(1) *Root–affix segmental qualities*

		roots		*affixes*	
a.	*Arabic*	faʕal	'do (PERF)'	-tu	(1SG COM, PERF)
		ħmarar	'redden'	na-	(3PL COM, IMP INDIC)
		dħaraʒa	'to roll (CONJ I)'	-aj	(DUAL, GEN/ACC)
b.	*Lushootseed*	q'axʷ-	'to freeze'	-alap	'you (PL)'
		k'ʷas-	'to roast'	-icid	'you (SG)'
		gəq'-	'to open something'	-txʷ	'CAUSATIVE'

In terms of combinations of sounds, in Sanskrit roots can have complex onsets (CHAPTER 55: ONSETS), while affixes cannot (Steriade 1988; no glosses given for affixes).

(2) *Sanskrit prosodic structure*

roots		*affixes*
krand	'cry out'	-ur
dwis	'hate'	-ta
djaut	'shine'	-ita

In terms of assimilatory processes, many languages exhibit what is referred to as "root-controlled" vowel harmony. In Turkish, a suffix vowel agrees in backness and rounding with a preceding root vowel, as can be seen in the plural words in (3a) below (CHAPTER 118: TURKISH VOWEL HARMONY). And in English, when the productive form of the plural /-z/ is suffixed to roots, the affix consonant, not the root consonant, becomes voiceless (though there are some exceptions, like *leaves, wives*, and *clothes*).

(3) *Root–affix alternation asymmetries*

		singular	*plural*		
a.	*Turkish*	dev	devler	'giant'	
		cɑn	cɑnlɑr	'soul'	
		ip	ipler	'rope'	
		son	sonlɑr	'end'	
b.	*English*	kɪt	kɪts	'kit'	*[kɪdz]
		tɑp	tɑps	'top'	*[tɑbz]
		dɑg	dɑgz	'dog'	*[dɑks]

This cursory view of root–affix asymmetries belies a more complex situation in which general trends interact with other factors. The aim of this chapter is twofold. The first is to tease apart some of these factors by presenting a survey of a range of root–affix asymmetries to determine the scope and kinds of asymmetries found in the phonological patterns of languages. The second is to present various theoretical approaches, with a brief discussion of how the predictions match the patterns found. In doing so, criteria for establishing what are authentic *root–affix asymmetries* (RAAs) will be provided, while also identifying what could be alternative explanations for the patterns when relevant. For instance, with the Turkish example, one could reasonably ask whether the pattern of vowel harmony is specifically due to the root maintaining the contrast at the expense of the suffix, or whether it is due to directionality, i.e. progressive assimilation. A striking finding in reviewing the literature on RAAs is that authentic cases of RAA reversals (affix dominance) seem to share a common property, that of being dominant–recessive systems. The two areas reported to have affix dominance are morphological accent and vowel harmony. Finally, the preferences found are part of a larger trend where certain prominent positions are known to exhibit more contrasts and resist changes (Trubetzkoy 1939; Steriade 1995; Beckman 1999).

A number of secondary themes have arisen from compiling this research. The first relates to methodological considerations when investigating root–affix asymmetries, discussed in §2. A second theme that has emerged relates to the mapping of form and function, with homophony avoidance or effective contrast (Ussishkin 2006) playing a potential role. This relates to contrasts in the following way. It is common to find far more roots in a language than affixes, due to the larger number of meanings expressed in roots – typically an open class of morphemes. If a language is to have roots that are not homophonous, it is necessary to draw upon a wider set of contrasts, delving into more marked segmentism. This brings the study of root–affix asymmetries full circle with Jakobson's (1965) work, which also examines the relation of form and meaning. Finally, this study touches on the topic of understanding what a root is, as well as morphemes more generally. In cases where the morphology is not strictly concatenative, it is not a straightforward matter to identify root and affix.

After a discussion of methodological considerations, the paper outlines some of the basic patterns found, which can be grouped into two broad categories: the expression of contrasts, and the preservation of contrasts. The expression of contrasts will be the focus in §3, along with some discussion of whether or not markedness, complexity, or grammaticization plays a role in the patterns. The trend to preserve contrasts in roots at the expense of alternations in an affix will be addressed next in §4, under the label root–affix alternation asymmetries. After surveying the types of patterns found, a review of several theoretical approaches is presented (§5). Finally, because one characteristic of roots *vs.* affixes relates to the types of meanings conveyed by each, a section on form–function mappings will delve into more morphological domains of investigation (§6), including non-concatenative morphology, the role that the number of affixes a language has plays in its phonological system, morphologically conditioned alternations, and a brief overview of research on how morphological processing favors roots.

2 Methodological considerations

In terms of methodological considerations, issues that arise in identifying asymmetries, and identifying roots and affixes, will be discussed. This is followed by a brief discussion of the scope and limitations of the current study.

2.1 *Identifying asymmetries*

The first point to consider in undertaking a study of RAAs is determining what an asymmetry is. Two situations can be identified: a complement set relation and a subset relation. In the former, an asymmetrical pattern would have the following distribution, where two domains have entirely different patterns. The two basic domains are root and affix, and the two parameters of investigation are contrasts (represented below by letters {a, b, c, d}), and alternations (represented below with a function sign and subscripts to indicate different alternations $\{f_1, f_2, f_3, f_4\}$).

(4) *Root–affix asymmetry: Complement sets*

	Contrasts	Alternations
Root	{a, b}	$\{f_1, f_2\}$
Affix	{c, d}	$\{f_3, f_4\}$

As illustrated with shading in one set *vs.* another, the asymmetry shows a complementarity of patterning. What is found in roots is not found in affixes and vice versa; there is no overlap at all in terms of the contrasts or the types of alternations found. No examples of this type of RAA have been found. But this is not surprising, because the general trend in language patterning is one of subset relations. Nonetheless it is important to point this out as a potential, logically plausible asymmetry, as it is relevant in the discussion of the types of predictions made by different theoretical models.

The more usual type of asymmetry in language, such as found in research on language universals (Greenberg 1966), is that of a subset relation. For any two domains, what is found in one domain is a subset of what is found in the other domain. With two domains and two parameters, and two possible values for each parameter, the actual number of possibilities is eight (2^3), as illustrated below.

(5)	*root*	*affix*	
a.	*segmental inventory*		
	{a}	{a}	same segmental inventory
	{a}	{a, b}	root is a subset of affix inventory
	{a, b}	{a}	affix is a subset of root inventory
	{a, b}	{a, b}	same segmental inventory
b.	*alternations*		
	$\{f_1\}$	$\{f_1\}$	same alternations
	$\{f_1\}$	$\{f_1, f_2\}$	root alternations are a subset of affix alternations
	$\{f_1, f_2\}$	$\{f_1\}$	affix alternations are a subset of root alternations
	$\{f_1, f_2\}$	$\{f_1, f_2\}$	same alternations

If we eliminate the systems that do not show root–affix asymmetries, that leaves us with four possible systems of root–affix subset type asymmetries: affixes have a subset of the contrasts found with roots, roots have a subset of the contrasts found with affixes, affix alternations are a subset of root alternations, and root alternations are a subset of affix alternations. If the only principle of RAAs is that of a subset relation, then all four systems are expected to occur in the languages of the world.

In terms of the types of asymmetries investigated, as noted above, the basic patterns include contrasts and alternations. Contrasts are somewhat static, while alternations are not.[1] Alternations are very common in phonological systems, and issues arise as to what sorts of alternations to investigate. For example, in some languages there can be a set of morphemes that induce alternations only to the base, such as with English apophony found in the pair *sing vs. sang*. Do we include morphophonologically triggered alternations in understanding phonological patterning, or not? I have left discussion of these out of the section on alternations, because of the morphological conditioning.

2.2 *Identifying roots and affixes*

Several issues arise in identifying roots and affixes. When examining morphologically related words, one can see that there are core elements of form and function that are shared between them. For example, in the Halkomelem examples below, the words in (6a) and (6b) have the meaning of 'to sew', an example of what Jakobson referred to as a lexical morpheme. The second member is inflected for progressive aspect, referring to how the activity is unfolding. The words in (6a) are intransitive, while those in (6b) and (6c) are transitive.

[1] As will be discussed below, in vowel harmony systems there is some interaction between these two dimensions.

(6) *Halkomelem* (Central Salish; Suttles 2004)

	perfective		*progressive*	
a.	'pet'$^{\theta}$	'sew'	'pepət'$^{\theta}$	'sewing'
b.	'pet'$^{\theta}$ət	'sew it'	'pepət'$^{\theta}$ət	'sewing it'
c.	'qamət	'bend it'	'qaqəm'ət	'bending it'

In terms of identifying the parts of the word, one would say that the form associated with the core lexical meaning 'sew' is the root in (6a) and (6b). In the first column the form associated with the meaning 'to sew' is ['pet'$^{\theta}$], while in the second column it is [pət'$^{\theta}$], with the root changing to schwa because of a regular process of unstressed vowel reduction. In terms of the other meanings expressed above, there is a transitiving suffix [-ət]. Looking at the words in (6) above, standard linguistic analysis would identify the roots as /pet'$^{\theta}$/ 'to sew' and /qam/ 'to bend', and the transitivizing suffix as /-ət/. Issues arise, however, in identifying the progressive "affix" in the second column. One exponent of *progressive* can be described as a C_1V_1- reduplicant (see CHAPTER 100: REDUPLICATION), which is formed with segments from the root. Another exponent is glottalization of the sonorant consonant in the root in (6c). What then is the affix here? As this simple example demonstrates, the issue of identifying roots and affixes in languages depends on the type of morphological operations that are used to create words. It is much more straightforward with concatenative morphology than with non-concatenative morphology (see CHAPTER 105: TIER SEGREGATION and CHAPTER 108: SEMITIC TEMPLATES on various issues related to tier segregation and non-concatenative morphology). As a result, this study will treat these types of morphological systems differently, presenting the bulk of generalizations regarding root–affix asymmetries in concatenative systems first, followed by a brief examination of non-concatenative systems.

A second set of issues arises when one considers content words *vs.* function words (lexical *vs.* inflectional, in Jakobson's terms). The issue relates to whether function words pattern with roots or affixes. For example, Willerman's (1994) cross-linguistic study of pronouns found that they have less complex segments (CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION) than content words, suggesting that function words pattern with affixes in this regard. Studies by Casali (1997) on vowel hiatus (CHAPTER 61: HIATUS RESOLUTION) and Alderete (2003) on Navajo have also found that function words (and functional categories) pattern with affixes in terms of being the locus of an alternation, while content words are similar to roots in terms of retaining contrasts and resisting alternations. Both types of RAA are therefore attested with function words. In terms of specific types of roots, different lexical classes of content words have also shown asymmetrical patterns along these same lines (see CHAPTER 102: CATEGORY-SPECIFIC EFFECTS).

Finally, there are many languages in which affixes can have a lexical type of meaning as well as a grammatical type of meaning. For example, in Salish languages there is a class of lexical suffixes that illustrates semantic, phonological, and syntactic properties of roots, and researchers on this topic have proposed that they are bound roots (Blake 1998; Urbanczyk 2001; Wiltschko 2005). Thus we can see that being bound is not a sufficient condition for being an affix. Many languages have obligatory bound roots, and many have free roots, and so it seems likely that a language can have both. Similarly, Alderete (2003) has found that the disjunct domain of prefixes patterns with the stems in Navajo in terms of having more lexical meanings, more marked phonological structure, and resisting alternations. A cursory examination of a language may fail to note these subtle patterns, so care must be taken in understanding how to categorize affixes, as they could be bound roots.

To summarize, the term root will be used as a convenient if not entirely accurate label for those parts of words that have lexical meaning, whether that be obligatorily bound or potentially free. Likewise, affixes will be understood as that portion of the word that has a more grammatical or morphosyntactic meaning associated with it, whether it be bound or not.

2.3 Summary and limitations

The limitations of this study extend to the body of data that was compiled in order to determine the range and types of root–affix asymmetries attested in the languages of the world and to issues discussed above in terms of analyzing words into roots and affixes.

Two types of research have been surveyed: research that is focused on root–affix asymmetries, such as Noske (2000), Alderete (2003), and Bybee (2005), and research that is broadly typologically based, such as Casali (1997), Hansson (2001), and Rose and Walker (2004). In surveying the research, I first attempted to find as comprehensive information on contrasts as possible, and Bybee (2005) was taken as a representative study. To find comprehensive information on alternations, I surveyed typological studies that focused on particular phonological processes, like morphological accent (Alderete 2001c) and lenition (Kirchner 1998). It is hoped that the examination of a range of different types has yielded an adequate sampling. A balance of studies was also hoped for in terms of detailed examination of a single language (Noske 2000; Alderete 2003) and broader surveys (Bybee 2005). It should be pointed out that a survey of the research shows that the broader typological studies have limitations that are not found with the detailed studies.

A further limitation on the examination of affixes is that no systematic comparison of affix type was attempted. Thus there was no specific examination of how prefixes, suffixes, and infixes could differ from each other. It should be pointed out that there may be significant differences between prefixes and suffixes, because the beginnings of words often show different phonological patterning than the ends of words and there are some well-known prefix–suffix asymmetries.

As was discussed above, the types of root–affix asymmetries that are possible fall into several categories of complement sets, and four different subset possibilities. If root–affix asymmetries are entirely unconstrained, then we expect to find an approximately equal number of each type of asymmetry. However, as will be discussed below, two types of subset root–affix asymmetry have been found much more frequently than any others: affix segment inventory is a subset of those found in roots (as pointed out by Jakobson), and affixes illustrate more alternations than roots. The reverse patterns – more contrasts in affixes and more alternations in roots – are not as robust. The following sections outline these findings for root–affix contrasts (§3) and for alternation asymmetries (§4).

3 Root–affix contrasts

As a means of investigating contrast asymmetries, this section deals with segmental contrasts and morpheme shapes.

3.1 *Segmental contrasts*

As Jakobson's observation reveals, contrasts found in affixes can be a subset of those found in roots. Bybee (2005) investigates this observation, by asking the following questions: how robust are the findings, and what are some hypotheses that account for the findings? In doing so, Bybee provides a comprehensive examination and discussion of segmental asymmetries, as found in a variety of typologically unrelated languages. As pointed out above, all segments found in affixes were a subset of those found in the language as a whole, although fewer than one quarter of the languages in her study illustrated restrictions that could be considered natural classes. The discussion of these restrictions is organized according to features classes, starting with place of articulation. The section ends by outlining some of the hypotheses put forward by Bybee in accounting for the patterns.

There are a number of well-known place markedness hierarchies, whereby sounds like coronals and glottals are considered less marked (CHAPTER 12: CORONALS; CHAPTER 4: MARKEDNESS; CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION), while complex sounds like pharyngeals and uvulars are considered marked (see de Lacy 2006 on markedness hierarchies; CHAPTER 25: PHARYNGEALS on pharyngeals). In terms of coronal unmarkedness, while Bybee (2005: 183) reports that "there is no particular evidence for the favoring of coronal consonants in affixes," Alderete (2003) observes that only coronal consonants are found in conjunct prefixes in Navajo. A difference between Bybee's and Alderete's statistical methods should be touched on at this point. Bybee was interested in determining the likelihood of specific segments occurring in affixes *vs.* roots, so just one instance of a segment in one affix is sufficient to include it as occurring in affixes. However, Alderete's study is based on frequency of occurrence. He found that some affix-class morphemes include non-coronals, but the presence of them is not statistically significant compared to other affixes in Navajo. Bybee (2005: 183) also observes that systematic exclusions of "place of articulation are rare if both obstruents and nasal consonants are taken into account," noting that labials are the ones most commonly excluded. She reports that, while they occur in roots in Abipon, Kanuri, and Tohono O'odam, alveo-palatal affricates are excluded from affixes; and in Shuswap, uvulars are excluded from affixes, even though they are found in roots. And in English, while roots have the entire range of contrasts, inflectional affixes have only coronal consonants, with the notable exception of *-ing*.

In terms of laryngeal features, the strongest tendency in Bybee's study relates to the absence of glottalization (both ejectives and implosives) in affixes. This finding is also supported in research elsewhere. In Cuzco Quechua, roots can have ejectives and plain and aspirated plosives, while affixes only have plain plosives (Parker and Weber 1996). In Amharic and Salish languages, ejectives appear only in roots, and never in affixes (Urbanczyk 2001; Usshkin 2006). The restriction is quite striking also when combined with Rose and Walker's (2004) finding that instances of consonant glottal harmony (CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS) are only found within roots, not between roots and affixes. In terms of other laryngeal contrasts, out of the core set of languages Bybee examined, only Dakota and Tojolabal have affixes that ban voiced obstruents in affixes, though the former permits voiceless aspirated stops in affixes. It is interesting to note that the majority of languages in Bybee's corpus do not show any symmetrical patterning with respect to voicing or aspiration. For example, Kanuri has [b t k g] in affixes, but not [p d]. Bybee examined seven other languages to confirm that the restriction of glottalization on affixes is robust. Interestingly, of this set of supplementary languages, Kanakura and Tigre also banned voiced obstruents from affixes and Krongo banned implosives.

In addition to the points above regarding alveo-palatal affricate and nasal/place restrictions, not much is known about manner contrast RAAs. An interesting pattern is found in Chaha, in which laterals appear only in loanwords and affixes, never in roots (see work by Banksira, cited in Golston 1996). In reviewing Bybee's appendices, one can see that there are many languages which restrict laterals in affixes (Maidu, Tojolabal, Kanuri, Baluchi, and others).

In terms of asymmetric vowel contrasts, languages were found to exclude long vowels (Kui, Kanakuru, Krongo), nasalized vowels (Ngambay, Slave), front rounded vowels (Guaymí), and non-peripheral vowels (Buriat, Krongo) from affixes, even though they are found in roots.

In terms of understanding the patterns of root–affix segmental asymmetries, Bybee outlines several approaches one can take: markedness, complexity, and grammaticization. The first two share the property that the restrictions found with affixes are related to being less marked or less complex (see CHAPTER 4: MARKEDNESS). Bybee (2005) points out that, without a suitable theory of markedness, it is difficult to assess whether affixes are truly less marked than roots. Having found only six languages which show classes of restrictions from affixes, Bybee proposes that neither markedness nor complexity offers a satisfactory account of the range of patterns found. Instead, she proposes that grammaticization is the cause of the asymmetry. Under this view, affixes are reduced forms of lexical morphemes, which start off with the full inventory of sounds. As morphemes are grammaticized, they tend to become reduced phonologically, thus eliminating some of the more complex articulations. The randomness of some affix segments is thus a reflection of the randomness of the earlier form of the morpheme or root, and is not the result of synchronic alternations to eliminate marked structure.

Bybee considers a very early conception of markedness, as presented in *SPE* (Chomsky and Halle 1968), alongside Willerman's (1994) proposed "complexity" metric, finding the latter to offer a more satisfactory explanation of some of the patterns found in her survey than markedness theory. However, there are two aspects of markedness constraints that make them well suited for understanding some of the patterns that Bybee found. First, markedness constraints can either be context-free or context-sensitive. Thus it is not the case that segments are simply indicated as marked or unmarked, as in *SPE*; context is crucial. Secondly, for some researchers markedness constraints can be grounded in the sense that there is often a phonetic explanation for why segments or segmental sequences are less common (Archangeli and Pulleyblank 1994; cf. de Lacy 2006 for a different view).

Let us take an example from Bybee (2005), presented to illustrate randomness. The consonants not found in affixes are indicated in parentheses, including [p g].

(7) *Consonant inventory of Tohono O'odham* (Bybee 2005: 178)[2]

(p)	t		(ʧ)	k	ʔ
b	d	ɖ	(ʤ)	(g)	
			(s)	ʂ	h
m	n		ɲ		
(w)	r (l)		j		

Given a rudimentary view of markedness, where voiced obstruents are more marked than voiceless (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION), the pattern above does indeed seem to be arbitrary: voiceless bilabial stops and voiced velar stops do not form a natural class. However, if one takes a phonetically grounded view of markedness, then we can consider the phonetic conditions under which voicing occurs. Whether or not the vocal folds "vibrate" has an aerodynamic underpinning, and the pattern of contrasts in Tohono O'odham appears to be a case of more natural (unmarked) segments emerging in affixes. The explanation is as follows. For voicing to occur at all, the vocal folds must be closed, and the air pressure above the glottis must be lower than the subglottal air pressure (Kingston and Diehl 1994). By Boyle's Law, the higher air pressure below the glottis will force the vocal folds open, causing periodic pulses (vocal fold vibration) to occur. With oral stops, air cannot leave the vocal tract during the closure phase. Because the total volume of air is held constant for stops, there is an interaction between the place of articulation (related to volume of air above the glottis) and the effort needed to maintain voicing. When a stop consonant is made closer to the front of the mouth, as with bilabial place, the volume of air is larger above the glottis and less effort is needed to maintain voicing than when a stop consonant is made closer to the glottis, as with velar place. Therefore, the more natural stops are voiced bilabial – more volume of air above the glottis facilitates voicing – and voiceless velar – less volume of air above the glottis inhibits voicing. The marked stops [p g] are missing from affixes. Because this explanation is based on a principle of aerodynamics, and there is only one feature involved in the contrast – [voice] – it is a good case in which we can separate complexity from phonetic markedness. A complexity approach cannot address why [voice] would be present with [labial] but not [dorsal].

Recall that for vowel contrasts, length, nasalization, and other features were excluded from affixes. If one views complexity as correlated with the number of features or structures present (μ, [nasal]), then this seems to support a complexity approach. However, when we consider languages with front rounded vowels, the issue is not as clear. If a language has [−back, −round] and [+back, +round] vowels, there is no a priori reason to assume that [−back, +round] is more complex than [+back, −round].[3]

[2] I represent the inventory following Saxton (1963), but with IPA symbols.

[3] This discussion is somewhat simplistic, as a full analysis would also need to consider whether features are binary or unary, and whether they are fully specified or not, as in research on underspecification theory (see CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION).

3.2 *Root–affix shapes*

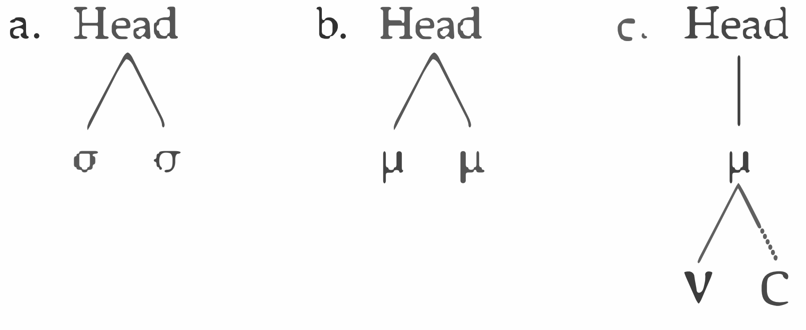
In many languages, affix morphemes have simple onsets, though complex and simplex onsets are widely attested in roots (Sanskrit, Tibetan). I have found no examples where only affixes have complex onsets while roots have simplex onsets. Once again this is an instance of affixes having a subset of the patterns found in roots.

A growing body of research indicates that roots have a special status in terms of their shape. In particular, Downing (2006) has argued extensively that roots, being heads of words, demand more complexity to their shape.

(8) HEADSBRANCH (Downing 2006: 122, adapted from Dresher and van der Hulst 1998)

Lexical heads (roots) must prosodically branch.

The representations in (9) all satisfy HEADSBRANCH.

(9)


Downing also proposes that roots, being monomorphemic heads, are predicted to be monosyllabic and to optimally satisfy branching by matching (9b) or (9c). Monosyllabicity follows from the MORPHEME–SYLLABLECORRELATION constraint, in which "each morpheme contains exactly one syllable" (Downing 2006: 120, adapted from Russell 1997: 121). The representation in (9a) would be ruled out in many languages if MORPHEME–SYLLABLECORRELATION is active. Downing's approach can account for different sizes of morphemes, but it is not clear how to extend this to syllable onset complexity. While one can view complex onsets as having branching structure, there doesn't seem to be evidence for a constraint that *requires* root syllables to have branching onsets.

A second type of asymmetry in morpheme shapes is related to a type of conspiracy in ensuring well-formed syllable structure over the word as a whole. In Salish and Bantu languages, the canonical root shape is CVC, though other root shapes exist, such as CVCC, CVCVC, etc. In these languages suffixes frequently have a -VC shape, so there is an asymmetry whereby roots are consonant-initial and suffixes are vowel-initial. When one considers that affixes by their very nature never occur in isolation, and that syllabification occurs to produce well-formed syllables, the end result is that no further codas are added to a word. So the addition of vowel-initial affixes to consonant-final roots seems to conspire to keep the number of codas the same. The following is a schema from a Salish type language to illustrate this (cf. (6a) *vs.* (6b) above).

(10)		UR		SR	
	a.	/CVC/	→	[CVC]	(1 coda)
	b.	/CVC + VC/	→	[CV.CVC]	(1 coda)

Finally, one might wonder about whether or not to include morphological accent in the section on contrasts. I have opted to include them in the section on alternations because the determination of whether a morpheme is accented or unaccented, and of how morphemes interact with each other, is a type of alteration.

In terms of morpheme shape, it is difficult to tease apart markedness from complexity. Clearly, complex onsets branch, which correlates with complex structure. However, it is not clear how that differs from our conceptions of syllable markedness.

4 Root–affix alternation asymmetries

In terms of investigating root–affix alternation asymmetries, the following processes have been examined: vowel harmony, consonant harmony, vowel elision, lenition, nasal place assimilation, voicing assimilation, and morphological accent systems. Vowel harmony seems to have special properties, which are related to how affixes are specified; this is the area where the most root–affix reversals have been pointed out in the theoretical literature (Baković 2000; Noske 2000; Pensalfini 2002; Ussishkin 2006). Therefore more space is dedicated to discussing this process than others. One preliminary finding of this survey is that there appear to be some processes that are more likely to be sensitive to the root–affix distinction than others. For example, I haven't found examples where nasal place assimilation, voicing assimilation, or lenition occur in affixes, but not in roots. If these types of processes occur at all, then they seem to occur across the board.

4.1 Vowel harmony

Vowel harmony is a process whereby the vowels within a particular domain (typically a phonological word) agree in their values of some (rarely all) vocalic features (see CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS; CHAPTER 118: TURKISH VOWEL HARMONY; CHAPTER 123: HUNGARIAN VOWEL HARMONY). There are two types of vowel harmony systems in the languages of the world: root- (or stem-) controlled and dominant–recessive systems (Baković 2000). An illustrative example of root-controlled harmony can be found in Akan (Niger-Congo), which has both prefixes and suffixes (Clements 1980). Following standard format, the affix vowels which are able to harmonize are indicated in capital letters in the input (indicating that they lack the harmonizing feature); the root is indicated with √.[4] As can be seen, both prefix and suffix vowels agree in [ATR] with the neighboring root vowel.

(11) *Akan* [ATR] *harmony* (Clements 1980, as cited in Baković 2000)

a. /E + √bu + O/ → [ebuo] 'nest' [+ATR] root
CLASS + nest + SFX

b. /E + √bʊ + O/ → [ɛbʊɔ] 'stone' [–ATR] root
CLASS + stone + SFX

[4] Based on the hypothesis that vowel features are only fully specified in roots, one could also argue that this is a case where roots show more contrasts than vowels.

c. /O + bE + √tu + I/ → [obetui] 'he came and dug (it)' [+ATR] root
3SG + come + dig + PAST

d. /O + bE + √tu + I/ → [ɔbɛtʊɪ] 'he came and threw (it)' [−ATR] root
3SG + come + throw + PAST

With dominant–recessive systems, a dominant vocalic feature is realized, regardless of whether it is in the root or affix. Furthermore, if it is present in both root and affix, then the affix feature prevails.[5]

A case of affix dominance is identified by Noske (2000), who proposes that Turkana (Eastern Nilotic) is an example of a language in which some suffixes trigger [ATR] harmony in the root. The relevant data are presented below. Observe in (12) that the root vowel is [−ATR]. However, when some suffixes are added with a [+ATR] vowel, then the root vowel alternates.

(12) *Turkana* (Noske 2000)

unsuffixed root		*suffixed root*	
a-k-dɔk	'to climb'	e-dok-eːn-e	's/he always climbs'
		e-dok-e	'way of climbing'
a-k-ɪmʊj	'to eat'	a-k-imuj-eːn	'to eat regularly'
		e-k-imuj-e	'way of eating'
a-dɛm-ar	'to take away'	e-dem-eːn-e	's/he always takes'
		e-dem-e	'way of taking'

The examples below crucially illustrate that the suffix's harmonizing feature can be [−ATR] as well as [+ATR].

(13) *Non-alternating* [−ATR] *suffixes in Turkana: Roots become* [−ATR]

unsuffixed root		*suffixed root*	
a-ki-lep	'to milk'	a-lɛp-ɔr	'to milk out'
a-ki-gol	'to close'	a-gɔl-ɔr	'to close out'
a-ki-bon	'to return'	a-bɔn-ɔr	'to return to a place'
a-ki-rem	'to spear'	ɛ-rɛm-ɛ-rɛ	'(why) is it speared?'
a-ki-mor	'to insult'	ɛ-mɔr-ɛ-rɛ	'(why) is he insulted?'

In trying to understand this system as a whole, it is important to compare the root vowels that enter into the harmony system with the affix vowels that participate in the [ATR] harmony. In doing so, we can see that the pattern of suffix-controlled harmony differs in a crucial way from root-controlled harmony. While the suffixes that undergo harmony can be triggered by both high and mid vowels (low vowels do not participate in harmony), only mid vowels on roots will agree with the [ATR] value of the suffix, as the following examples illustrate.

[5] Hansson (2001) observes that the majority, if not all, of the dominant–recessive harmony systems involve tongue root features [ATR] or [RTR].

(14) *High vowels are not affected* (Noske 2000: 78)

a.	a-bu-ɛt	'swelling (SG)'
	a-tub-ɛt	'judgment (SG)'
	a-tiŋ-ɛt	'handle (SG)'
b.	a-duŋ-ɔr	'to cut open'
	a-rɪp-ɔr	'to skim off'
	a-buk-ɔr	'to pour out'

This suggests that Turkana is actually a case of root alternations being a subset of affix alternations. Note that the set of vowels that participate in vowel harmony in the root is a subset of those that undergo harmony in the affix.

(15) *Summary of Turkana vowel harmony*

a. *Root-controlled*
Mid and high affix vowels alternate.

b. *Affix-controlled*
Mid root vowels alternate.

Secondly, there are several different types of suffix in Turkana: those in which the vowel freely participates in [ATR] harmony, and those in which it does not, so the set of affixes that trigger harmony is a subset of the affixes as a whole. To be a case of a true reversal, or even a complement set asymmetry, the alternation is expected only in the root, or the affix alternations should be a subset of the root alternations. These observations lead us to determine that the pattern found in Turkana is not an authentic instance in which the affix has more alternations than the root, but rather a specific type of RAA involving a dominant–recessive distinction with a specific, small set of affixes.[6]

4.2 Consonant harmony

While vowel harmony systems illustrate root control and affix dominance, only root control is found with consonant harmony systems (Hansson 2001). There are many instances in which affix consonants assimilate to root consonants in terms of place, manner, and voicing. Much rarer are cases of consonant harmony, whereby affix consonants assimilate entirely to the features of root consonants in the word. See also work on long-distance consonant assimilation by Rose and Walker (2004).

4.3 Other assimilations

Finally, we will make a quick examination of other types of assimilation. Vowels can assimilate to consonants in terms of manner of articulation – becoming nasalized vowels – and changes in vowel height. In many Salish languages, vowels lower in the context of a uvular consonant, which has a low tongue body. This

[6] Baković (2000) analyzes the Turkana system as a dominant-recessive harmony system, invoking the cycle to derive the correct harmony patterns. He also needs to refer to high-ranking specific suffix faithfulness constraints. Cf. Noske (2000).

process is almost universal in those languages that have uvular place of articulation. As far as I am aware, there are no languages in which the assimilation occurs with affix vowels but not roots; nor with root vowels but not affix vowels. In all the cases that I am aware of vowel lowering happens across the board.

In terms of consonants assimilating to vowels, some common processes include lenition (CHAPTER 66: LENITION) and palatalization (CHAPTER 71: PALATALIZATION). In a typological survey of lenition patterns, based on over 200 languages, Kirchner (1998) does not identify any root–affix asymmetries, though he does observe that lenition can be blocked in two environments: word-initially, and in the onset of a stressed syllable. These are two prominent positions identified by Beckman (1999) in her Positional Faithfulness model of phonological asymmetries (to be discussed further below).

Finally, some processes seem to be quite ubiquitous, applying freely anywhere within a word. Nasal place assimilation is an example. In this instance the nasal is almost universally the target of the assimilation rather than the following stop articulation. However, when nasal–stop sequences result in coalescence, whereby the output is a single nasal consonant that has the same place of articulation as the stop, Pater (1999) has found that coalescence does not always occur with root-initial segments. The failure of nasals to assimilate to the place of articulation of a following consonant can also be found in English with the prefix *un-*, which is external to the prosodic word.

4.4 Deletions

Deletions can affect both consonants and vowels (CHAPTER 68: DELETION). Let us start with vowel elision and the typological survey reported on in Casali (1997), based on 68 Niger-Congo languages and 19 other languages. The effects of vowel elision can be seen when morphemes or words combine. Casali observes that while it is widely claimed that V_1 elides universally in V_1–V_2 sequences, cases in which V_2 elides form a natural grouping in terms of belonging to a non-prominent position (see also CHAPTER 61: HIATUS RESOLUTION). The following summarizes the relevant findings presented in Casali (1997). It is important to point out that the summary below combines all findings; thus it represents the universal preference for V_1 to elide, as well as the preference for affixes and/or function words (FunctWord) to elide over roots and/or content words.

(16) *Vowel elision patterns* (Casali 1997: 496–497)

a.	LexWord LexWord	$V_1\ V_2 \rightarrow V_2$	no language in which only V_2 elides
b.	LexWord FunctWord	$V_1\ V_2 \rightarrow V_2$	
		$V_1\ V_2 \rightarrow V_1$	for 12 languages
c.	Prefix–root	$V_1\ V_2 \rightarrow V_2$	no language in which only V_2 elides
d.	Root–suffix	$V_1\ V_2 \rightarrow V_2$	
		$V_1\ V_2 \rightarrow V_1$	for 14 languages

As can be seen from the summary above, when both categories are the same as in (16a) only the first vowel elides: this illustrates the universal preference for V_1 to delete. However, when a root (or LexWord) is followed by an affix (or FunctWord), as in (16b) and (16d), then in some languages the second vowel elides:

the preference to elide a vowel in the affix (FunctWord) in these languages overrides the other tendencies. The pattern in (16c) illustrates that root vowels do not elide when they follow affix vowels. To summarize, the only condition in which V_2 does elide is when it is part of an affix (or FunctWord). This is consistent with the RAA subset pattern of alternations, where root vowel alternations are a subset of affix vowel alternations.

In terms of consonant deletion, St'at'imcets coronal–coronal sequences are resolved differently, based on the morphological affiliation of the segments. The data in (17a) illustrate that the "indirective" suffix /-xit/ ends with a coronal stop. In (17b) this suffix-final /t/ is deleted when followed by a coronal. In (17c), the roots retain the final /t/ when followed by a coronal.

(17) *St'at'imcets coronal–coronal sequences* (Interior Salish; Blake 1998: 2–3)[7]

a.	*Suffix–suffix*		
	/√nas-xit-aʃ/	[nas.xi.taʃ]	's/he gave it to him/her'
	/√hal'a-xit-aʃ/	[ha.l'a.xi.taʃ]	's/he showed it to him/her'
b.	*Suffix–suffix*	$C_1 C_2 \rightarrow C_2$	
	/√tɬ'ɛq-xit-tumuɬ/	[tɬ'ɛq.xi.tu.muɬ]	'bring it to us'
	/√tɬ'ɛq-xit-ʧ/	[tɬ'ɛq.xiʧ]	'bring it to me'
c.	*Root–suffix*	$C_1 C_2 \rightarrow C_1 C_2$	
	/n-√ʕʷɔj't=tən/	[nʕʷɔj't.tən]	'bed'
	/n-√pət=ʧ-an'/	[npət.ʧan']	'to cover an opening'

The pattern above parallels that found with vowel elision in terms of the first instance of an affix consonant deleting, while root segments are retained.

4.5 Morphological accent

Just as with vowel harmony, there are two types of morphological accent systems: root-controlled and dominant–recessive (Alderete 2001c). In morphological accent systems, morphemes are specified as accented or unaccented and/or as dominant or recessive. Accented morphemes have a stress in the input, while unaccented ones do not. Dominant morphemes are those that exert an influence over recessive morphemes, regardless of whether the affix is accented or unaccented.[8]

In terms of understanding morphological accent RAAs, Alderete (2001c) examines a wide array of morphological accent systems, finding that there is a preference for expressing root accents rather than affix accents. The system found in Cupeño (Uto-Aztecan) illustrates this. As can be seen below, when both roots and affixes are accented, the root accent will surface (18b).

[7] I have modified Blake's representation of the data, so that her original underlying representations have the same segmental allophones as the surface forms.

[8] Much as in harmony systems, dominant–recessive morphological accent systems are analyzed as involving cyclic application, which can erase previous morphological structure. See Czaykowska-Higgins (1993) for a detailed analysis of Nxa'amxcin (Interior Salish) stress.

(18) *Cupeño accented roots with accented affixes* (Alderete 2001a: 456)

a. /ˈpə + ˈ√miʔaw + lu/ → [pə-ˈmiʔaw-lu]
3SG + come + MOTION 'he came'
b. /ˈ√ʔaju + ˈqa / → [ˈʔajuqa]
want + PRES SG 'he wants'

The following illustrates that affix accent surfaces if the root is unaccented, providing evidence that affixes have an underlying accent.

(19) *Cupeño unaccented root with accented affixes*

a. /ˈpə + √jax/ → [ˈpə-jax]
3SG + say 'he says'
b. /nəʔən √jax + ˈqa/ → [nəʔən ja-ˈqaʔ]
1SG say + PRES SG 'I say'

However, recent work by Hargus and Beavert (2006) on Sahaptin has revealed a case of affix dominance. The following data illustrate that, when an accented root is preceded by an accented prefix, the prefix accent surfaces (20a). If the prefix is unaccented, then the root accent surfaces (20b).

(20) *Yakima Sahaptin morphological accent* (Hargus and Beavert 2006: 179)

a. /ˈpa-/ (INVERSE) /ˈpa-ˈ√watʼa-na/ [ˈpawatʼana] 'he struck at him'
b. /pa-/ (3PL NOM) /pa-ˈ√watʼa-na/ [paˈwatʼana] 'they struck'

This is clearly a case to look at more closely. In examining the system, a key difference was found between Sahaptin and a system like Cupeño, with root-controlled accent. While in Cupeño only some roots are accented, it turns out that in Sahaptin all roots have some form of lexical accent (Jacobs 1931; Hargus and Beavert 2006). Combining this with the fact that only some affixes have a lexical accent (20), it would seem that once again the affixes that are lexically specified for accent represent a subset of contrasts found on roots. If all roots are accented and root dominance prevailed, one would never see the effect of the affix accent at all.

To sum up the findings with respect to alternations in general, it appears that the types of alternations that roots undergo are a subset of the types of alternations affixes undergo. The reverse was not found in which affix alternations are a subset of root alternations. The only exceptions to this are instances of dominant–recessive vowel harmony or accent systems. The latter seem to be characterized as permitting only one value to appear within a particular domain, so they seem to have different properties than other types of alternations.

5 Approaches to root–affix asymmetries

Having provided an overview of the types of patterns found (or not) with respect to root–affix asymmetries, one may well ask whether they should be accounted for within a generative theory of grammar, i.e. as part of our linguistic competence (i-language; Chomsky 1986), or whether the randomness found by Bybee (2005) suggests that the patterns are the result of linguistic performance (e-language). By finding that some possible asymmetries are not attested, I hope to have teased apart some of the issues involved in undertaking research on RAAs and I suggest that this provides some evidence that grammatical principles are active in accounting for aspects of RAAs (though more research is needed to clarify what can be attributed to competence *vs.* performance).[9] The discussions of the approaches taken to accounting for RAAs thus assume that the asymmetry can be accounted for with grammatical principles. The approaches can be grouped into two broad categories: implicit approaches, which do not refer to root and affix directly, *vs.* explicit approaches, which do.

5.1 *Implicit approaches*

In implicit approaches to root–affix asymmetries, patterns arise due to mechanisms that do not refer directly to roots and affixes. This section surveys some of the ways in which RAAs can be achieved.

One example of an implicit approach was mentioned briefly at the outset of the chapter: directionality. In autosegmental phonology (Leben 1973; Goldsmith 1976) contrasting features (such as tones) are listed in the lexicon along with the relevant morphemes, but are not attached to segments. The features become properties of segments once they are linked to them via association conventions. The default direction of association is left-to-right. This direction of association is able to derive spreading of some listed feature of a morpheme to its right that lacks that feature. In the case of vowel harmony, this can account for root-controlled vowel harmony, when there is a following suffix, as in Turkish, but not for cases where prefixes also participate in the harmony, as in Akan.

A significant amount of research on the phonology–morphology interface has been undertaken in the theory of Lexical Phonology (Kiparsky 1982). In Lexical Phonology, words are built up incrementally by interspersing morphological operations with phonological processes; the various levels are often referred to as strata. There is evidence for (at least) two kinds of affixes in many languages: stem-level (non-neutral) are associated with changes to a base, and word-level (neutral) are not associated with changes to a base (CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME). Given that phonological rules can apply after affixes are added, this approach can account for processes that affect affixes, but not roots: for the root to not be affected, the rule does not apply until after an affix is added. A key issue in the application of phonological rules is how to prevent an alternation from applying to the whole word, leaving the root unaffected. The failure of some rules to be able to look back into the root is known as non-derived environment blocking (NDEB; see CHAPTER 88: DERIVED ENVIRONMENT EFFECTS). A well-known example that illustrates the phenomenon is found in Finnish, in which assibilation occurs when roots and affixes are concatenated, turning /t-i/ into [si], as can be seen in (21a). This process does not occur root-internally, but only at root–affix junctures (21b).

[9] See de Lacy's (2006) proposal and discussion of evidence that some types of phonological markedness are part of competence and others are part of performance.

(21) *Finnish NDEB* (Kiparsky 1993)

a.	/tilat-i/	→	[tilasi]	'order-PAST'	*[tilati, silasi]
b.	/vaati-vat/	→	[vaativat]	'demand-3PL'	*[vaasivat]

The failure of root-internal /ti/ sequences to undergo an alternation could also be viewed as a subtype of root–affix asymmetry.

Another approach to maintaining the identity of roots falls under the general umbrella of paradigm uniformity effects. Because roots form the basis of many paradigms, and we know that paradigm pressure can inhibit the application of processes (see CHAPTER 83: PARADIGMS), this is also a potential area to investigate RAAs.

Finally, some patterns of root–affix reversals are related to analyses that propose a dominant–recessive system for roots and affixes. As mentioned in footnotes 6 and 8, cyclicity is one way to account for dominant–recessive morphemes. Another way to view dominant–recessive systems is that only the feature associated with the dominant affix will surface within a phonological word. If this is true, then the types of alternations found with dominant–recessive systems are fundamentally different from other types of alternations. Having a fundamentally different approach to analyzing root–affix reversals, distinct from other types of processes, is therefore not unexpected, given the larger picture of the properties of RAAs.

5.2 *Explicit approaches*

In explicit approaches to accounting for root–affix asymmetries, reference is made directly to roots and affixes, as is the case with morpheme structure conditions/constraints (Halle 1959; Stanley 1967) and in research within Optimality Theory (OT; Prince and Smolensky 1993).

Morpheme structure conditions (MSCs) involve statements about the phonotactic patterns of morphemes in a language. As such, they encode information about roots and affixes. The formulation of MSCs relates also to what is redundant or predictable. For example, with Turkish vowel harmony, roots would be specified for all their vocalic features, while affixes would not. By directly encoding predictable information in MSCs, differences between roots and affixes emerge, even though MSCs do not directly refer to root–affix asymmetries. (See CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS for further discussion of MSCs.)

Perhaps the largest body of research that analyzes RAAs directly has been carried out in OT. McCarthy and Prince (1995, 1999) propose that the identity of morphologically related strings (input–output, base–reduplicant, and word–word) is mediated by a correspondence relation. They refer to this approach as Correspondence Theory (CT). CT is not a necessary component of OT; however, this framework has led to a growing body of theoretical research on RAAs. Monitoring the identity of strings, such as between input and output (underlying representation and surface representation), is done through the use of faithfulness constraints, so any alternation correlates with a violation of some faithfulness constraint. In identifying the types of correspondence relations that can exist, McCarthy and Prince propose that the correspondence relation is relativized to roots and affixes. Several OT approaches make use of CT in analyzing RAAs; these will be discussed in the remainder of this section.

In order to account for some of the observations noted above – that roots tend to have more marked contrasts and to resist alternations – McCarthy and Prince propose the following universal ranking.

(22) *Root-affix faithfulness metaconstraint* (RAFM; McCarthy and Prince 1995: 364)

FAITH-Root >> FAITH-Affix

In OT, constraints are ranked and violable; any deviation from identity between strings results in a violation of FAITH. Lower-ranked constraints are more likely to be violated, and the location of some markedness constraint with respect to FAITH-Root or FAITH-Affix can compel alternations resulting in asymmetrical patterns. For example, the following constraint ranking accounts for the Sanskrit RAA presented in (2), whereby roots have complex onsets, but affixes do not (McCarthy and Prince 1995: 365).

(23) FAITH-Root >> *COMPLEX >> FAITH-Affix

Because markedness constraints (among others) can intervene between these two FAITH constraints, the universal ranking in (23) predicts that affixes will be the target of alternations and will exhibit less marked structure than roots. Thus this general approach has a way to capture exactly the two types of asymmetries found. While a great deal of literature has supported this general approach (Casali 1997; Beckman 1999; Alderete 2001a, 2001b, 2001c, 2003; Urbanczyk 2001; Pensalfini 2002), there has also been a growing body of research which proposes that the ranking in (22) is not universal (Noske 2000; Ussishkin 2005, 2006; Hargus and Beavert 2006).

A modification of how to obtain root–affix asymmetries in OT, known as positional faithfulness (PF), is developed by Beckman (1999) to accommodate a range of privileged positions. Beckman proposes that roots are just one of several positions within a word, such as root, initial syllable, and stressed syllable, which illustrate more contrasts and are resistant to change. Instead of having faithfulness specific to affixes and adopting the RAFM in (22), Beckman proposes that there is generalized faithfulness along with several specialized faithfulness constraints, which are relevant to different prominent positions. To recast the constraints in (23) into a PF approach, one would have the following ranking.

(24) *Positional faithfulness approach to Sanskrit*

FAITH-Root >> *COMPLEX >> FAITH

The way that the FAITH constraints are evaluated in PF is that *every* change in the identity of a string – regardless of whether it is a privileged position or not – will incur a violation of FAITH. If there is a change in the identity of the privileged position – in this case roots – that will incur a violation of both FAITH-Root and FAITH. This means that violations of privileged positions will incur two violations, inherently worse than violations in a non-privileged position. The ranking of FAITH-Root over FAITH is not proposed to be universal, and is instead emergent. This key difference between how faithfulness violations are calculated in RAFM and PF approaches translates into a difference of complement set *vs.* subset asymmetries. While in PF the violations of FAITH-Root are literally a subset of the violations of FAITH, in the RAFM approach the violations of FAITH-Root cannot be a subset of the violations incurred by FAITH-AFFIX.

It is significant to note that PF is well suited to analyzing instances of RAAs. Let us consider how constraint re-ranking – a core feature of OT – can derive just the subset pattern. In OT, unless there is a fixed universal ranking (as with the RAFM), all possible permutations of constraint rankings are possible. If, for any phonological phenomenon being investigated, there are two types of faithfulness constraints (FAITH and FAITH-Root) and one markedness constraint (*M), this gives us 3!, i.e. six (3 × 2 × 1), possible rankings. It turns out that several of these rankings will result in the same pattern. For example, as long as FAITH dominates *M, the language will have the marked structure M throughout the entire language – roots and affixes alike (25a). The following table summarizes how the six different rankings map onto the patterns of a language, where specification of root *vs.* affix is indicated in the first two columns: M indicates that the marked feature is found, while ¬M indicates that it is not found. For alternations, A indicates that an alternation occurs, while ¬A indicates no such alternation.

(25) *Factorial typology*

	root	*affix*	*constraint ranking*		
a.	M, ¬A	M, ¬A	FAITH	>> FAITH-Root	>> *M
			FAITH	>> *M	>>FAITH-Root
			FAITH-Root	>> FAITH	>> *M
b.	¬M, A	¬M, A	*M	>> FAITH	>> FAITH-Root
			*M	>> FAITH-Root	>> FAITH
c.	M, ¬A	¬M, A	FAITH-Root	>> *M	>>FAITH

Only three patterns are generated. In three rankings, (25a), roots and affixes have the marked structure, but not the alternation. Two constraint rankings, (25b), produce a pattern where both roots and affixes do not have the marked structure, but do have the alternation. And one ranking, (25c), results in unmarked structure emerging in the affix, along with an alternation; this is "the Emergence of the Unmarked" (TETU) effect showing up in non-roots (for more about TETU, see CHAPTER 58: THE EMERGENCE OF THE UNMARKED). As one can see, there is no ranking such that affixes will have the marked structure (M) not found in roots or such that roots will exhibit an alternation (A) not occurring in affixes. This matches the types of RAAs found. The complement set type of asymmetry cannot be derived, because either there is a difference between roots and affixes, as with the TETU ranking in (25c), or roots and affixes are the same, as in (25a) and (25b). There is nothing that will favor markedness in affix classes by constraint ranking alone (though see Pater 2000, 2009 and Gouskova 2007 for discussion of indexed constraints for specific morphemes).

An alternative OT approach to privileged positions is Positional Markedness (Zoll 1998). Rather than having faithfulness constraints relativized to position, markedness constraints are relativized to positions. So the Sanskrit pattern would have the following constraints.

(26) *Positional Markedness approach to Sanskrit*

*COMPLEX >> FAITH >> *COMPLEX$_{ROOT}$

See Alderete (2003) for arguments against the Positional Markedness approach for Navajo RAAs.

A further approach to RAAs in OT includes research on domains and morpheme-specific alternations within a co-phonology framework (Inkelas *et al.* 1997; Inkelas 1999). In co-phonologies, each morpheme can be associated with a different constraint ranking. There can be one ranking for roots and another for affixes. The ability to freely re-rank constraints for a specific morpheme can result in systems where complement set asymmetries can be derived. A language in which voiced stops are present only in affixes can be derived by the following ranking.

(27)	a.	Root co-phonology	*VOICED-STOP	>> FAITH
	b.	Affix co-phonology	FAITH	>> *VOICED-STOP

The survey conducted here was not able to verify any instance of a complement set asymmetry. So it would seem that, without some internal constraints, the co-phonology model overgenerates possible RAAs.

6 Form–function mappings

One of the secondary themes regarding RAAs relates to form–function mappings. Three topics will be examined briefly: non-concatenative morphology (§6.1), morphologically conditioned alternations (§6.2), and psycholinguistic grounding (§6.3).

6.1 Non-concatenative morphology

As discussed in §2, there is a range of non-concatenative morphological processes that make the identification of roots and affixes a challenging task. Some of these have been touched upon above. However, numerous other processes are relevant as well, including root-and-pattern morphology, reduplication, vowel and consonant mutations, subsegmental affixes, and tone (see CHAPTER 105: TIER SEGREGATION for a survey of many of these patterns from the perspective of tier segregation).

The discussion begins by examining a purported case of affix dominance in Hebrew (Ussishkin 2005, 2006). As is well known, Hebrew words can be formed by a system of intercalcating root consonants and affix vowels, in the form of *binyanim* (see CHAPTER 108: SEMITIC TEMPLATES). The following illustrates the basic system for /gadal/ 'to grow'.

(28) *Hebrew verbal paradigm* (Ussishkin 2006: 114)

underived lexical representation	*base form + affix*		*derived form*	
/gadal/	[gadal]			'he grew'
		[i e]	[gidel]	'he raised'
		[u a]	[gudal]	'he was raised'
		[hi i]	[higdil]	'he enlarged'
		[hu a]	[hugdal]	'he was enlarged'

Working within an OT framework, Ussishkin proposes that the affix vowels in the input must be parsed in the output, arguing that this is a case in which Faith-Affix dominates Faith-Root.

However, there are two reasons to believe that this is not an authentic counterexample to RAAs. First, the expression of the affix vowels is compelled by a morphological criterion, not a phonological one. Failure to parse the input vowels would result in homophony throughout the system, as discussed by Ussishkin. Secondly, there is an assumption that the vowels in the input stem are part of the stem and not part of any affix. However, the entire Semitic *binyanim* system of morphological contrasts is based on the assumption that the pattern of contrasts in consonants is linked with roots, while that of vowels is associated with affixal meanings. Why would the vocalic melody /a-a/ in /gadal/ not be associated with any meaning in just this one word, while in all the other words in the paradigm it is associated with some type of aspectual meaning? As indicated in Ussishkin (2005: 200), the /a-a/ cell in the system of vowel contrasts is associated with "PAST TENSE" meaning. Therefore, this is not an actual counterexample to the trend for affixes to be more likely to be subject to alternations. In actuality, the vowels that are lost are associated with affixal meaning as well.

Reduplication is a morphological process whereby part of a word is copied, in order to indicate a difference in meaning (CHAPTER 100: REDUPLICATION; CHAPTER 119: REDUPLICATION IN SANSKRIT). There is a great deal of evidence that reduplication targets roots rather than affixes (see McCarthy and Prince 1993; Urbanczyk 2001, 2007). For example, in Axininca Campa root segments are copied over affix segments (McCarthy and Prince 1993). There is a disyllabic minimum for reduplication. If the root meets the disyllabic target, then only root segments are copied (29a). However, if the root is subminimal, then it is augmented to meet minimality requirements, or the prefix material is recruited for copying (29b).

(29) *Axininca Campa reduplication* (McCarthy and Prince 1993: 63–64)

	stem size	*without prefix*	*with prefix*		
a.	stem ≥ disyllable	kawosi-kawosi-	noŋ-kawosi-kawosi	'bathe'	
		kintha-kintha-	noŋ-kintha-kintha	'tell'	
b.	stem < disyllable	paa-paa-	no-wa-no-wa-	'feed'	/p-/
		nata-nata-	no-naa-no-naa-	'carry'	/na-/

6.2 *Morphologically conditioned alternations*

Languages which exhibit consonant mutations and other types of stem modifications have been excluded from the discussion of purely phonological alternations (see CHAPTER 82: FEATURAL AFFIXES; CHAPTER 65: CONSONANT MUTATION; CHAPTER 117: CELTIC MUTATIONS). It is interesting to note that, for the most part, these types of alternations seem to target roots rather than affixes, and also to create marked segments rather than eliminate them. Recall that, in Halkomelem, progressive morphology is accompanied by resonant glottalization (6), creating an extremely rare class of sounds cross-linguistically.

6.3 *Psycholinguistic grounding*

There is a range of psycholinguistic evidence to show that roots are accessed differently from affixes as reported in Beckman (1999). In addition, Ussishkin and Wedel (2002) discuss a psycholinguistically grounded explanation for RAAs on the basis of the concept of effective contrast. On the basis of the hypothesis that the ease of lexical retrieval is facilitated by token frequency and neighborhood density (Luce 1986; Luce and Pisoni 1988), Ussishkin and Wedel propose that an effective contrast is one that balances these two factors. High-frequency words with few neighbors are the easiest to access, while low-frequency words with many similar neighbors are the most difficult to access. Because there are many more meanings of roots to be encoded, having sparse neighborhoods and distinct sounds will aid in lexical recovery. This has implications for patterns found in Hebrew and Nuu-chah-nulth.

In Hebrew, the affixal vocalic melodies utilize the bulk of the available affix space. With five distinctive vowels and two possible slots, there are 25 possible combinations, of which there are 17 distinct meanings (Ussishkin 2006).

Languages like Nuu-chah-nulth (Southern Wakashan), with over 400 suffixes, illustrate more patterns than languages with relatively few affixes, suggesting that it is also important to consider the proportion of roots to affixes when considering markedness patterns. Lee and Urbanczyk (2006) found a strong correlation between phonetic similarity of the affixes (parallel to neighborhood density) and whether or not an affix triggers stem modifications, such as vowel lengthening or reduplication.

7 Summary

This paper has investigated RAAs, with two key goals: to tease apart some of the complications associated with studying RAAs, and to discuss the various theoretical approaches. In doing so, some unexpected patterns emerged. First, apparent cases of RAA reversals are correlated with dominant–recessive morphological accent and vowel harmony systems. Analyses which aim to fit these into the RAFM approach – a model of root–affix correspondences which is complementary – are forced to give up the universal ranking of Faith-Root over Faith-Affix. Secondly, some types of processes were found to be less susceptible to RAAs. And finally, the range of RAAs found seems to suggest that some patterns are due to competence while others are due to performance. It is hoped that further research into root–affix asymmetries will undertake significantly detailed case studies, in an attempt to verify the findings reported here.

ACKNOWLEDGMENTS

Many thanks to Ewa Czaykowska-Higgins and Sunghwa Lee for discussion of some of the points in this paper, and to the editors of the *Companion* for helpful feedback in revising it. Research was funded in part by a University of Victoria Internal Research Grant and SSHRC 410-2003-1523 awarded to the author, who bears the responsibility for any factual errors or oversights.

REFERENCES

Alderete, John. 2001a. Root-controlled accent in Cupeño. *Natural Language and Linguistic Theory* 19. 455–502.

Alderete, John. 2001b. Dominance effects as transderivational anti-faithfulness. *Phonology* 18. 201–253.

Alderete, John. 2001c. *Morphologically governed accent in Optimality Theory*. New York and London: Routledge.

Alderete, John. 2003. Structural disparities in Navajo word domains: A case for LEXCAT-FAITHFULNESS. *The Linguistic Review* 20. 111–157.

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Baković, Eric. 2000. Harmony, dominance, and control. Ph.D. dissertation, Rutgers University (ROA-360).

Bates, Dawn, Thom Hess & Vi Hilbert. 1994. *Lushootseed dictionary*. Seattle: University of Washington Press.

Beckman, Jill N. 1999. *Positional faithfulness: An optimality theoretic treatment of phonological asymmetries*. New York: Garland.

Blake, Susan J. 1998. The OCP and root-affix faithfulness in St'at'imcets (Lillooet Salish). Unpublished ms., University of British Columbia.

Bybee, Joan. 2005. Restrictions on phonemes in affixes. A crosslinguistic test of a popular hypothesis. *Linguistic Typology* 9. 165–222.

Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: Chicago University Press.

Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73. 493–533.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1980. *Vowel harmony in nonlinear generative phonology: An autosegmental model*. Bloomington: Indiana University Linguistics Club.

Czaykowska-Higgins, Ewa. 1993. Cyclicity and stress in Moses-Columbia Salish (Nxa'amxcin). *Natural Language and Linguistic Theory* 11. 197–278.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.

Downing, Laura J. 2006. *Canonical forms in prosodic morphology*. Oxford: Oxford University Press.

Dresher, B. Elan & Harry van der Hulst. 1998. Head–dependent asymmetries in phonology: Complexity and visibility. *Phonology* 15. 317–352.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.

Golston, Chris. 1996. Direct Optimality Theory: Representation as pure markedness. *Language* 72. 713–748.

Gouskova, Maria. 2007. The reduplicative template in Tonkawa. *Phonology* 24. 367–396.

Greenberg, Joseph H. 1966. *Language universals, with special reference to feature hierarchies*. The Hague: Mouton.

Halle, Morris. 1959. *The sound pattern of Russian: A linguistic and acoustical investigation*. The Hague: Mouton.

Hansson, Gunnar Ólafur. 2001. Theoretical and typological issues in consonant harmony. Ph.D. dissertation, University of California, Berkeley.

Hargus, Sharon & Virginia Beavert. 2006. High-ranking affix faithfulness in Yakima Sahaptin. *Proceedings of the West Coast Conference on Formal Linguistics* 25. 177–185.

Inkelas, Sharon. 1999. Exceptional stress-attracting suffixes in Turkish: Representations versus the grammar. In Kager *et al.* (1999), 134–187.
Inkelas, Sharon, Cemil Orhan Orgun & Cheryl Zoll. 1997. The implications of lexical exceptions for the nature of grammar. In Iggy Roca (ed.) *Derivations and constraints in phonology*, 393–418. Oxford: Clarendon Press.
Jacobs, Melville. 1931. A sketch of Northern Sahaptin grammar. *University of Washington Publications in Anthropology* 4. 85–291.
Jakobson, Roman. 1965. Quest for the essence of language. *Diogenes* 13. 21–37.
Kager, René, Harry van der Hulst & Wim Zonneveld (eds.) 1999. *The prosody–morphology interface*. Cambridge: Cambridge University Press.
Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70. 419–454.
Kiparsky, Paul. 1982. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.
Kiparsky, Paul. 1993. Blocking in nonderived environments. In Sharon Hargus & Ellen M. Kaisse (eds.) *Studies in Lexical Phonology*, 277–313. San Diego: Academic Press.
Kirchner, Robert. 1998. An effort-based approach to consonant lenition. Ph.D. dissertation, University of California, Los Angeles.
Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT.
Lee, Sunghwa & Suzanne Urbanczyk. 2006. Reduplication and vowel lengthening/shortening in Nuuchahnulth. Paper presented at the 2nd International Conference on Wakashan Languages, University of Victoria.
Luce, Paul A. 1986. Neighborhoods of words in the mental lexicon. Ph.D. dissertation, Indiana University.
Luce, Paul A. & David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19. 1–36.
McCarthy, John J. 2005. Optimal paradigms. In Laura J. Downing, T. A. Hall & Renate Raffelsiefen (eds.) *Paradigms in phonological theory*, 170–210. Oxford: Oxford University Press.
McCarthy, John J. & Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University (ROA-482).
McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.
McCarthy, John J. & Alan Prince. 1999. Faithfulness and identity in prosodic morphology. In Kager *et al.* (1999), 218–309.
Michelson, Karin. 1989. Invisibility: Vowels without a timing slot in Mohawk. In Donna Gerdts & Karin Michelson (eds.) *Theoretical perspectives on Native American languages*, 38–69. New York: SUNY Press.
Noske, Manuela. 2000. [ATR] harmony in Turkana. A case of Faith-Suffix >> Faith-Root. *Natural Language and Linguistic Theory* 18. 771–812.
Parker, Steve & David Weber. 1996. Glottalized and aspirated stops in Cuzco Quechua. *International Journal of American Linguistics* 62. 70–85.
Pater, Joe. 1999. Austronesian nasal substitution and other NÇ effects. In Kager *et al.* (1999), 310–343.
Pater, Joe. 2000. Non-uniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* 17. 237–274.
Pater, Joe. 2009. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In Steve Parker (ed.) *Phonological argumentation: Essays on evidence and motivation*, 123–154. London: Equinox.
Pensalfini, Rob. 2002. Vowel harmony in Jingulu. *Lingua* 112. 561–586.
Piggott, Glyne L. 1995. Epenthesis and syllable weight. *Natural Language and Linguistic Theory* 13. 283–326.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University and University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Rose, Sharon & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80. 475–531.

Russell, Kevin. 1997. Optimality Theory and morphology. In Diana Archangeli & D. Terence Langendoen (eds.) *Optimality Theory: An overview*, 102–133. Cambridge, MA & Oxford: Blackwell.

Saxton, Dean. 1963. Papago phonemes. *International Journal of American Linguistics* 29. 29–35.

Stanley, Richard S. 1967. Redundancy rules in phonology. *Language* 43. 393–436.

Steriade, Donca. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5. 73–155.

Steriade, Donca. 1995. Underspecification and markedness. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 114–174. Cambridge, MA & Oxford: Blackwell.

Suttles, Wayne. 2004. *Musqueam reference grammar*. Vancouver: University of British Columbia Press.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.

Urbanczyk, Suzanne. 2001. *Patterns of reduplication in Lushootseed*. New York: Garland.

Urbanczyk, Suzanne. 2006. Reduplicative form and the root–affix asymmetry. *Natural Language and Linguistic Theory* 24. 179–240.

Urbanczyk, Suzanne. 2007. Reduplication. In Paul de Lacy (ed.) *Cambridge handbook of phonology*, 473–493. Cambridge: Cambridge University Press.

Ussishkin, Adam. 2005. A fixed prosodic theory of nonconcatenative templatic morphology. *Natural Language and Linguistic Theory* 23. 169–218.

Ussishkin, Adam. 2006. Affix-favored contrast inequity and psycholinguistic grounding for non-concatenative morphology. *Morphology* 16. 107–125.

Ussishkin, Adam & Andrew Wedel. 2002. Neighborhood density and the root–affix distinction. *Papers from the Annual Meeting of the North East Linguistic Society* 32. 539–549.

Willerman, Raquel. 1994. The phonetics of pronouns: Articulatory bases of markedness. Ph.D. dissertation, University of Texas, Austin.

Wiltschko, Martina 2005. The syntax of precategorial roots. *UBC Working Papers in Linguistics* 17. 245–258.

Zoll, Cheryl. 1998. Positional asymmetries and licensing. Unpublished ms., MIT (ROA-282).

105 Tier Segregation

Adam Ussishkin

1 Introduction: Tier segregation as phonological representation

At or near the top of the job description of the phonologist is the goal of finding, describing, and explaining patterns of behavior among features, sounds, and other phonological elements in languages of the world. For theories stemming from the generative approach that began most conspicuously with Chomsky and Halle (1968, henceforth *SPE*), the most highly valued explanations are rooted in Universal Grammar (UG). *SPE*-based phonology centered on alternations in phonological patterns, and one of its primary contributions to linguistic theory was to shift the burden of the explanatory basis for such patterns away from the lexicon and onto the grammar (the hypothesized computational system responsible for the relationship and mapping between the hypothesized lexical representations and observed surface forms). Therefore, all evidence for such impoverishment of lexical representations is *indirect* evidence adduced from patterns and alternations observed on the surface. The cost of reliance on such indirect evidence was nonetheless thought by generative phonologists to be outweighed by the explanatory benefit that resulted from the ability to pare down lexical representations and concomitantly endow the grammar with a rich array of rules (or constraints, depending on one's theoretical commitment) that provided structure necessary for the mapping from underlying to surface form.

With a firm commitment to the bare bones, abstract lexical representation with no redundant information encoded, *SPE*-based phonology and its practitioners in the 1960s and 1970s focused on the content and form of grammatical rules, where the majority of the phonological action was to be found. A crucial (and quite obvious) benefit of this approach was the ability to grammatically relate variant forms of the same underlying abstract elements, such as the phoneme and the morpheme (CHAPTER 11: THE PHONEME; CHAPTER 1: UNDERLYING REPRESENTATIONS). A simple example at the phoneme level can be found in the following pairs of words from Maltese, a Semitic language spoken on the island nation of Malta, where voiceless and voiced obstruents turn out to be in complementary distribution:

(1) *Maltese voicing assimilation* (data from Borg 1975)[1]

voiceless obstruent		*voiced obstruent*	
hapta	'knock (N)'	habat	'he knocked'
aatsa	'ducking'	oodos	'he ducked'
hasat	'he startled'	hazda	'fright'
fadal	'it remained'	vdaal	'remains'

Such examples of alternation are quite common cross-linguistically, and can be further subdivided into various principled types; such details need not concern us here. The important point about such examples for the generative phonologist is that they establish the need for abstract, underlying elements such as the phoneme and the morpheme, which are the primary ingredients for word formation. Rather than listing the underlying form of the Maltese morpheme for 'knock' as /hapta/ and 'he knocked' as /habat/, the fact that the two words share some core set of semantics is captured by the fact that they are formed from the same lexical element. Whatever this lexical element turns out to be, it should nonetheless contain only the unpredictable information. Therefore, the voicing specification of the labial obstruent in the lexical representation of these words will either be underspecified or specified in a marked way (see CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION for more on underspecification). The surface form of the consonant in question is context-dependent, and determined by the computational system, or grammar, that operates on the concatenation of elements in word formation, and not on the underlying representation, which is consequently simpler. From a morphological perspective, the paired words above can be observed to share their consonants, and so in a similar vein the underlying morphological representation of each form might reflect this by storing a set of non-contiguous consonants as a lexical item; more on this type of representation will be discussed below.

Early generative phonology maintained this focus on representational issues, though for the most part the debates on representation centered on morphophonemic structure and the abstract representations of phonemes and morphemes in the lexicon, as illustrated in the Maltese examples above. However, the groundbreaking work on tone and tonal alternations by Leben (1973) literally introduced a new level of analysis, in which tonal elements were viewed not as segment-delimited specifications within a feature matrix, but rather as freed from the grasp of the segment and allowed to behave as independent units on a tonal *tier*. An important result from this work is the genesis of the Obligatory Contour Principle (OCP), taken to be responsible for tonal distribution patterns (e.g. surface tone patterns in Mende, as analyzed by Leben 1973). In its original guise, the OCP was defined by Leben as follows:

(2) *Obligatory Contour Principle*[2]

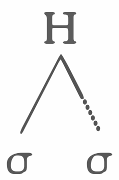
At the melodic level, adjacent identical elements are prohibited.

[1] Both voiced and voiceless obstruents are phonemic in Maltese. This alternation is an example of contrast neutralization.

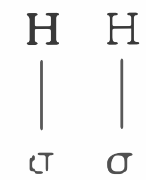
[2] While the OCP is a crucial motivating ingredient in the development of tier segregation, it is nonetheless not the main focus of this chapter. See CHAPTER 14: AUTOSEGMENTS.

The OCP is invoked as an explanation for the existence of tone patterns such as H+, HL+, LHL, L+, and LH+ in Mende,[3] and the exclusion of patterns such as HHL and LLH in the language. Note the repeating elements in these latter, unattested cases. It is only under a view in which tones are freed from their segmental manifestations that the argument becomes clear: if tones occupy a distinct tier, it is possible to capture the asymmetry in distribution of tone patterns. Thus, Leben's work invoked a more complex, non-linear style of representation. In this representation, sequences of the same tone are represented by a multiply linked tonal element (3a), rather than multiple tonal elements with single linking (3b).

(3) a. *Multiple linking* (respects OCP)



b. *Single linking* (violates OCP)



In addition to the OCP, non-linear representation was also capable of capturing the fact that, when some phonological alternation requires a change in tone, all elements in that sequence of tone tend to change. The argument requires additional assumptions regarding representation; namely, that tones and segments are arrayed on separate tiers, and the invocation of the various conditions:

(4) a. *Association Convention*
Associate tones to syllables one-to-one and left-to-right.

b. *Well-Formedness Condition*
At each stage in the derivation, all syllables are associated with at least one tone, and all tones are associated with at least one syllable.

c. *No-Crossing Condition*
Association lines may not cross.

This is the original use of tier segregation, though not the one focused on in this chapter. Nonetheless, its genesis was a crucial step in the development of non-linear models of phonology, and led to the subsequent application of tier segregation at the featural level in phonology and to the segregation of vowels and consonants into separate tiers in Semitic and other languages.

In the tier-segregated view, a representation consists not of a string of concatenated, independent feature matrices, but rather is composed of a set of multi-tiered elements linked by association lines. This more complex and non-linear representation permits principles of association, spreading and the OCP as explanatory devices for patterns difficult to explain in the strictly linear model. The non-linear view found further support in subsequent work, most importantly Goldsmith (1976), McCarthy (1979, 1981, 1989, 2000), and Clements and Keyser (1983). Goldsmith (1976) generalized Leben's (1973) model to all phonological interactions, capturing much more intuitively the notion of featural alternations, especially phonological assimilation. The behavior of certain phonological features appears suspiciously similar to tonal behavior, so that in fact the extension to these

[3] Kleene "+" in these representation refers to a tone that associates to a syllable and to any following syllables.

features seems logical once the data are viewed under the autosegmental lens. In fact, this logical extension applied not only to local (or adjacent) environments, but additionally to seemingly non-adjacent sets of elements, thus allowing for a unified explanation of both local and long-distance phonological behavior.

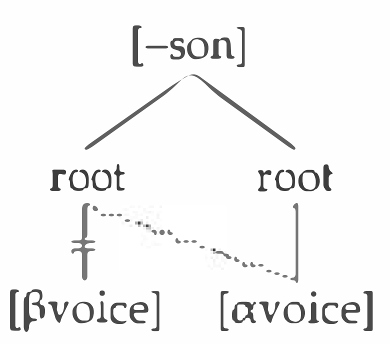
The primary focus of this chapter on tier segregation is the segregation of consonants and vowels into distinct tiers. This idea has its roots in the autosegmental approach of Goldsmith (1976), who, based on work on the behavior of tone in tone languages, proposed extending the notion of tiers beyond tones to the phonological domain in general. The earlier generative approach to phonological behaviors such as Maltese voicing assimilation was the application of linear rules to an underlying or intermediate representation. Phonological rules capturing this sort of behavior have the generic form A → B / C __ D, where A refers to a structural description, B to a structural change and the portion of the rule following the / to the phonological environment. For instance, returning to the case of voicing assimilation in Maltese as presented above, a linear rule could be written as follows (CHAPTER 81: LOCAL ASSIMILATION):

(5) *Maltese voicing assimilation* (*SPE*-style)

[−sonorant] → [αvoice] / __ [−sonorant, αvoice]

Such a rule successfully describes the alternation, and for *SPE*-style phonologists would be seen as explanatorily adequate, but proponents of autosegmental phonology would argue that it misses a crucial fact. Note the double occurrence of the feature [αvoice] in the rule: once in the structural change, and a second time in the phonological environment, or trigger. Use of the Greek letter variable notation forces the phonologist to hypothesize that, in fact, there are not two separate voicing specifications in the representation (as the linear rule would lead us to believe) but rather that the voicing feature of the trigger *spreads* to the target under this type of assimilation behavior. In other words, despite the double occurrence of the changing feature, there really is only a single token of it at work here. Such an approach involves a more complex representation, which is needed to capture the idea that segments are composed of features that themselves have autonomy; that is, that are not strictly bound to the segment they are underlyingly associated with (hence the name *autosegmental*). An autosegmental rule, then, involves a type of decomposition along with a higher degree of complexity, and for the case of voicing assimilation could appear as follows:

(6) *Maltese voicing assimilation* (autosegmental)



Note that unlike the linear rule, which explicitly denotes the input and output of a phonological process by explicitly referring to a structural description (to the left of the arrow) and a structural change (to the right of the arrow), the autosegmental rule collapses these into the same representation, where the structural change is denoted by the dashed line indicating the spreading of the voicing feature regressively from the second segment (segments are represented here as non-sonorant root nodes) and by the dissociation of the first segment's original voicing specification (which is not explicitly indicated in the linear rule).

In the late 1970s, the representational force of this sort of rule was carried even further by McCarthy (1979, 1981, 1989, 2000), whose prosodic model of non-concatenative morphology resulted in yet further unification: Semitic languages and their well-known root-and-pattern morphology could also be analyzed as involving autosegments; in other words, the idea of autosegmental representation was lifted to the level of the morpheme. The result of this work is consonant–vowel tier segregation, the best-known example of tier segregation, and the one that this chapter will focus on. On McCarthy's view, various patterns specific to Semitic were explained (e.g. OCP effects, morpheme structure constraints, and long-distance consonantal spreading/copying), while invoking a simple yet natural additional extension of the autosegmental view to the morpheme level. The notions of the consonantal root, the vocalic pattern affix, and the prosodic template follow naturally from this extension. At the same time, the tier-segregated representation opened the door to a more advanced understanding of complex phonological behaviors in another well-known set of data outside Semitic: templatic languages that lack root-and-pattern morphology, such as Yawelmani Yokuts (Archangeli 1983, 1984, 1985).

In fact, the influence of autosegmental representation and tier segregation was so great that many of the most important results in phonology in the 1980s stemmed from this revolutionary representation. During this period, an enormous amount of attention was devoted to the representational details of the tier segregation model, leading to a growth industry in feature geometry, the nature of UG-provided structures, and ultimately to other models, such as Articulatory Phonology (Browman and Goldstein 1986, 1989; among others), in which the notion of the segment itself was almost so diluted that it seemed epiphenomenal.

While interest in the representation of the internal structure of segments has comparatively waned since the advent of models such as Articulatory Phonology (see CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATION), and especially since Optimality Theory (henceforth OT; Prince and Smolensky 1993), the issue shaped the field of phonology quite dramatically for a significant period. In this chapter, tier segregation will be explored in detail: §2 reviews evidence for tier segregation from Semitic data; §3 provides further evidence, with data from non-Semitic languages; §4 presents some of the more current debates regarding tier segregation, along with external evidence that bears on the issue; and finally, §5 concludes the chapter.

2 The evidence for tier segregation from Semitic

2.1 Roots and patterns

In his 1979 dissertation and subsequent 1981 article, McCarthy took non-linear representation an enormous step further. The basis of his contribution comes from Semitic languages (McCarthy focuses mainly on Classical Arabic and Tiberian

Hebrew), which have long been characterized by their non-concatenative morphological behavior (dubbed "root-and-pattern morphology" by McCarthy). In these languages, much of the word structure displays a strikingly unusual characteristic: rather than forming words from sequences of contiguous elements, words appear to have a non-concatenative structure in which the consistent phonological material within a given paradigm is non-contiguous. This can be illustrated with a typical paradigm from the verbal system of Modern Hebrew:

(7) *Hebrew verbal paradigm*

verb	*template*	
ganav	CVCVC	'to steal'
higniv	CVCCVC	'to smuggle in'
hitganev	CVCCVCVC	'to sneak in'
nignav	CVCCVC	'to be stolen'

Basic morphological analysis reveals that the consistent phonological material in the paradigm is the discontiguous string of consonants /gnv/, which in the traditional root-and-pattern approach to Semitic is called a consonantal root and has the status of a morpheme. Prior to McCarthy, the consonantal root had long been recognized as the likely center of gravity in the Semitic lexicon (cf. Chomsky 1951; but see also work from much earlier, including de Alcala 1505; de Volney 1787; de Sacy 1810; Bopp 1824; Ewald 1827; Gesenius 1910; Gaon 1932, 1942, 1969; Harris 1941). It is clear why this is the case: superficially at least, the consonantal root appears to have an obvious and consistent semantics associated with the examples given above (something to do with stealing or sneaking). What's more, a similar case can be made for the parts of each word that form the complement to the consonantal root. For instance, in the word /higniv/, there is clearly a causative meaning, and causatives frequently occur with the same /hi-i/ pattern combined with a consonantal root (for the sake of comparison, a related (non-causative) form is also provided for each example):

(8) *Hebrew causatives with* /hi-i/

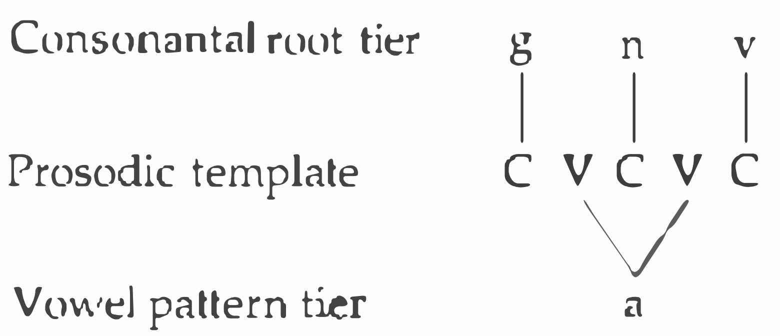
verb		*related form*	
higniv	'to smuggle in'	ganav	'to steal'
hizrik	'to inject'	zarak	'to throw'
hixnis	'to bring in'	nixnas	'to enter'
hirdim	'to put to sleep'	nirdam	'to fall asleep'

Not every verb with /hi-i/ is causative (e.g. /hirgiʃ/ 'to feel'; /himtin/ 'to wait'), and not every causative verb has /hi-i/ (e.g. /limed/ 'to teach, cause to study'; /simeax/ 'to make happy'), but nonetheless verbs in a particular class (known as a *binyan*, the Hebrew word for building or structure) all demonstrate the same phonological behavior and pattern together morphologically.[4] In order to formalize the notions of consonantal root morphemes and pattern morphemes,

[4] See Laks (2006) for a contemporary view of such relations from a syntactic-morphophonological point of view.

McCarthy's (1979, 1981) theory of non-concatenative morphology proposes C/V tier segregation based on morphological grounds: each morpheme is said to occupy its own tier in the representation. Lexically, as far as stem formation is concerned, three types of morpheme are found: consonantal root morphemes, pattern morphemes, and prosodic templates.[5] Word formation involves the concatenation of these units according to autosegmental principles (association of segmental elements to appropriate prosodic positions, going one-to-one from left to right). A form like Hebrew /ganav/ therefore has the following representation in the phonology after association of morphemes:

(9) *Tier-segregated representation of /ganav/*



There are several things worth noting about this type of representation. First of all, note the multiply associated vowel /a/, which is linked to both V positions in the prosodic template. This multiple association is a necessary consequence of the OCP, which would prohibit a representation containing two tokens of the vowel /a/ located on the same tier (because identical adjacent elements are prohibited by the OCP). Secondly, this representation does not capture the pronunciation of the word /ganav/ directly, because it is not a linear representation. In later work, McCarthy (1986) cites Younes's (1983) concept of tier conflation, a mechanism responsible for linearizing a tier-segregated representation in order to pronounce it. Tier conflation is essentially the bridge between the autosegmental, tier-segregated non-linear representation, and the surface form, which is viewed as linear.[6]

2.2 *Templates*

Another motivation for McCarthy's approach – and the reason tier segregation was so widely adopted following his theoretical innovations – was that it was able to capture the fact that consonants and vowels occur in different combinations, depending on the particular pattern that a consonantal root is found in. In other words, the precedence relations between consonants and vowels change, depending on the template they associate to. If rules of association govern these precedence relations, then it would be redundant to record this information in the lexicon as well. Tier segregation achieves the goal of redundancy avoidance by remaining silent on the issue of precedence relations among elements belonging to distinct morphemes. To illustrate the argument, recall the paradigm above for

[5] Hebrew also has the more familiar type of concatenative morphology in the form of prefixes and suffixes (as do all Semitic languages), in addition to its non-concatenative structure. These prefixes and suffixes tend to be correlated with inflectional morphology.

[6] See Bat-El (1988) for an important discussion and review of tier conflation and its consequences.

/gnv/: sometimes all three consonants are separated by vowels (e.g. /ganav/ 'to steal'), other times the first two consonants of the root are adjacent (e.g. /nignav/ 'to be stolen'). The ordering relationships between elements from different morphemes are determined by the prosodic template, and therefore fall into the category of predictable information (for a review and current assessment, see CHAPTER 108: SEMITIC TEMPLATES). Even in a more current view (one beginning with McCarthy and Prince 1986, 1990, 1993), in which the prosodic template is no longer lexically specified but rather is derived from independent grammatical principles (cf. Ussishkin 2000, 2005), it is nonetheless possible to predict the relative placement of consonants and vowels. No matter the template to which a consonantal root morpheme or a vowel melody morpheme is associated, the ordering of elements *within each morpheme* is the only unpredictable and consistent information, and thus the only ordering-related information encoded lexically. Essentially, the benefit of tier segregation is that since the grammar will provide the information regarding ordering of consonants and vowels via association, that information can be removed from the lexicon. As McCarthy (1989: 88) points out:

> planar V/C segregation emerges from the redundancy of linear order relations between vowels and consonants in languages with templatic morphology [not just Semitic but also others like Yawelmani] or sufficiently rigid constraints on canonical form.

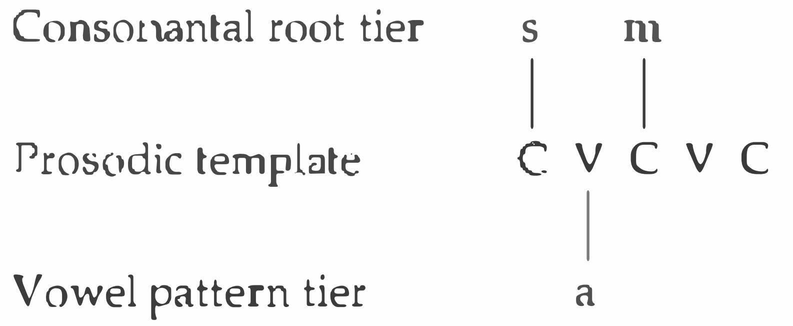
2.3 *The OCP, morpheme structure constraints and alternation*

Further motivation for tier segregation is offered by some well-known distributional facts regarding Semitic word structure (McCarthy 1979, 1981; and many others): a word never contains the same consonant tautomorphemically in both the first and second C positions of a template, while words frequently contain the same consonant tautomorphemically in the second and third positions.[7] The well-known examples cited by McCarthy from Arabic are the non-existent form */sasam/ and the commonly encountered type of form such as /samam/ 'to poison'. Under an approach that recognizes the consonantal root and the segregation of morphemic tiers, forms such as /samam/ are formed from the root /sm/ associated to a CVCVC template in which the /m/ is multiply associated to both the second and the third C positions. This is a natural outcome under autosegmental phonology, in which the default direction of association is left to right. Once the /s/ associates to the first C and the /m/ associates to the second C, the third C is left without segmental content, so the /m/ spreads again to the third C position (spreading the /s/ to that position is impossible because doing so would violate the constraint against line-crossing). This mechanism is known as "long-distance consonantal spreading." The following diagrams illustrate each step in the derivation.

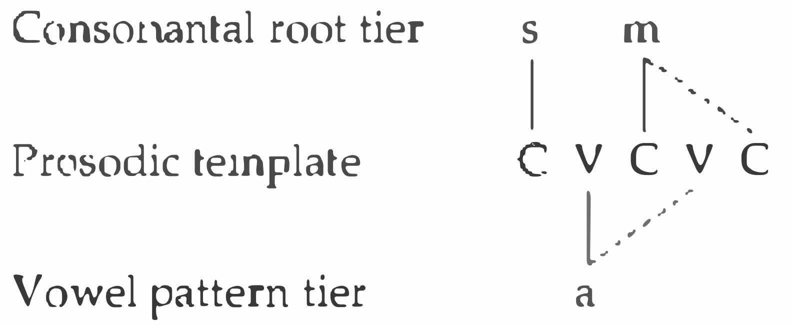
[7] Many thanks to an anonymous reviewer for pointing out that some Semitic languages have exceptions to this generalization; e.g. Hebrew /mimen/ 'to finance'. Since these exceptions are fairly uncommon, most previous research has tended to ignore them.

(10) *Deriving Arabic* /samam/

a. *Associate elements one-to-one, left-to-right*



b. *Fill empty templatic positions via long-distance spreading*



Under this model, it is impossible to generate unattested */sasam/-type forms, but attested /samam/-type forms should be the only possible output from biconsonantal roots. To complete the argument, it is necessary to rule out unattested */sasam/-type forms generated from a root like /ssm/ – such a root is easily ruled impossible, since its underlying representation violates the OCP by the presence of two adjacent, identical consonants. The same holds for an underlying representation like /smm/ (for a more general discussion of morpheme structure constraints, see CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS). This argument is perhaps the strongest argument in favor of tier segregation, since without tier segregation the OCP violation is not triggered.[8] The only way to generate the attested type (/samam/) is with a biconsonantal root and one-to-one, left-to-right association, resulting in the multiple linking of the second consonant to the second and third C positions.

Another well-known piece of evidence for the consonantal root discussed by McCarthy (1981) and subsequently cited by many authors (e.g. Hume 1994; Odden 1994; Suzuki 1998; Rose 2000) is found in Akkadian labial dissimilation. In Akkadian, the prefix /ma-/ is realized as [na-] if the stem it attaches to contains a labial consonant in the root. That is, the place specification of the prefix alternates depending on whether or not the root of the stem it attaches to contains a labial consonant. This type of alternation can be viewed as triggered by the OCP:

(11) *Akkadian labial dissimilation* (data from Rose 2000)

Akkadian noun	
ma-zuukt	'mortar'
ma-ʃkanu-m	'place'
na-phar	'totality'
na-rkabt	'chariot'

[8] See Rose (2000) for a reformulation of the OCP that would count both */sasam/ and /samam/ as OCP-violating forms.

The first two examples above contain roots with no labial consonants, and thus the nominalizing prefix is realized with the labial [m] (note that the suffix /in/ in the second case does not trigger dissimilation of the prefix). In the last two examples, there is a labial root consonant, and the prefixal consonant is realized as [n] as a result.

2.4 *Language games and other external evidence*

McCarthy (1981) also discusses evidence from languages games that appears to argue in favor of the consonantal root occupying its own tier at some level of representation. At issue are forms from the Bedouin Hijazi dialect of Arabic brought to light by Al-Mozainy (1981). As an example, McCarthy cites the variant forms of the verb /difaʕna/ 'we pushed': [daʕafna, fidaʕna, ʕadafna, faʕadna, ʕafadna]. The argument here is that without tier segregation, there is no way to array just the root consonants separately in a way that allows for their permutation as a target of the language game process. This is because the game distinguishes between root consonants (e.g. /dfʕ/ in these examples), which can permutate, and non-root consonants (e.g. /n/ in these examples), which cannot.

A similar type of argument is put forward by Prunet *et al.* (2000) and Idrissi *et al.* (2008). These researchers document a series of metathesis errors in a bilingual French–Arabic aphasic patient; of relevance is the observation that these errors only take place in the subject's Arabic utterances and never in his French utterances. Further, the errors involve metathesis of root consonants only; it is never the case that vowels or affixal consonants are metathesized. The authors take this as clear evidence for tier-segregated consonantal roots along the same lines as McCarthy and others, though Ussishkin (2006) rejects their account and proposes a constraint-based alternative that does not require reference to the consonantal root.

In closing, this section has provided many cases of evidence in favor of tier segregation based on phonological and morphological behavior in Semitic languages. In the following section, evidence from other languages is considered.

3 Non-Semitic evidence for tier segregation

3.1 *Yawelmani*

Evidence from non-Semitic languages has also provided arguments in favor of tier segregation. Archangeli's work on a number of dialects of the California Penutian language Yokuts is the best known, and will be the focus of the discussion and data here. In Yawelmani Yokuts, morphology looks strikingly Semitic; there appears to be, as in Semitic, a distinction between consonants, vowels, and the prosodic templates in which the consonants and vowels are realized. The template selected is dependent on morphology: essentially, the suffix selects the template. Archangeli (1983, 1984, 1985) provides stem alternations from Yawelmani illustrating multiple templates for verbs with three consonants:

(12) *Yawelmani CV templates* (adapted from Archangeli 1983: 351) *and data* (stems are underlined)

template		
CVCC	ʔamtsʼit	'be near'
CVVCC	deejilhin[9]	'guard'
CVCVVC	bineetit	'ask'

This pattern, in which the stem templates alternate, is persistent throughout the language, and it is typically the case that the suffix appended to the stem determines which template is chosen for the surface form. A tier-segregated representation, in which consonants and vowels of a morphological root occupy distinct tiers, captures the fact that the order of vowels and consonants is template-dependent and predictable, thus allowing for lexical representations that are rendered simpler by the absence of prespecified precedence relations between vowels and consonants. If the templates supply those orderings, and the templates are independently necessary, Yawelmani can be viewed much like Semitic, where the templates also supply the orderings between consonants and vowels. A crucial difference, of course, is that Yawelmani lacks the morphological distinction between consonants and vowels; both the vowel(s) and the consonants are viewed as part of the same morpheme in the cases above, yet nonetheless occupy separate tiers.[10]

3.2 Rotuman

Although he later rejects it (McCarthy 2000; see discussion below) as an example of tier segregation, McCarthy (1989) strongly suggests that the Oceanic language Rotuman provides evidence from outside the Semitic languages in favor of tier segregation. In this case, the data come from an alternation involving morphological metathesis (CHAPTER 59: METATHESIS).[11] Rotuman has a morphological distinction between what are known as different phases, of which there are two: the complete phase and the incomplete phase,[12] their occurrence being governed by syntactic and semantic principles (see McCarthy 2000: App. B for a review of Churchward's 1940 rules for use of the different phases). The two phases can be distinguished phonologically by a number of different alternations: final vowel deletion, umlaut, diphthongization, and, importantly for our purposes, metathesis (there are also cases in which forms in the two phases are homophonous). McCarthy (2000) gives the following examples of phase distinction by metathesis.

[9] The first [i] in this form is epenthetic, and therefore appears to disrupt the template. For much discussion on this issue, see Archangeli (1983).

[10] While the use of the term "tier" is maintained here, it is important to acknowledge Archangeli's (1985) more precise formulation of the distinction between "tier" and "plane." Much of the literature following Archangeli (1985) subsequently adopted the term "plane" for what is being termed here a "tier," and for the sake of consistency in this chapter I will continue to refer to "tiers."

[11] Though see Hume (1991) for metathesis evidence from Maltese that argues against McCarthy's (1989) claims for planar segregation.

[12] McCarthy (1986) states that the morphological difference between the complete *vs.* the incomplete phase corresponds to "a kind of free versus bound form, respectively."

(13) *Rotuman complete and incomplete phases distinguished by metathesis*

complete	*incomplete*	
iʔa	iaʔ	'fish'
seseva	seseav	'erroneous'
hosa	hoas	'flower'
pure	puer	'to rule'
parofita	parofiat	'prophet'

McCarthy (1989) considers Rotuman as a possible (but not, in his words, "iron-clad") case favoring tier segregation (referred to in McCarthy 1989 as "planar V/C segregation"). Despite the fact that this language has no templatic morphology, and no morphological distinction between consonants and vowels, allowing tier segregation in this language can explain the metathesis evidence in the phase distinctions above. In terms of linear order, with tier segregation in place in Rotuman, the metathesis can be viewed as templatic, since for each of the pairs of forms above the consonants always occur in the same linear order, as do the vowels; this is true because metathesis always involves a consonant and a vowel in this language. In other words, the precedence relations between consonants are preserved, as are the precedence relations between vowels. Tier segregation allows for this, though McCarthy acknowledges that other mechanisms could explain it as well. What differentiates the phases, then, is something template-like. Depending upon which phase is being selected by the grammar, the consonant–vowel ordering will differ.

This concludes the section in which non-Semitic evidence for tier segregation is discussed. In the following section, some debates concerning tier segregation are reviewed.

4 Debates

The extension of tier segregation beyond Semitic is not uncontroversial. The principal reasons behind this controversy can be divided into two principal types. One type concerns the non-morphological motivations for tier segregation in non-Semitic languages like Yawelmani and Rotuman. The second type is based on more recent developments in phonological theory.

4.1 *Non-Semitic languages*

4.1.1 Yokuts

Steriade (1986) attacks ideas from Archangeli (1984, 1985) concerning tier segregation in Yokuts. In particular, Steriade argues against separate tiers for consonants and vowels when the language has no morphological distinction between them. The idea advanced in Steriade (1986) is that only morphologically motivated tier segregation is legitimate. Since Yokuts does not distinguish consonants and vowels morphologically, her view is that tier segregation in Yokuts is not licit. An important consequence of this argument is that Steriade must explain the templatic effects so easily captured via tier segregation. She proposes an alternative analysis, in which a mechanism of "melody copy" (analogous to copying processes in reduplication) explains the templatic patterns found in this type of language. Prince (1987) responds to Steriade's approach with a vigorous argument against Steriade and in favor of Archangeli's original proposals for tier segregation. Again focusing primarily on Yokuts, with some additional Arabic and Hebrew evidence, Prince (1987) defends Archangeli's claims that tier segregation is *phonologically* motivated.[13]

4.1.2 Rotuman

Arguing against claims found in previous work, McCarthy (2000) rejects Rotuman as a possible case of tier segregation. Recall that the metathesis process encountered in the complete and incomplete phases in this language seems like a good candidate for tier segregation, since the two phases could be thought of as templatic. McCarthy's more recent (2000) account is couched within OT, and argues that

> OT takes much of the burden of explanation off of representations (e.g. tier segregation) and places it on substantive constraints . . . [t]hus, consonant–vowel tier segregation is completely superfluous in an Optimality Theoretic analysis of Rotuman, and in fact it is antithetical to fundamental premises of OT.

This is because of OT's assumption of Richness of the Base, which states that restricting the form of input material is disallowed. Under this assumption, every potential input form must be considered a legitimate input form, and despite this potential variation in inputs the actual attested output must nonetheless emerge as optimal. But with tier segregation, an OT analysis yields undesirable results, since with tier segregation it becomes impossible to capture the contrast between pairs such as /usi/ 'bush (sp.)' and /sui/ 'bone', which form a contrastive pair despite sharing the same consonant and the same vowel sequence in a tier-segregated view. That is because, under the tier-segregated view, the vowels would be represented separately from the consonants, yielding the same input (/ui + s/) for these two surface-contrastive pairs. McCarthy goes on to show that various attempts to patch the holes resulting from such an approach turn out to be unsuccessful, and concludes that "[n]o superficial technical fix is appropriate, because the problem derives directly from the core assumption of tier segregation theory."

Essentially, tier segregation is incompatible with OT, provided the basic assumption in OT that input forms may not be restricted, and that linguistic properties must be viewed as the result of output-oriented constraints. The argument continues, McCarthy suggests, as an even more fundamental result of this output-oriented focus: all underspecification, tier segregation included, is highly undermined by OT's principle/ingredient of output-based constraints. Since tier segregation underspecifies the linear order relations between consonants and vowels (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION), any tier-segregated input is an underspecified input. In earlier approaches, on the one hand, this was viewed as beneficial, since the linear order relations in question are predictable, and in pre-OT generative phonology no predictable information

[13] For another paper arguing against consonant–vowel tier segregation on a strictly phonological basis, see Odden (1987), who argues that McCarthy's (1982) proposals for phonologically based tier segregation in the Munda language Gta? are unmotivated and predict unattested data. Odden (1987) proposes an alternative analysis based on two rules: one of initial melody linking, and another (optional) rule of melody spreading.

is granted the status and luxury of residing in underlying representations. But OT, on the other hand, argues that the content of constraints, rather than the elegance of representations, bears responsibility for phonological behaviors, and thus tier segregation should be dispensed with.

4.2 *Semitic languages*

A number of researchers have proposed rejecting tier segregation even for Semitic languages, which formed the original case for tier segregation in the late 70s and early 80s. These researchers have various reasons for rejecting tier segregation, ranging from concerns regarding the highly abstract nature of lexical repres-entation under tier segregation (e.g. Ussishkin 2000, 2005, 2006) to the need to reformulate principles that tier segregation seemed to stem from originally, like the OCP (Rose 2000). Some of these arguments will be reviewed here.

Tier segregation relates directly to lexical storage of linguistic units and much scholarly activity on Semitic languages over the last decade or so has focused on lexical representations in Semitic. Bat-El (1994) was among the earliest researchers to propose that Semitic consonantal roots were explanatorily inadequate, and based her arguments on Hebrew data in which consonant clusters are preserved in the formation of verbs from nouns, such as the following:

(14) *Hebrew denominal verb formation with cluster transfer* (Bat-El 1994: 577–578; preserved clusters underlined)

noun		*denominal verb*	
transfer	'transfer'	trinsfer	'to transfer'
sandlar	'shoemaker'	sindler	'to make shoes'
psanter	'piano'	psinter	'to play the piano'
nostalgia	'nostalgia'	nistelg	'to be nostalgic'

If denominal verb formation were dependent on consonantal roots, it would be impossible to explain why the sets of adjacent consonants in these pairs are preserved in the related verbs, rather than all conforming to a uniform template. This is because, under an account based on tier segregation, the consonantal roots, on their own tier, are not stored lexically with any information regarding which consonants are surface-adjacent. Further, Ussishkin (1999) noted that another set of denominal verbs in Hebrew consistently preserves information from the vowel in the base – information that is not available from the base consonants alone:

(15) *Hebrew denominal verb formation with vowel feature transfer* (Ussishkin 1999: 405; corresponding vowel–consonant transfer effects underlined)

noun		*denominal verb*	
tik	'file'	tijek	'to file'
ˀir	'city'	ˀijer	'to urbanize'
sug	'kind, type'	siveg	'to classify'
hon	'capital, wealth'	hiven	'to capitalize'

Again, reliance on the consonantal root is insufficient to explain this phenomenon, because if the consonants are the only material that matters, then information about the vowel of the base is absent. Nonetheless, for these forms, there is a clear correlation between the vowel in the base and the medial consonant in the verb. Bat-El and Ussishkin both conclude that the Hebrew lexicon may not encode consonantal roots, thus invalidating an original motivation for tier segregation. They propose that, for denominal verbs, the lexicon stores stems or whole words, in which case no level of representation involves segregations of vowels and consonants into distinct tiers. In subsequent work, both Bat-El (2003) and Ussishkin (2000, 2005, 2006) extend these results and propose such a view not just for these verbs but for the entire language. That is, the lexicon stores stems or words, not consonantal roots, and these stems or words form the basis for morphological relationships and derivations.

Similar conclusions are reached for Semitic by Gafos (1998), who focuses on apparent spreading in the Malaysian language Temiar. Gafos (1998) concentrates on the theoretical device known as "long-distance consonantal spreading," which was previously thought (cf. McCarthy 1979, 1981, 1989, 2000) to be responsible for patterns such as doubled consonants in Semitic verbs (e.g. the discussion of Arabic /samam/ above). In Gafos's proposal, however, such long-distance spreading is rejected as an explanation for forms like /samam/, which are instead reformulated as examples of segmental copying rather than segmental spreading. Under the alternative copying-based account, no level of representation involves a multiply linked consonant (seen in (10b) above). The penultimate section of Gafos (1998) argues against tier segregation. The argument, like that of McCarthy (2000), is theory-internal from an OT perspective, and centers on the fact that within OT, properties of lexical forms do not result from underspecification, but rather from the interaction of universal constraints on surface forms.[14] This is essentially the same Richness of the Base argument proposed by McCarthy regarding Rotuman. Since tier segregation was originally motivated as a way to avoid lexical redundancy by avoiding lexical storage of predictable information, it falls under the rubric of underspecification and is therefore, at the very least, subject to deep scrutiny under OT. And indeed, Gafos concludes that tier segregation should be eliminated altogether. Gafos goes on to argue that not only does this apply at the phonological level (e.g. for languages with no morphological distinction between consonants and vowels, such as Yokuts), but that it should also apply at the morphological level, to Semitic languages in which one might posit a morphological distinction between consonants and vowels. Even this distinction can be captured via universal, output-oriented constraints, Gafos claims, and thus tier segregation receives no support.

In order for such a view to be maintained, alternative explanations for various phenomena are required. Gafos, for instance, proposes that, rather than viewing doubled consonants in Semitic as resulting from a spreading operation, we should see them as the result of a copying operation analogous to the copying seen in reduplication phenomena (CHAPTER 100: REDUPLICATION; CHAPTER 119: REDUPLICATION IN SANSKRIT). This argument echoes the proposals of Steriade

[14] See also Hudson (1986) for a tier-free analysis of Semitic, with a focus on Arabic. A related aspect of Hudson's (1986) analysis is that it foreshadows later OT accounts with its emphasis on output representations.

(1986) in her argument against Archangeli's phonological tier segregation, and finds further support in Ussishkin's (1999) treatment of doubled consonants in Hebrew denominal verbs, where a subset of verbs double the second base consonant to fill a verbal template:

(16) *Doubled consonants in Hebrew denominal verb formation* (Ussishkin 1999)

noun		*denominal verb*	
dam	'blood'	dimem	'to bleed'
sam	'drug'	simem	'to drug, poison'
xad	'sharp'	xided	'to sharpen'
mana	'portion'	minen	'to apportion'

Again, rather than viewing the repeated consonant in each denominal verb as resulting from long-distance spreading over a vowel position in a template, the repeated consonant is simply viewed as a copy. The template still plays an important role here, in that it is responsible for inducing the copy, but no spreading is invoked to explain the doubled consonant.

Another OT-based argument arguing against tier segregation can be found in Rose's (2000) reformulation of the OCP, which bears direct relevance to this discussion. In Rose's view, the OCP applies across intervening vowels to penalize identical consonants such as those seen above in consonant doubling denominal verbs in Hebrew. Rose additionally claims that consonant doubling results from copying, and not spreading, and also relies on a rejection of segregating vowels and consonants into distinct tiers. Focusing on evidence from Ethio-Semitic languages, Rose shows that a reformulation of the OCP into a surface-based constraint provides exactly the explanatory adequacy that is needed while at the same time avoiding the various pitfalls resulting from a combination of tier segregation and the input-based version of the OCP that would be incompatible with any output-oriented approach.

4.3 *Behavioral evidence*

Various experiments in Semitic have been reported in the literature that bear on tier segregation, and the results of these experiments are mixed, so it is impossible to adjudicate definitively based on these results. It is also important to note that these experiments have the explicit goal of testing the nature of lexical representations; in particular, whether lexical representations are root-based or word-based. The assumption being made in such work, and in this chapter as well, is that root-based representations are congruent with tier segregation, while word-based representations are not. The best-known studies have been carried out with Hebrew and Arabic (Frost *et al.* 1997; Deutsch *et al.* 1998; Frost *et al.* 2000 on Hebrew; Boudelaa and Marslen-Wilson 2000, 2001, 2004 on Arabic) using a variety of techniques, most commonly a masked priming lexical decision task using visual stimuli.[15] These

[15] In this task, subjects respond to pairs of stimuli consisting of a masked prime that they do not consciously perceive, followed by a target that they do perceive, and are asked to make a lexical decision about the target (i.e. decide whether the target is a word of their language or not). For instance, in a single trial the masked prime might consist of the Hebrew word /higdil/ 'to enlarge' followed by the target /gidel/ 'to raise'.

studies consistently demonstrate that when a target word shares a consonantal root with a masked prime, there is a significant priming effect manifested in a faster reaction time for such target items than when the masked prime does not share the consonantal root. This robust finding supports lexical storage of roots, and therefore supports a form of tier segregation. Conflicting results exist with respect to the decomposition of vocalic patterns: Frost *et al.* (1997) found no priming effect for vocalic patterns, while Deutsch *et al.* (1998) did find such priming. As for results that might argue against tier-segregated representations lexically, one reported study on Hebrew using visual stimuli (Moscoso del Prado-Martín *et al.* 2005) found a word frequency effect (which supports word-based storage and therefore does not support tier segregation), though the same study also found an effect of morphological family size, which supports representing consonantal roots in the lexicon.

These studies are not without their problems. First, all stimuli were presented visually. Because the orthographies of Hebrew and Arabic do not contain graphemes for many of vowels, visually presented primes and targets may bias participants to process consonants preferentially. For example, because there is no letter for the sound [a] in Hebrew, the word for 'to write', pronounced [katav], is written as *ktv* in Hebrew orthography. Visual presentation therefore introduces a very serious confound for any study that aims to see whether consonants play a special role in processing. Secondly, these studies have been carried out on languages with a consonantal orthography in which all educated native speakers receive explicit pedagogical instruction based on the consonantal root. Because of this, it is not entirely clear that studies are probing online processing of lexical units; earlier work has shown that literacy affects lexical organization (e.g. Morais *et al.* 1986). Another weakness of these earlier studies is that they do not manipulate root or word frequency, which would be the ideal probing mechanism for distinguishing root from whole-word access.

For Arabic, Boudelaa and Marslen-Wilson (2000, 2001, 2004) report consistent effects of priming by consonant-vowel patterns as well as by roots and patterns separately in a series of auditory lexical decision tasks with auditory primes (Marslen-Wilson and Zhou 1999). These results come from work on both Standard Arabic and Southern Tunisian colloquial Arabic, and, due to the fully auditory nature of both primes and stimuli, they avoid orthographic confounds inherent in any study that employs visual stimuli. The results suggest a decompositional route for lexical processing, thus supporting a tier-segregated view. One published study on Maltese (Ussishkin and Twist 2009) reports mixed results: in a visual masked priming experiment, a significant priming effect for consonantal roots was found, but the same paper also reports an auditory experiment whose results support storage of whole forms.

In recent work on Hebrew, Berent *et al.* (2007) provide experimental evidence that whole-word representations in the lexicon are needed in order to account for speakers' acceptability ratings of words in different paradigms that share identical consonantal roots. Such results argue against tier segregation, and although they are consistent with lexical storage of words, previous work on acceptability ratings (Berent *et al.* 2001, 2002, 2004) is consistent with both word-based and root-based storage. Finally, other work on Hebrew that supports word storage is reported in Sumner (2003), who ran auditory lexical decision experiments with auditory priming to test lexical access for a class of irregular verbs in Hebrew known as weak verbs. These are verbs whose final root consonant does not occur in pronounced forms. Sumner's results show that, for one class of weak verbs, there is evidence for whole-word storage, while, for the second class of weak verbs, age turns out to be a factor: younger speakers exhibit evidence of whole-word storage, while older adult speakers show an inhibitory effect of weak verb primes.

Despite their importance, these psycholinguistic studies leave open the issue of lexical representation. Given the range of evidence in favor and against tier segregation based on these studies, one tenable hypothesis is that the lexicon can store both root-based and word-based representations, thus allowing for the possibility of tier segregation.

More general studies also leave open this possibility. In a series of experiments testing learnability of statistically regular patterns, Newport and Aslin (2004) exposed subjects to invented words in a nonsense language, and manipulated the parts of the words that were consistent across stimuli. When subjects were exposed to stimuli consisting of non-adjacent regularities in the consonants (Newport and Aslin 2004: Experiment 2), they readily learned the patterns. These results contrast with those from words where non-adjacent syllables, rather than the consonants, were consistent. In the non-adjacent syllable case, subjects performed poorly. The fact that subjects were able to learn the pattern when it was manifested in non-adjacent consonants shows at the very least that humans can attend to such regularities, thus indirectly supporting an ability to segregate consonants from vowels. Similar results obtained by Newport and Aslin (2004: Experiment 3) show that, when non-adjacent vocalic patterns were held constant, subjects readily learned the pattern. Again, this result shows at the very least that subjects are capable of attending to regularities that hold across non-adjacent sets of segments.[16]

5 Conclusion

This chapter has focused on introducing and discussing the theoretical and representational concept of tier segregation. As the introductory paragraphs explain, this mechanism furthered the generative phonologist goal of reducing the burden on the lexicon by removing from lexical specification the ordering relationships between consonants and vowels. This was seen as a positive development by most phonologists, since it allowed for a highly grammatical explanation of patterns not only in Semitic, where it was first invoked, but also in non-Semitic languages such as Yokuts.

As discussed, the benefits associated with tier segregation were not without their costs. For the non-Semitic cases in which tier segregation has been called upon, some reactions have been negative, proposing that tier segregation should be limited to cases in which it is fully justified solely on morphological grounds. From the standpoint of phonology, tier segregation has also been challenged as imposing too strict a limit on potential underlying representations, thus conflicting with some principles that underlie Optimality Theory.

[16] Interestingly, Newport *et al.* (2004) report similar studies carried out with non-human primates (cottontop tamarin monkeys) as the subjects. The results showed that these subjects demonstrated learning of the non-adjacent syllable and non-adjacent vowel patterns, but were unable to learn the non-adjacent consonant patterns that adult human subjects learned.

And, finally, the external evidence brought to bear on the issue is unable at this point to provide a final answer regarding the psychological reality of tier segregation. While some studies suggest that tier segregation is cognitively viable, others demonstrate that it can be done without. This is not a surprising situation; after all, only a limited number of such studies have been carried out at this point. It is clear that many more experiments could and should be conducted in order to further our understanding of the patterns that tier segregation was proposed to account for, and to help determine whether tier segregation should be maintained or discarded in phonological theory.

ACKNOWLEDGMENTS

Discussions with Dafna Graf, Lior Laks, Hadas Velan, and Andy Wedel have helped greatly with portions of this chapter. Many thanks are also due to Marc van Oostendorp, Beth Hume, and two anonymous reviewers. All errors remain the author's responsibility.

REFERENCES

Alcala, Pedro de. 1505. *Arte parta ligera mente saber la lengua araviga, emendada y añada y segunda mente imprimida.* Granatae: Juan Verala de Salamanca.

Al-Mozainy, Hamza Qublan. 1981. Vowel alternations in a Bedouin Hijazi Arabic dialect: Abstractness and stress. Ph.D. dissertation, University of Texas, Austin.

Archangeli, Diana. 1983. The root CV-template as a property of the affix: Evidence from Yawelmani. *Natural Language and Linguistic Theory* 1. 347–384.

Archangeli, Diana. 1984. Underspecification in Yawelmani phonology and morphology. Ph.D. dissertation, MIT. Published 1988, New York: Garland.

Archangeli, Diana. 1985. Yokuts harmony: Evidence for coplanar representation in nonlinear phonology. *Linguistic Inquiry* 16. 335–372.

Bat-El, Outi. 1988. Remarks on tier conflation. *Linguistic Inquiry* 19. 477–485.

Bat-El, Outi. 1994. Stem modification and cluster transfer in Modern Hebrew. *Natural Language and Linguistic Theory* 12. 571–596.

Bat-El, Outi. 2003. Semitic verb structure within a universal perspective. In Joseph Shimron (ed.) *Language processing and acquisition in languages of Semitic, root-based, morphology*, 29–59. Amsterdam & Philadelphia: John Benjamins.

Berent, Iris, Daniel L. Everett & Joseph Shimron. 2001. Do phonological representations specify variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology* 42. 1–60.

Berent, Iris, Gary F. Marcus, Joseph Shimron & Adamantios I. Gafos. 2002. The scope of linguistic generalizations: Evidence from Hebrew word formation. *Cognition* 83. 113–139.

Berent, Iris, Vered Vaknin & Joseph Shimron. 2004. Does a theory of language need a grammar? Evidence from Hebrew root structure. *Brain and Language* 90. 170–182.

Berent, Iris, Vered Vaknin & Gary Marcus. 2007. Roots, stems, and the universality of lexical representations: Evidence from Hebrew. *Cognition* 104. 254–286.

Bopp, Franz. 1824. *Vergleichende Zergliederung des Sanskrits und der mit ihm verwandten Sprachen. Erste Abhandlung. Von den Wurzeln und Pronominen erster und zweiter Person. Gelesen in der Königlich-preussichen Akademie der Wissenschaften.* Berlin: Druckerei der Königlichen Akademie der Wissenschaften.

Borg, Albert. 1975. Maltese morphophonemics. *Journal of Maltese Studies* 10. 11–28.

Boudelaa, Sami & William Marslen-Wilson. 2000. Non-concatenative morphemes in language processing: Evidence from Standard Arabic. In Anne Cutler, James M. McQueen & Rian Zondervan (eds.) *Proceedings of the Workshop on Spoken Word Access Processes*, 23–26. Nijmegen: Max Planck Institute for Psycholinguistics.

Boudelaa, Sami & William Marslen-Wilson. 2001. Morphological units in the Arabic lexicon. *Cognition* 81. 65–92.

Boudelaa, Sami & William Marslen-Wilson. 2004. Abstract morphemes and lexical representation: The CV-skeleton in Arabic. *Cognition* 92. 271–303.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6. 201–251.

Chomsky, Noam. 1951. Morphophonemics of modern Hebrew. M.A. thesis, University of Pennsylvania.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Churchward, C. Maxwell. 1940. *Rotuman grammar and dictionary, comprising Rotuman phonetics and a Rotuman–English dictionary*. Sydney: Australasian Medical Publishing Company.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Deutsch, Avital, Ram Frost & Kenneth I. Forster. 1998. Verbs and nouns are organized and accessed differently in the mental lexicon: Evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24. 1238–1255.

Ewald, Georg Heinrich August. 1827. *Kritische Grammatik der hebräische Sprache*. Leipzig: Hahn.

Frost, Ram, Avital Deutsch & Kenneth I. Forster. 2000. Decomposing morphologically complex words in a nonlinear morphology. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26. 751–765.

Frost, Ram, Kenneth I. Forster & Avital Deutsch. 1997. What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23. 829–856.

Gafos, Adamantios I. 1998. Eliminating long-distance consonantal spreading. *Natural Language and Linguistic Theory* 16. 223–278.

Gaon, Sa'adia. 1932. Kutub al-Lughah (S. L. Skoss, Fragments of unpublished philological works of Saadia Gaon). *Jewish Quarterly Review* (new series) 23. 329–336.

Gaon, Sa'adia. 1942. A study of inflection in Hebrew from Sa'adia Gaon's grammatical work Kutub al-Lughah. *Jewish Quarterly Review* (new series) 33. 171–212.

Gaon, Sa'adia. 1969. *Ha-Egron*. In Nehemia Allony (ed.) *Ha-egron kitab usul al-shir al-Ibrani* [*The book of the principles of Hebrew poetry*]. Jerusalem: Academy of the Hebrew Language.

Gesenius, Wilhelm. 1910. *Gesenius' Hebrew grammar*. Edited by E. Kautzsch; revised by A. E. Cowley. Oxford: Clarendon Press. Original edition: Halle, 1813.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT. Published 1979, New York: Garland.

Harris, Zellig S. 1941. Linguistic structure of Hebrew. *Journal of the American Oriental Society* 62. 143–167.

Hudson, Grover. 1986. Arabic root and pattern morphology without tiers. *Journal of Linguistics* 22. 85–122.

Hume, Elizabeth. 1991. Metathesis in Maltese: Implications for the Strong Morphemic Plane Hypothesis. *Papers from the Annual Meeting of the North East Linguistic Society* 21. 157–172.

Hume, Elizabeth. 1994. *Front vowels, coronal consonants and their interaction in nonlinear phonology*. New York: Garland.

Idrissi, Ali, Jean-François Prunet & Renée Béland. 2008. On the mental representation of Arabic roots. *Linguistic Inquiry* 39. 221–259.
Laks, Lior. 2006. The interface of prosodic morphology and arity operations: Lexical vs. syntactic operations in Arabic and Hebrew. M.A. thesis, Tel-Aviv University.
Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT. Published 1980, New York: Garland.
Marslen-Wilson, William & Xiaolin Zhou. 1999. Abstractness, allomorphy, and lexical architecture. *Language and Cognitive Processes* 14. 321–352.
McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.
McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12. 373–418.
McCarthy, John J. 1982. Prosodic templates, morphemic templates, and morphemic tiers. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part I, 191–223. Dordrecht: Foris.
McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17. 207–263.
McCarthy, John J. 1989. Linear order in phonological representation. *Linguistic Inquiry* 20. 71–99.
McCarthy, John J. 2000. The prosody of phase in Rotuman. *Natural Language and Linguistic Theory* 18. 147–197.
McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.
McCarthy, John J. & Alan Prince. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory* 8. 209–283.
McCarthy, John J. & Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University.
Morais, José, Paul Bertelson, Luz Cary & Jesus Alegria. 1986. Literacy training and speech segmentation. *Cognition* 24. 45–64.
Moscoso del Prado-Martín, Fermin, Avital Deutsch, Ram Frost, Robert Schreuder, Nivja de Jong & Harald Baayen. 2005. Changing places: A cross-language perspective on frequency and family size in Dutch and Hebrew. *Journal of Memory and Language* 53. 496–512.
Newport, Elissa & Robert Aslin. 2004. Learning at a distance I: Statistical learning of non-adjacent dependencies. *Cognitive Psychology* 48. 127–162.
Newport, Elissa, Marc Hauser, Geertrui Spaepen & Robert Aslin. 2004. Learning at a distance II: Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology* 49. 85–117.
Odden, David. 1987. Arguments against the vowel plane in Gta?. *Linguistic Inquiry* 18. 523–529.
Odden, David. 1994. Adjacency parameters in phonology. *Language* 70. 289–330.
Prince, Alan. 1987. On planes and copying. *Linguistic Inquiry* 18. 491–509.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.
Prunet, Jean-François, Renée Béland & Ali Idrissi. 2000. The mental representation of Semitic words. *Linguistic Inquiry* 31. 609–648.
Rose, Sharon. 2000. Rethinking geminates, long-distance geminates, and the OCP. *Linguistic Inquiry* 31. 85–122.
Sacy, Sylvestre Antoine Isaac de. 1810. *Grammaire arabe à l'usage des élèves de l'école spéciale des languages orientales vivantes*. Paris: Imprimerie Royale.
Steriade, Donca. 1986. Yokuts and the vowel plane. *Linguistic Inquiry* 17. 129–146.

Sumner, Megan. 2003. Testing the abstractness of phonological representations in Modern Hebrew weak verbs. Ph.D. dissertation, State University of New York, Stony Brook.

Suzuki, Keiichiro. 1998. A typological investigation of dissimilation. Ph.D. dissertation, University of Arizona.

Ussishkin, Adam. 1999. The inadequacy of the consonantal root: Modern Hebrew denominal verbs and output-output correspondence. *Phonology* 16. 401–442.

Ussishkin, Adam. 2000. The emergence of fixed prosody. Ph.D. dissertation, University of California, Santa Cruz.

Ussishkin, Adam. 2005. A fixed prosodic theory of nonconcatenative templatic morphology. *Natural Language and Linguistic Theory* 23. 169–218.

Ussishkin, Adam. 2006. Affix-favored contrast inequity and psycholinguistic grounding for nonconcatenative morphology. *Morphology* 16. 107–125.

Ussishkin, Adam & Alina Twist. 2009. Auditory and visual lexical decision in Maltese. In Bernard Comrie, Ray Fabri, Elizabeth Hume, Manwel Mifsud, Thomas Stolz & Martine Vanhove (eds.) *Introducing Maltese linguistics: Selected papers from the 1st International Conference on Maltese Linguistics*, 233–249. Amsterdam & Philadelphia: John Benjamins.

Volney, Constantin François de Chasseboeuf de. 1787. *Voyages en Syrie et en Egypte pendant les années 1783, 1784, 1785*. Paris: Volland, Desenne.

Younes, R. 1983. The representation of geminate consonants. Unpublished ms., University of Texas, Austin.

7 Feature Specification and Underspecification

Diana Archangeli

1 Introduction

The concept of minimal feature specification has been part of the phonological dialogue throughout the modern era, from the introduction of archiphonemes (Jakobson 1929; Martinet 1936; Trubetzkoy 1936) through various types of specification: Markedness (Kean 1975), Radical Underspecification (Archangeli 1984, 1988), Contrastive Underspecification (Steriade 1987; Mackenzie and Dresher 2004; Dresher 2009), and Combinatorial Specification (Archangeli and Pulleyblank 1994) to Richness of the Base (Prince and Smolensky 1993; Itô *et al.* 1995; Smolensky 1996). What is common to the various approaches is that features may be absent from at least some segments at any level of representation. Theories differ on when – or whether – the missing features are provided. For example, the Markedness rules of Chomsky and Halle (1968) provide values (the unmarked specifications) for all features on every segment before the application of any phonological rules, whereas under more recent perspectives, some segments exit the entire phonological-phonetic component without receiving specifications for every feature (Keating 1988; Pierrehumbert and Beckman 1988; Archangeli and Pulleyblank 1994).

A challenge in evaluating underspecification is to separate the essence of the arguments for minimal feature specification from the theoretical context in which the feature specification model is presented. In fact, the very term "underspecification" is testament to this challenge: the model was developed in the early 1980s in opposition to a model of full feature specification, and the term "underspecification" emphasized this changed role. Changing the focus to the motivation for specifications led to the term "Combinatorial Specification" (Archangeli and Pulleyblank 1994), while Optimality Theory's "Richness of the Base" makes no direct reference to feature specification at all (Prince and Smolensky 1993).

In retrospect, developing a theory of "feature specification" rather than of "underspecification" is more important. Which features are *necessarily present* regardless of the theory one adopts? Those which are not necessarily present can be unspecified, if the theory allows it. In a very basic way, then, the essence of feature specification hinges on the "necessary and sufficient" criterion: is a feature specification necessary to account for this pattern? Is that same specification sufficient to account for the pattern? Where the answer to both questions is "yes," the feature is critical to an analysis of that part of the language; where the answer to both is "no," whether the feature is specified depends on the theory in which the analysis is couched, not on the requirements of the data.

In this survey, we consider a wide array of phenomena, ranging from the morphological to the phonological to the phonetic, which demonstrate that at all levels, there are features that meet the "necessary and sufficient" criterion and there are others which do not. Those that do must be specified, regardless of the theory one adopts. Those that do not are features that may be left unspecified if one's theory allows for unspecified features.

In many of the examples, there is a secondary argument. If segments are fully specified, then the analysis includes a component simply to deal with those specified features – features which are not necessary to account for the basic pattern. The analysis then is more complex, and the necessary elements of the pattern are obscured. This provides a second set of arguments in favor of minimal specification: omitting unnecessary feature specifications reveals patterns; including them obscures patterns.

2 Morphological evidence

The first cases we consider are those in which a single feature differentiates one morphological class from another (see also CHAPTER 82: FEATURAL AFFIXES). That single feature is both necessary and sufficient. No other features are necessary, and so have no empirical role; they can be omitted from the representation. No fully specified segment is needed in these cases; in fact, positing such a representation obscures the patterns observed. Cases from Chaha (McCarthy 1983), Tiv (Pulleyblank 1986), and Yokuts (Archangeli 1991) make the argument.

2.1 *Single-feature morphemes*

In Chaha, a Western Gurage (Semitic) language of Ethiopia, the 3rd person singular object is marked by labialization of the rightmost labializable consonant (1) (McCarthy 1983).[1]

(1) *Chaha* (McCarthy 1983; CHAPTER 82: FEATURAL AFFIXES)

		no object	*with object*	
a.	*final*	dænæg	dænæg^w	'hit'
		nækæb	nækæb^w	'find'
b.	*medial*	mækær	mæk^{w}ær	'burn'
		s'æfær	s'æf^{w}ær	'cover'
c.	*initial*	qætær	q^{w}ætær	'kill'
		mæsær	m^{w}æsær	'seem'
d.	*none*	sædæd	sædæd	'chase'

[1] Zoll (1996) presents a similar case from Inor; Rose (1994) delves further into the similar palatalization pattern, also explored in McCarthy (1983) and CHAPTER 82: FEATURAL AFFIXES.

There are two aspects to the analysis. First, the sound component of the lexical representation need only include a labialization feature, [labial]. Should this affix be represented as a fully featured segment, rather than simply as the feature [labial], then it is critical (i) that [labial] be free to dissociate from that segment and associate to another segment and (ii) that all other features of the segment disappear completely regardless of whether [labial] successfully associates or not (1d). In short, all features other than [labial] are unnecessary; [labial] alone is sufficient.

Second, the segments of the roots need no [−labial] specification despite the fact that they are not round unless the object marker is added. Any [−labial] specifications complicate the analysis because some mechanism must be introduced to eliminate those specifications when the [labial] object marker is added.[2]

A similar example is found in Tiv, a Niger-Congo language of Nigeria, where the general past is indicated by downstep of a verb-initial high tone (2a); there is no overt marker when the verb is low-toned (2b).

(2) *Tiv general past* (Pulleyblank 1986)

	a. H-stem		b. L-stem	
1 syllable	ꜜH	ꜜvá 'came'	L	dzà 'went'
2 syllable	ꜜHL	ꜜúngwà 'heard'	LL	vèndè 'refused'
3 syllable	ꜜHLL	ꜜjévèsè 'fled'	LLL	ngòhòrò 'accepted'

These contrast with stems in other tenses, such as the recent past (3), where there is no downstep. Note the effect of vowel ablaut on the vowels /a/ and /u/, and the addition of a recent past high tone.

(3) *Tiv recent past* (Pulleyblank 1986)

	a. H-stem		b. L-stem	
1 syllable	H	vé 'came (recently)'	H	dzé 'went (recently)'
2 syllable	HH	óngó 'heard (recently)'	LH	vèndé 'refused (recently)'
3 syllable	HHL	jévésè 'fled (recently)'	LHL	ngòhórò 'accepted (recently)'

Treating the general past morpheme as a low tone directly results in these patterns since high tones regularly downstep after low tones in Tiv, while low tones are unaffected. Representing the general past affix as a low tone is necessary in order to account for the sound alternations; it is also sufficient. What is not necessary is to represent the general past marker with a fully specified vowel – crucially including the tone specification, but with no other critical features; all that would surface of such a vowel would be the tone.

[2] In some models, [labial] is a privative feature, and there is no "[−labial]" specification. In other models, [round] (or [±round]) might be used instead of [labial[. The name of the feature is irrelevant to the issues of feature specification.

The location of the H-tone recent past marker also supports a level of representation without specifications, as pointed out in Pulleyblank (1986). Were all vowels fully specified for either H-tone or L-tone, then it would be surprising for the recent past H-tone to dock on the second vowel of polysyllabic forms. However, if only the initial vowel is specified for a tone – which is the only location for contrastive tone in Tiv verbs – then the second vowel is the leftmost available vowel, and the location of the recent past H-tone is no surprise.

2.2 *Ghost segments*

A different type of case involves a "ghost segment," a segment that surfaces in some environments, but that is not realized in other environments.[3] For example, in French, many final consonants are realized only when there is a following vowel or there is a feminine suffix, *peti*[ ] but *peti*[t] *oiseau* and *la peti*[t]*e* (Tranel 1981; CHAPTER 112: FRENCH LIAISON). Archangeli (1984, 1991) and Zoll (1996) argue that such segments have feature representations, but crucially lack other structure: for Archangeli, the ghost segment is a floating feature, while for Zoll it is a feature or set of features without a root node. (While this chapter is focused on degrees of feature specification, we briefly visit the issue of prosodic specification in §3.5.) A particularly intriguing case of ghost segments is found in Yokuts, a Penutian language of Northern California. In Yokuts, there are several morphemes with ghost segments that may or may not surface. In one class of cases, there is good evidence that the ghost should be viewed as a single feature, rather than as a full-fledged segment. In this class of cases, a glottal stop surfaces in some environments (4a) but does not in other environments (4b), the typical ghost behavior. However, there is a third set of environments, where the ghost is realized as glottalization of a sonorant consonant, (4c): "the floating glottal stop of suffixes . . . may infect only the second consonant of a stem, if that consonant is a nasal, semivowel, lateral" (Newman 1944: 15). Thus, in this third class of cases, the Yokuts ghost [ʔ] is similar to the single-feature morphemes. However, in Yokuts, there are other segments in the morpheme (in this example, the sequence *-in'aj*): it is not simply a single feature morpheme. There are several Yokuts suffixes with the "floating glottal"; there are also suffixes with a floating /l/, /n/, /h/, or /m/. See Newman (1944), Archangeli (1984, 1991), and Zoll (1996) for examples with other suffixes.[4]

(4) *Yokuts contemporaneous gerundial* -ʔin'aj (Archangeli 1984)

	root	*affixed form*		
a.	root with 2 Cs gets glottal stop			
	dub	dubʔun'aj	*dubun'aj, *dub'un'aj	'while leading by the hand'
	doos	dosʔin'aj	*dosin'aj, *dos'in'aj	'while reporting'

[3] These are called "latent segments" in Zoll (1996).

[4] Other effects can be found in (4), both phonological (rounding harmony giving *-un'aj* after roots with /u/) and templatic. The root column shows the syllabic pattern taken by the root with about half of the suffixes in the language. The other half of the suffixes determine the prosodic structure to the root, and *-in'aj* is among these: in each case, the root has the minimal prosody possible given its segments. See Prince (1990) and Archangeli (1991).

b. root with 3 Cs gets nothing

lihim	lihmin'aj	*lihmʔin'aj, *lih'min'aj	'while running'

c. root with sonorant C2 gets glottalization, whether root has 2 Cs or 3

panaa	pan'in'aj	*panin'aj, *panʔin'aj	'while arriving'
c'oow	c'ow'in'aj	*c'owin'aj, *c'owʔin'aj	'while grasping'
taan	tan'in'aj	*tanin'aj, *tanʔin'aj	'while going'
hiwiit	hiw'tin'aj	*hiwtin'aj, *hiwtʔin'aj	'while walking'

The feature that is necessary to account for this pattern is a glottal constriction feature, e.g. [+constricted glottis]. Interestingly, this feature is sufficient to account for the sonorant behavior in (4c). This feature is also necessary and sufficient to account for the realization of a glottal stop, all other features being predictable once it is known that the segment has [+constricted glottis].

2.3 *Summary*

These morphological examples demonstrate two points. First, in some cases, a single feature is necessary and sufficient to explain the phonological behavior of some morpheme: [+constricted glottis] is sufficient to explain the three-way Yokuts patterns while [labial] accounts for the distribution of the object marker in Chaha. Second, while additional features might be included in the representation of such morphemes, any such additional features serve no necessary phonological purpose; the only reason to include such features is to satisfy the demands of the theory, not the demands of the data. The data themselves make no such requirement.

In the next section, we examine phonological cases which make the same point, with the difference being that the motivation for active *vs.* inert features is phonological, not morphological.

3 Phonological evidence

There are a variety of types of phonological patterns that support the absence of at least some feature specifications. The cases are similar in that there are asymmetric behaviors among the segments of a language, and that asymmetry is consistent with the absence or presence of feature specification. Variety arises when the effects of this asymmetry are considered: morphemes with invariable feature specifications *vs.* those with variable specifications, revealed *vs.* obscured phonological patterns, inert segments *vs.* involved segments, and segments that behave ambiguously.

3.1 *Variable and invariable feature specification*

The first class of evidence comes from the completely phonological distribution of features. In these cases, a contrast is found in some part of the lexicon that does not occur in some other part of the lexicon. In the Margi example, roots and suffixes differ by whether they bear invariable tone or not (Williams 1976; Pulleyblank 1986, 2006). We focus on roots here; the same argument can be made based on suffix behavior.

Some Margi roots are always high-toned (5a), some always low-toned (5b), and some have a low–high contour (5c) (data from Pulleyblank 2006).

(5) *Margi roots*

a. *High-toned roots*

tá	'cook'	tábá	'cook all'
ŋgúlí	'roar'	ŋgúlíbá	'surpass in roaring'

b. *Low-toned roots*

ptsà	'roast'	ptsàbá	'roast thoroughly'
tsàvè	'pierce'	tsàvèbá	'pierce thoroughly'

c. *Low–high contoured roots*

fǐ	'swell'	fìbá	'make swell'
və̌l	'fly, jump'	və̀lbá	'make jump over, across'

These forms reveal a three-way contrast among roots: those which are invariably high-toned, those which are invariably low-toned, and those which surface with a contour in isolation, but simplify that contour when suffixed. However, there is a fourth class of roots, the alternating roots. In this class, the tone of the root varies depending on the tone of the suffix that is attached. The root is high-toned with a high-toned suffix, and otherwise low-toned.

(6) *Margi alternating roots* (Hoffman 1963; Evans 2008)

a. *With high-toned suffixes*

ɗə̀l	'buy'	ɗə́lbá	'buy'
ɲàl	'bite'	ɲálbá	'bite a hole'

b. *With low-toned suffixes*

pʧì	'wash'	pʧìnà	'wash'
tlǎ	'cut'	tlànà	'cut off'

Pulleyblank (1986, 2006) argues that in the latter case, the roots are toneless in lexical representation and receive their surface tones through the phonological component. Note that requiring all roots to be fully specified, and so to bear lexical tone, would require either a completely arbitrary third tone (which never surfaces) or non-phonological diacritics on the variable-tone class to distinguish the roots with low tones in isolation that are always low-toned (5b) from those that are low-toned in isolation, but are high-toned when they have high-toned suffixes (6a).

A similar case is found in Tiv, and has already been illustrated in (2). The only location in a word where tones vary unpredictably is the first syllable. All other tones are predictable. Consequently, the two patterns in (2) are the extent for verb roots in Tiv. Requiring full specification of tones leads to the expectation of far more patterns than are attested in the language.

There are numerous cases of completely predictable features, with behavior comparable to that of Margi tone. For example, the distribution of [ATR] in Pulaar (Paradis 1986; Archangeli and Pulleyblank 1994) and Kinande (Archangeli and Pulleyblank 1994, 2002) follows this pattern, as does the distribution of all vowel features in Tiv (Archangeli and Pulleyblank 1994).

3.2 *Patterns revealed*

The second class of cases involves phonological patterns which are seen to be quite simple patterns if only the necessary and sufficient features are present in representations. These same patterns appear to be rather complex patterns if all features are present on all segments. Two brief examples from Yokuts are presented. In each case, the essence of the phenomenon is revealed when only the necessary and sufficient features are present. One case involves vowel harmony, a feature phenomenon, and the other involves epenthesis, a prosodic phenomenon.

3.2.1 *Yokuts vowel harmony*

The Yokuts vowel harmony system is relatively simple. The language has a four-vowel system, /i a o u/, and vowels of like height agree for roundness. See Newman (1944), Archangeli (1984, 1985), and Steriade (1986) for more on Yokuts harmony. In (7a), the vowels in the non-initial syllables arise either from epenthesis or from the suffix *hin* (aorist).[5]

(7) *Yokuts vowel harmony*

a. [u] *after* /u/, *else* [i] *in* -hin *(aorist)*

	ʔugunhun	*ʔuginhin	'drank'
but	lihimhin	*lihumhun	'ran'
	baṭ'inhin	*baṭ'unhun	'fell down'
	hoginhin	*hogunhun	'floated'

b. [o] *after* /o/, *else* [a] *in* -(h)atin *(desiderative)*

	doshotinhin	*doshatinhin	'was trying to tell'
but	bintatinxokʾ	*bintotinxokʾ	'be trying to ask!'
	hudhatinxoʔ	*hudhotinxoʔ	'wants to know about'
	ṭʾawhatinxoohin	*ṭʾawhotinxoohin	'was trying to win from'

It is possible to represent the vowel system as the result of the logically possible combinations of the features [round] and [non-high]: /o/ is specified as [round, non-high] while /i/ has neither feature; /a/ is [non-high] and /u/ is simply [round]. Under this view, the harmony effect is achieved by a manipulation of the feature [round] where [non-high] values are identical.

Were features fully specified, the harmony process would need to change both [round] and [back] features on /i/, for /i/ to surface as [u]. Additionally, it would need to change [round] and [low] for /a/ to surface as [o]. Thus, with full specification, the harmony process must mention not only [round], but also [back] and [low], a far more complex process than one that involves only [round]. The complexity is brought on by the fully specified segments, not by properties inherent to the process itself. The essential properties of the alternation are revealed by minimally specified representations.

[5] The suffix *(h)atin* (desiderative) requires the minimal prosody for the root segments; see footnote 4 and §3.5. It also has a "floating /h/" which surfaces only with biconsonantal roots. See §2.2 for discussion of this phenomenon.

3.2.2 *Yokuts epenthesis*

A further phenomenon in Yokuts is that of vowel epenthesis; the inserted vowel surfaces as [i], or [u] in harmonic environments. Vowel epenthesis occurs when consonants would otherwise be unsyllabifiable: the motivation is purely prosodic. See Kisseberth (1970) and Archangeli (1984, 1991) for more on Yokuts syllabification.[6]

(8) *Yokuts epenthesis* (Archangeli 1984, 1991)

(epenthetic vowels are underlined; forms without epenthesis have "__" showing the absence of a high vowel in that environment)

	root	-hin (aorist)	-t (passive aorist)	
a.	*no epenthesis with either affix*			
	hojoo	hojoohin	hojot	'name'
b.	*epenthesis with* -t *(passive aorist)*			
	caw	cawhin	cawit	'shout'
	bok'	bok'hin	bok'it	'find'
	ṣil'	ṣil'hin	ṣil'it	'shout'
	dub	dubhun	dubut	'lead by the hand'
	biniit	binethin	bineetit	'ask'
c.	*epenthesis in different places with* -hin *(aorist) and* -t *(passive aorist)*			
	diijl	deejil__hin	dej__lit	'guard'
	bilṣ	biliṣ__hin	bil__ṣit	'finish'
	ʔamc	ʔamic__hin	ʔam__cit	'be near'
	diijl	deejil__hin	dej__lit	'guard'
	ʔogl	ʔogil__hin	ʔob__lit	'emerge'
	luk'l	luk'ul__hun	luk'__lut	'bury'

With a fully unspecified /i/, the formal account of epenthesis is solely prosodic. No features are necessary: the resultant vowel is equivalent to the lexical representation of /i/, fully unspecified.

These examples are representative of the way in which minimal feature specification reveals phonological patterns. In both cases, requiring full specification obscures the essence of each process, in the harmony case making [round], [low], and [back] formally equivalent while only [round] is critical and in the epenthesis case making the prosodic phenomenon appear to critically involve specific features.

3.3 *Inert segments*

Inert segments are segments that do not participate in some alternation, despite surface appearances suggesting that they would be involved. These segments function in contrast to their apparent specifications.

[6] Forms show the effect of other, interacting phenomena. Items with /u/ exhibit round harmony, for example *dubhun* and *dubut*, **dubhin*, **dubit*; see discussion of (7). Long vowels typically surface as non-high (*deejilhin*, **diijilhin*), and closed syllables typically have short vowels (*dejlit*, not **deejlit*).

3.5 *Prosodic specification*

Prosody includes at least syllable and foot structure.[8] In contrast to features, the standard approach is to assume that input representations do not include syllables or feet, and sometimes moras: these are provided by the phonology. For run-of-the-mill cases, the primary argument is that it isn't possible to tell what the syllabification or footing of a particular sequence is until it is placed in context: whether a morpheme-final consonant is an onset or a coda depends on whether it is followed by a vowel (*dual vs. duality*, etc.); stress placement also depends on the root and any affixes (*'atom vs. a'tomic*, etc.). Similarly, whether a high sonorant is a vowel or a consonant depends on context which may be provided through morphological concatenation, for instance in Berber (Guerssel 1986). In such cases, lexically specified syllable or foot structure is neither necessary nor sufficient.

Morphological patterns sometimes exploit underspecified prosodic structure, particularly reduplication (Marantz 1982; McCarthy and Prince 1986, 1993b, 1995b; CHAPTER 100: REDUPLICATION) and templatic morphology (Archangeli 1984, 1991; McCarthy 1985; CHAPTER 108: SEMITIC TEMPLATES). The hallmark of prosodic morphology is that (a significant part of) the prosody is determined by one morpheme, distinct from that which provides the segments. Examples of templatic morphology appear in (4) and (7b); see also footnote 4. (13) includes some key examples from (4). The first column in (13) shows the segments and the prosodic pattern taken by the verb root with about half the suffixes in the language; the second column shows suffixation with a couple of these "normal" suffixes. (Note that the [ee] in *hiweeten* is due to the lowering of long high vowels to mid, a well-documented process in the language. See references cited.) Comparison of forms in the column headed "supplied CVC" shows what happens to those roots when the following suffix imposes a prosody on the root, supplanting its default prosodic structure.

(13) *Yokuts templatic morphology* (Archangeli 1984, 1991)

	root prosody	*affixation*	*supplied* CVC ...	*non-occurring*
a.	panaa	panaahin	pan'in'aj	*pan'aan'aj
	'arrive'	'arrived'	'while arriving'	
b.	hiwiit	hiweeten	hiw'tin'aj	*hiw'eetin'aj
	'walk'	'will walk'	'while walking'	

If the suffix *(?)in'aj* did not provide a prosodic structure for the verb root, the forms in the column headed "non-occurring" would be expected. Both roots typically begin with an iambic pattern preceding the suffix; with *(?)in'aj*, the root takes as little prosodic space as possible. See references cited for further exemplification.

There are, in addition, cases of "a-templatic" morphology, where the morphology provides no template (Bat-El 1989, 1994; McCarthy 1993). Putting this together with minimal feature specification suggests instances where there is a morpheme with neither prosodic nor segmental structure. Gafos (1998) argues for just such

[8] This section is included at the recommendation of an anonymous reviewer who pointed out the relevance of prosodic cases for generalizing the notion of necessary and sufficient specification.

an analysis of reduplication in Temiar, noting that the "invariable" aspects of reduplication can be explained without reference to imposed prosody or features.

While these brief paragraphs can do little justice to the diversity involved with prosodic specification, the intent of including this discussion is more modest: to demonstrate that the concept of considering whether a representation is both necessary and sufficient is not limited to features, but extends to the prosodic aspects of phonological representations.

3.6 *Summary*

In this section, we have reviewed a number of examples from phonological patterning. In each case, specifying only the necessary and sufficient features provides a more revealing explanation of the phonological patterns than does full specification of the relevant segments.

Phonological accounts must be able to handle cases of the type presented in the next section, where phonological features are irrelevant for the phonetic realization of a form – a possibility that is expected given minimally specified representations.

4 Phonetic evidence

Phonological representations with minimal feature specifications result in segments that are unspecified for unnecessary features. The minimization of feature specifications leaves open the possibility that these specifications are provided phonetically. Of interest are cases where the phonetic feature specification has an impact that is different from the phonological specification of features, for example the striking difference between the phonological (categorical) specification *vs.* the phonetic (gradient) realization such as a cline, a feature whose impact gradually increases or decreases phonetically over a span of one or more segments.

The cases of most interest are where the cline correlates to some phonological feature, such as tongue body height, tone, lip rounding, etc. In the relevant cases, some feature is phonologically active for some segments (either because it is distinctive, because it plays a role in a phonological pattern, or both), while for other segments there is a cline effect, with the greatest articulation of the feature found on the segment(s) for which it is distinctive, otherwise gradually building to and/or decreasing from that point.

4.1 *The cline effect in vowel harmony*

In Akan (a Niger-Congo language), [ATR] harmony has a phonological effect on all non-low vowels, rendering them categorically either advanced or retracted. [ATR], then, is necessary on non-low vowels (Clements 1981; Archangeli and Pulleyblank 1994). By contrast, low vowels are typically retracted. Since this is a general state for low vowels, [RTR] (or [–ATR] – the difference is irrelevant to the point being made) is not necessary. Furthermore, specifying low vowels for a tongue root feature would be contradictory, because to the *left* of an advanced vowel, a series of normally retracted vowels (whether low or not) exhibits a gradual increase and, in particular, no phonological or phonetic tone specification specifically associated with moras between the initial mora with its low boundary tone and the pitch-accented mora with its high tone. As in other cases, while it is possible to specify phonological tone on these moras, the choice of tone is arbitrary for it never surfaces: the surface pitch contour derives from the interpolation from the initial low tone to the high pitch accent tone.

4.3 *Co-articulation* vs. *allophonic variation*

A final example of this type is the pattern of tongue body height in Russian VCV sequences (Keating 1988). As Keating points out, in Russian most consonants have both a palatalized and a plain counterpart. An exception is the velar fricative /x/. The backness or frontness of /x/ is dependent on the context. Because there are no lexical requirements on the backness dimension for /x/, it might be considered unspecified for [front/back], in contrast to other consonants.

Keating (1988) compares formant patterns in [x] in three environments: /ixa/, /axi/, and /axa/. The spectrograms clearly demonstrate a continuous transition throughout the articulation of the consonant from a high F2 to a low F2 in /ixa/, and no change in F2 in the fricative in /axa/. However, in /axi/, the fricative shows a sharp rise to a steady high F2, consistent with the consonant being [front]. Keating's conclusion (1988: 285) is that "[t]he /x/ before /i/ acquires a value for [back] by spreading from /i/, whereas other /x/'s remain unspecified for [back], and therefore show interpolatory effects."

This pattern arises only when the intervening consonant evinces no phonological or allophonic palatalization. When the /i/ follows the consonant, the contour of the formant is consistent with Russian palatalization that takes place between [i] and any preceding consonant. Otherwise, the contour of the tongue's trajectory is consistent with an interpolated gesture between the palatal and the non-palatal vowels, consistent with the intervening consonant having no specification for the interpolating tongue body features.

4.4 *Summary*

The three examples presented in this section converge on a set of properties consistent with minimal feature specification. The domain in which the interpolation pattern is found is segments that have no necessary phonological specification for a feature. The boundaries of each pattern are defined by specified features. And finally, the effect seen is consistent with simply interpolating the specified values from one boundary to the next. In each case, more phonological specifications would get in the way of the phonetic pattern, and would obscure any characterization of the process.

Evidence of this sort is relevant to the discussion of Optimality Theory in that the phonological constraints need to be constructed in a way that allows for this class of phonetic effects. This might be done by integrating phonetic constraints into the constraint hierarchy along with the phonological constraints; it might be done by developing a constraint hierarchy to force unspecified segments in the (phonological) output which then feeds into a phonetic component. In either case, it is critical that segments are without specifications for certain phonological features.

5.3 *Future directions*

Very recent developments in Evolutionary Phonology (Blevins 2004) and, in particular, Emergent Phonology (Mielke 2008; Mohanan *et al.* 2009) and "substance-free phonology" (Hale and Reiss 2000a, 2000b; Morén 2007; Blaho 2008) have implications for how our understanding of feature specification might develop in the future. Emergent Phonology intends to account for the growing body of evidence that the features and patterns found in language sound systems are strongly influenced by the physical system, i.e. the human's articulatory and perceptual systems (Archangeli and Pulleyblank 1994; Flemming 1995, 2002, 2005; Boersma 1998; de Boer 2000; Hume and Johnson 2001; Blevins 2004; Hayes *et al.* 2004). In order for the formal model to capture this concrete influence, Emergent Phonology proposes that there is no universal, innate set of features. Consequently, in an optimality approach, there can also be no innate constraints referring to features. Rather, the necessary entities are generated on an as-needed basis as each language is learned.

As laid out in Mohanan *et al.* (2009), if there is no specific inventory of innate distinctive features, then the necessary features emerge from the data encountered by the learner. With no inventory of innate features to map to, in the absence of evidence for a particular feature there will be no feature proposed. Minimal specification results. Lack of specification for irrelevant features is expected. There is no need to wonder about what happens with features that are not phonologically active in a language, since there is no motivation to learn such features. This same argument extends to subsets of the sounds of a language for which a feature is inactive: without evidence, the feature will not be learned for that subset, producing segments unspecified for a feature that is active elsewhere in the language.

A further implication under Emergence is due to the heavy responsibility placed on the learning process, a responsibility also exploited in Evolutionary Phonology (Blevins 2004), where mislearning leads to language sound change. Under Emergence, there is the potential also for overlearning, i.e. learning more than is necessary and sufficient in some cases. When overlearning creates a conflict with the ambient grammar, language change is predicted, leading to item-by-item language change until a critical mass is reached and the pattern is generalized.

6 Conclusion

The cases examined here present a variety of arguments in favor of minimal feature specification. Morphological arguments include accounting for single-feature morphemes and morphemes with floating features. Among the phonological arguments are accounting for inert segments, for ambiguous segments, and for the contrast between variable and invariable surface forms, as well as the demonstration that the essence of patterns is revealed when fewer feature specifications are present. Finally, the interpolation of phonologically active feature effects across spans of segments for which that feature is not necessary provides phonetic evidence for the absence of feature specification. Throughout, evidence is consistent with all and only the necessary and sufficient features being present in representations.

Clements, G. N. 1981. Akan vowel harmony: A non-linear analysis. *Harvard Studies in Phonology* 2. 108–177.

Davis, Stuart. 1990. An argument for radical underspecification in English. *Linguistic Inquiry* 21. 301–306.

de Lacy, Paul. 2002. The formal expression of markedness. Ph.D. dissertation, University of Massachusetts, Amherst.

Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology.* Cambridge: Cambridge University Press.

Evans, Jonathan. 2008. "African" tone in the sinosphere. *Language and Linguistics* 9. 463–490.

Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles.

Flemming, Edward. 2002. *Auditory representations in phonology.* London & New York: Routledge.

Flemming, Edward. 2005. Deriving natural classes in phonology. *Lingua* 115. 287–309.

Frisch, Stefan A. 1996. Similarity and frequency in phonology. Ph.D. dissertation, Northwestern University.

Gafos, Adamantios I. 1998. A-templatic reduplication. *Linguistic Inquiry* 29. 515–527.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory.* Cambridge, MA & Oxford: Blackwell.

Guerssel, Mohamed. 1986. Glides in Berber and syllabicity. *Linguistic Inquiry* 17. 1–12.

Hale, Mark & Charles Reiss. 2000a. Phonology as cognition. In Noel Burton-Roberts, Philip Carr & Gerard J. Docherty (eds.) *Phonological knowledge: Conceptual and empirical issues*, 161–184. Oxford: Oxford University Press.

Hale, Mark & Charles Reiss. 2000b. "Substance abuse" and "dysfunctionalism": Current trends in phonology. *Linguistic Inquiry* 31. 157–169.

Haraguchi, Shōsuke. 1977. The tone pattern of Japanese: An autosegmental theory of tonology. Tokyo: Kaitakusha.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42. 25–70.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology.* Cambridge: Cambridge University Press.

Hoffman, Carl. 1963. *A grammar of the Margi language*. London: Oxford University Press.

Hume, Elizabeth & Keith Johnson. 2001. A model of the interplay of speech perception and phonology. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 3–26. San Diego: Academic Press.

Inkelas, Sharon. 1995. The consequences of optimization for underspecification. *Papers from the Annual Meeting of the North East Linguistic Society* 25. 287–302.

Itô, Junko. 1984. Melodic dissimilation in Ainu. *Linguistic Inquiry* 15. 505–513.

Itô, Junko & Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17. 49–73.

Itô, Junko, Armin Mester & Jaye Padgett. 1995. Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26. 571–613.

Jakobson, Roman. 1929. Remarques sur l'évolution phonologique du russe comparée avec celle des autres langues slaves. *Travaux du Cercle Linguistique de Prague* 2. Reprinted (1962) in *Selected writings*, vol. 1: *Phonological studies*, 7–116. The Hague: Mouton.

Kaplan, Lawrence. 1981. *Phonological issues in North Alaskan Inupiaq.* Fairbanks: Alaska Native Language Center.

Kean, Mary-Louise. 1975. The theory of markedness in generative grammar. Ph.D. dissertation, MIT.

Keating, Patricia. 1988. Underspecification in phonetics. *Phonology* 5. 275–292.

Kisseberth, Charles W. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1. 291–306.

Lombardi, Linda. 2002. Coronal epenthesis and markedness. *Phonology* 19. 219–251.

Mackenzie, Sara & B. Elan Dresher. 2004. Contrast and phonological activity in the Nez Perce vowel system. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 29. 283–294.

Marantz, Alec. 1982. Re reduplication. *Linguistic Inquiry* 13. 435–482.

Martinet, André. 1936. Neutralisation et archiphonème. *Travaux du Cercle Linguistique de Prague* 6. 46–57.

McCarthy, John J. 1983. Consonantal morphology in the Chaha verb. *Proceedings of the West Coast Conference on Formal Linguistics* 2. 176–188.

McCarthy, John J. 1985. *Formal problems in Semitic morphology and phonology*. New York: Garland.

McCarthy, John J. 1993. Template form in prosodic morphology. *Proceedings of the Formal Linguistics Society of Midamerica* 3. 187–218.

McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University. Revised version available at http://ruccs.rutgers.edu/pub/papers/pm86all.pdf.

McCarthy, John J. & Alan Prince. 1993a. Generalized alignment. *Yearbook of Morphology 1993*. 79–153.

McCarthy, John J. & Alan Prince. 1993b. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst.

McCarthy, John J. & Alan Prince. 1993c. Prosodic morphology I: Constraint interaction and satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University.

McCarthy, John J. & Alan Prince. 1995a. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst, MA: GLSA.

McCarthy, John J. & Alan Prince. 1995b. Prosodic morphology. In Goldsmith (1995), 318–366.

McCarthy, John J. & Alison Taub. 1992. Review of Paradis & Prunet (1991). *Phonology* 9. 363–370.

Mester, Armin & Junko Itô. 1989. Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language* 65. 258–293.

Mielke, Jeff. 2005. Ambivalence and ambiguity in laterals and nasals. *Phonology* 22. 169–203.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Mohanan, K. P., Diana Archangeli & Douglas Pulleyblank. 2009. The emergence of Optimality Theory. In Linda Uyechi & Lian-Hee Wee (eds.) *Reality exploration and discovery: Pattern interaction in language and life*, 143–158. Stanford: CSLI.

Morén, Bruce. 2007. The division of labor between segment-internal structure and violable constraints. In Sylvia Blaho, Patrik Bye & Martin Krämer (eds.) *Freedom of analysis?*, 313–344. Berlin & New York: Mouton de Gruyter.

Newman, Stanley. 1944. *Yokuts language of California*. New York: Viking Fund.

Paradis, Carole. 1986. Phonologie et morphologie lexicales: Les classes nominales en Pulaar (Fula). Ph.D. dissertation, University of Montreal.

Paradis, Carole & Jean-François Prunet. 1989. On coronal transparency. *Phonology* 6. 317–348.

Paradis, Carole & Jean-François Prunet (eds.) 1991. *The special status of coronals: Internal and external evidence*. San Diego: Academic Press.

Pierrehumbert, Janet B. & Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.

Prince, Alan. 1990. Quantitative consequences of rhythmic organization. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26(2). 355–398.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

106 Exceptionality

MATTHEW WOLF

1 Introduction

The work of phonologists is, in part, to discover empirical generalizations about the sound shape of utterances in each of the world's languages, and to formulate theories which predict these generalizations. The "in part" comes from the mentalist stance of generative linguistics: its ultimate goal is not merely a theory which accounts for generalizations in surface data, but a theory that accounts for these data in the same way that native speakers' internalized grammars do. In a sense, then, generative analyses are not really theories of (corpus-internal) linguistic data per se; they are theories of native speakers' tacit theories about those data – the grammars which speakers arrive at during language acquisition. In Chomsky's (1965) terms, generative linguistics aims at theories that possess *explanatory adequacy* – the ability to account for how language acquisition occurs (and thus also for the knowledge resulting from acquisition) – and not merely *observational* adequacy – the ability to correctly predict the grammatical or ungrammatical status, for the language in question, of any given form.

This mentalist stance creates two complications for how phonologists deal with generalizations in linguistic data. First, a language's lexicon might contain generalizations which speakers have not incorporated into their mental grammars, perhaps because Universal Grammar does not give them the resources to do so (see Becker *et al.* 2008 and Becker 2009 for a proposed instance). Second, speakers might have mentally internalized a generalization, but in a way that seems to be sub-optimal according to the standards of theory comparison, for instance Occam's Razor, which phonologists are used to calling upon. The Māori passive (Hale 1973; McCarthy 1981; de Lacy 2004) is a well-known case in which the seemingly most simple account of the data is arguably not the way in which speakers have internalized those data.

An inverse situation obtains in the cases where a language displays an apparent generalization to which there are exceptions. Say, for instance, we observe in some language that all stems begin with a consonant, except for two stems [eda] and [opu]. Strictly speaking, the statement "stems must begin with a consonant" is not a true generalization about this language. Nevertheless, it is conceivable that speakers have mentally internalized this restriction, and have somehow marked [eda] and [opu] as exempt from the restriction. That is, we can imagine that speakers have internalized a theory which might be considered sub-optimal for the reason that it contains stipulations. Two questions therefore face the linguist. The first is to establish whether the imperfect generalization has in fact been internalized – we might check, for example, whether speakers productively apply a process of consonant epenthesis or vowel deletion to vowel-initial loanwords, or to nonce forms elicited in a wug test (Berko 1958; see also CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY). The second question, assuming that a restriction against vowel-initial words has been internalized, is *how* the exceptional status of [eda] and [opu] is diacritically indicated in speakers' grammar and/or lexicon. One possibility is that they bear an *abstract* diactritic feature like [−vowel deletion], stating that a certain rule should not apply to these forms. The other possibility is to make diacritic use of *phonological* features. Say for instance that our hypothetical language lacks the segment [x] on the surface. Then we might assume that underlying forms of the exceptional items are /xeda/ and /xopu/, and that the rule that deletes /x/ is ordered after whatever rule is responsible for ensuring that stems are consonant-initial. Either way, we are adding to an item's lexical representation something that causes it to behave phonologically in a special way. No matter how analyzed, then, exceptionality ultimately reflects the arbitrary nature of sound/meaning pairings, and the manner in which the underlying form (or some other stored representation) holds information about those aspects of an item's pronunciation which are not predictable from general principles (an idea traceable to work on exceptionality in structuralist morphophonemics by Swadesh and Voegelin 1939).

Despite this commonality, the two approaches to exceptionality can be implemented in different ways, which make different predictions about exactly what kinds of exceptional patterns languages can have. §2 of this chapter discusses approaches to exceptionality that have been proposed in rule-based generative phonology, and §3 considers approaches in Optimality Theory (Prince and Smolensky 1993). As we shall see, the abstract-diacritic and phonological-diacritic approaches have both appeared in a number of different forms in both rule-based and constraint-based phonology. §4 discusses predictions about exceptionality in Lexical Phonology and Stratal OT. §5 offers a concluding summary of some of the issues about how exceptionality works which we will encounter in examining the strengths and weaknesses of various proposals.

2 Exceptionality in rule-based phonology

2.1 *The theory of diacritic features in* SPE

Lexical exceptionality is treated in two sections of Chomsky and Halle (1968; hereafter *SPE*): §4.2.2, "Treatment of exceptions," and §8.7, "Diacritic features." The latter lays out three kinds of diacritic features with distinct uses. One kind are features which define conjugation classes, native *vs.* foreign lexical strata, and similar kinds of word classes. A second is what Coats (1970) dubs "alphabet features"; these are, as Zonneveld (1978: 132) notes, "rather vague[ly]" defined, and are called on in *SPE* to perform several tasks, which are reviewed in Zonneveld (1978: §3.2.3). As we shall see, and as Zonneveld (1978) argues, these features can be used to handle the various types of exceptionality which were argued in the late 1960s

An important question which can be raised in regard to minor rules is whether major *vs.* minor status is a strictly binary matter, or is instead a continuum. Most work on minor rules seems to assume binarity (Levy and Fidelholtz 1971 invoke a feature [±major]). An equally binary approach to minor-rule effects would be to assume that minor rules are not rules at all, and that the morphemes which "undergo" a minor rule simply have multiple listed allomorphs. This is argued for by Hudson (1974), and more recently, in an optimality-theoretical context, by Mascaró (2007) and Kager (2009). Kager's example is that of open syllable lengthening in Dutch. Normally, both short (3a) and long (3b) vowels stay the same length, even as the syllable they are in alternates between closed and open, as can be seen by comparing singular and plural forms of nouns:

(3)	a.	*kl*[ɑ]*s*	~	*kl*[ɑ].*sen*	'class(es)'
		k[ɪ]*p*	~	*k*[ɪ].*pen*	'chicken(s)'
	b.	*b*[aː]*s*	~	*b*[aː].*zen*	'boss(es)'
		p[oː]*t*	~	*p*[oː].*ten*	'paw(s)'

However, certain exceptional nouns show an alternation between a short vowel in the bare root and a long vowel in the plural:

(4)	*gl*[ɑ]*s*	~	*gl*[aː].*zen*	'glass(es)'
	sl[ɔ]*t*	~	*sl*[oː].*ten*	'lock(s)'

This appears to indicate a minor rule of open syllable lengthening. Kager's (2009) analysis is that the alternating roots have two listed allomorphs, one with the short vowel and one with the long vowel, with the appropriate form being selected in the appropriate environment.

This assumes that the minority alternation pattern in Dutch actually *is* unproductive and not really the default. In inflectional morphology, at least, minority patterns arguably can be used as the default (Marcus *et al.* 1995), so in cases of apparent "minor rules" like this we might need to run a wug test or the like to see whether the minority or the majority pattern were the one which speakers extended to nonce forms. If the minority pattern truly is not the default, then either a minor-rules or a listed-allomorphy treatment would reasonably lead us to expect that the minor alternation should be *perfectly* unproductive – it might seem odd for speakers to set up multiple listed allomorphs for a nonce word if they don't yet have evidence that it alternates. Is this expectation borne out?

Coetzee (2009) investigated this using an artificial language learning paradigm. If trained on novel-language data, in which some stems alternate and some do not, and then tested on novel stems, subjects tended to assume by default that the novel stems were non-alternating. However, this was not a categorical preference, and, moreover, the degree to which novel stems were assumed to be non-alternating decreased if the ratio of alternating to non-alternating stems in the learning data was increased.

Several responses can be given to such results. One is to pursue a model like that proposed in Zuraw (2000), discussed in §3.4 below, which combines fixed, listed pronunciations for existing forms with a stochastic grammar that will reproduce, for novel forms, statistical trends about items in the existing lexicon. This would be, effectively, a hybridization of a minor-rule approach and an allomorphy one. Alternatively, we could use minor rules (or indexed OT constraints), and assume that nonce words are assigned the diacritic features of a randomly chosen existing word (Becker 2009). Here again, existing forms would have a fixed behavior, and a nonce form's likelihood of taking on a certain mode of behavior would be in part determined by the proportion of existing forms which had that behavior.

Another option would be to assume that the difference between major and minor rules, between default and exception, is not one of kind but only of degree. Within a generative paradigm, this kind of move is exemplified by Jackendoff's (1975) theory of the lexicon. In this theory, the lexicon consists of fully specified entries for all words (even morphologically derived ones). Morphological rules, including those like English umlaut (2) which form irregulars, serve not to actually *produce* existing forms, but instead function as statements of speakers' knowledge of which properties of certain existing forms are predictable from those of other existing forms. These *lexical redundancy rules* are, however, available for creatively generating novel forms.

A more radical move in the same conceptual direction would be to adopt an exemplar-based theory (e.g. Bybee 2001; Pierrehumbert 2001), in which words are mentally represented as clusters of concrete, phonetically detailed tokens which have been heard and recorded in memory. Diachronic change is assumed to result from updating the exemplar clouds associated with certain words or phonological categories. A central argument for exemplar approaches is the observation (Fidelholtz 1975; Phillips 1984) that leniting changes tend to appear first in more frequent words (see CHAPTER 66: LENITION). In an exemplar model, this follows from the very fact of frequent words being encountered more often, which results in their exemplar clouds being updated more often. With the novel pattern entrenched in part of the lexicon, its diffusion over time to new forms (Wang 1969) is entirely expected. The case of apparent "minor rules" – patterns restricted to a few items – being extended to new words might be seen in such a model as simply a micro-scale manifestation of this process of lexically diffusing diachronic change.

For a comparison of exemplar models with classical generative feed-forward models in their predictions about historical change, including lexical diffusion and the role of frequency, and discussion of possible problems for a purely exemplar-based approach, see Bermúdez-Otero (2007a) (see also CHAPTER 93: SOUND CHANGE).

2.2.2 *Rule-environment features*

Suppose that rule *n* is:

(5) A → B / X __ Y

Under the convention proposed in *SPE* (1968: 173, 374), the [±rule *n*] value of the focus *A* is relevant to whether the rule applies or not, but the [±rule *n*] values of the matrices *X* and *Y* making up the context do not matter. Chomsky and Halle note (1968: 375) that this limitation could be relaxed, such that, for rule *n* to apply, all of the matrices making up the environment (*X* and *Y*), as well as the focus *A*, would have to be [+rule *n*]. While they report having "no clear cases in a real language" requiring this move, they do point out one prediction of restricting reference to rule features to the focus of the rule: that rules of epenthesis (CHAPTER 67:

VOWEL EPENTHESIS) cannot have lexical exceptions. In an epenthesis rule, the focus is zero, which can have no features (diacritic or otherwise) because it is, literally, nothing. Chomsky and Halle therefore suggest a small-scale tweak to their original convention, whereby a rule *n* (A → B / X __ Y) applies to a string *X'A'Y'* except when (i) A' is [−rule *n*] or (ii) the formative containing A' is [−rule n]. (Karen Jesney (personal communication) points out that, if the focus of a rule must be [+rule *n*] for the rule to apply, then by exactly the same argument, rules of epenthesis could never apply at all unless the [±rule *n*] features of the environment are allowed to count.)

Subsequently, Coats (1970) and Kisseberth (1970) proposed not only that rule features had to be relevant in the environment, but also that there had to be distinct rule features for foci and for environments. This was also advocated by Coats (1974) and Coats and Lightner (1975). Kisseberth's (1970) proposal is based on data from Yine (then referred to as Piro), an Arawakan language spoken in Peru; the data are from Matteson (1965).

Yine has a rule that deletes a preconsonantal, morpheme-final vowel, unless it is preceded by a consonant cluster (CHAPTER 68: DELETION):

(6) V → Ø / VC __ + CV

This is exemplified by alternations like the following:

(7)	*bare root*		*nominalization in* /-lu/	
	[jimaka]	'teach'	[jimaklu]	'teaching'
	[kakonu]	'build a hunting shelter'	[kakonru]	'hunting shelter'
	[kama]	'make'	[kamlu]	'handicraft'

Rule (6) has two kinds of exceptions. First, there are affixes of the appropriate phonological shape before which vowel drop does not take place:

(8)	[meji-ta]	please-VERBAL THEME	'to please'
	[meji-wa-ta]	please-INTRANS-VERBAL THEME	'to celebrate'
	[heta-wa]	see-DURATIVE	'still see'
	[hata-ta]	illuminate-VERBAL THEME	'to illuminate'
	[heta-nu]	see-ANTICIPATORY	'going to see'

As can be seen, a morpheme-final vowel preceding one of these suffixes will not delete. In addition, the intransitive suffix /wa/ not only fails to trigger vowel-drop, but also fails to undergo it. In this regard, intransitive /wa/ differs from durative /wa/, verbal theme /ta/, and anticipatory /nu/. These latter three suffixes fail to trigger vowel drop, but they do *undergo* vowel drop:

(9)	[meji-w-lu]	please-DURATIVE-NOMINALIZER	'celebration'
	*[meji-wa-lu]		
	[jona-t-nawa]	paint-VERBAL THEME-REFL	'to paint oneself'
	*[jona-ta-nawa]		
	[heta-n-ru]	see-ANTICIPATORY-NOMINALIZER	'going to see him'
	*[heta-nu-ru]		

the "detached exception" problem. Zonneveld (1978: 204) notes, however, that one of the rules of Russian discussed by Coats and Lightner (1975) in support of rule-environment features instantiates just this phenomenon. The rule in question is:

(12) V → Ø / __ jV [after certain bases; MW]

This rule affects suffixes such as comparative /-eje/ when they are attached to certain bases, but not when they are attached to other bases. The applicability of 12 to the suffix is thus being decided by another morpheme, even though /-eje/ by itself contains the rule's entire structural description (Coats and Lightner 1975: note 3; Zonneveld 1978: 193).

Detached exceptions may also exist in the genitive plurals of nouns in Czech and Serbo-Croatian (Halle and Nevins 2009). Slavic nouns have the structure root + theme vowel + case-number. In Czech, while vowels normally delete in the environment / __ + V, in certain declensions the theme vowel fails to delete before the yer of the genitive plural ending. Simply making the genitive plural ending [−environment vowel deletion] will not do, because it is only following roots of certain declension classes that the genitive plural ending fails to trigger deletion of the theme vowel. In terms of Zonneveld's (1978) proposal for Yine, we might set up for Czech a re-adjustment rule which marks the theme vowel [−vowel deletion] in the environment of the genitive plural ending and of roots of the appropriate declension classes (cf. Halle and Nevins 2009: 367 (26)). The ban on superheavy syllables in the perfect conjugation of Vedic Sanskrit (Cooper, forthcoming) may also represent an example of a detached exception.

2.3 *Representational approaches to exceptions in rule-based phonology*

In addition to abstract diactric features like [−rule *n*], *SPE* also calls upon diacritic use of phonological features to analyze exceptions to several rules (1968: 233–234). For example, in *righteous* [ɹaɪ.ʧəs], the first vowel is unexpectedly tense, rather than lax. This is attributed to the underlying form of the stem *right* being /ɹixt/, and there being a rule of Pre-/x/-tensing which is ordered after Laxing. Such analyses provoked debate over how abstract underlying forms could be. One of the earliest entries in this debate on the side of concreteness came from Kiparsky (1968, 1973), who proposed to ban absolute neutralization via the Alternation Condition (CHAPTER 80: MERGERS AND NEUTRALIZATION).

(13) Neutralization rules cannot apply to all occurrences of a morpheme.

Adopting (13) would close off the *SPE* analyses based on underlying /x/ in English, since English lacks [x] on the surface, and so a rule of /x/-deletion would have to apply to every instance of a morpheme with underlying /x/. Kiparsky (1968, 1973) accordingly proposes that all cases of exceptionality be handled via abstract rather than phonological diacritics.

In response to (13), a number of counterexamples were proposed, perhaps most notably by Brame (1972) for Maltese, and Selkirk and Vergnaud (1973) for French (see Kiparsky 1973 for references to additional examples). In this section I will focus on Selkirk and Vergnaud's example of French *h-aspiré*, as it is simpler than they don't meet the rule's structural description (they do not have a C immediately after the word boundary). Selkirk and Vergnaud note (1973: note 5) that one might entertain a theory in which diacritic features could endow morphemes with the ability to trigger a rule even without meeting the rule's structural description. They rebut this possibility by relating the following "old Hallean tale":

(15) In Armenia, there is a special kind of nightingale which sings when the moon shines.
– And when the moon does not shine?
– It sings anyway.

That is, if morphemes can be listed as triggering a rule without meeting its structural description, then the rule is stripped of content as a phonological generalization. This objection is the exact inverse of Kisseberth's (1970) and Coats' (1970) "detached exception" argument against the alphabet-feature alternative to [−environment rule *n*]. In both cases, the analysis being argued against is one in which the exceptional triggering or blocking of a process is attributed to a morpheme which has no phonological relevance to the process.

This idea, that exception features can allow a morpheme to trigger some rule without meeting its structural description, was advocated in contemporaneous works by Harms (1968; cf. Karttunen 1970: 150; Zonneveld 1978: 166–167) and Kenstowicz (1970). Drawing inspiration from the alphabet-feature analysis outlined by Zonneveld (1978) for Kenstowicz's (1970) Lithuanian example, we could rewrite the French VN coalescence rule as follows:

(16) $\text{VN} \rightarrow \text{V}_{[+\text{nas}]}$ / __ # {C, [+*h-aspiré*]}

That is, we can make words that fail to meet the phonological structural description of the unaltered rule by virtue of lacking an initial C nevertheless trigger the rule by using curly brackets to make that phonological element of the structural description disjunctive with a diactritic feature. While this does work, formally speaking, the availability of such an analysis does not diminish the force of the Armenian nightingale objection: it allows us to stipulate that phonological rules apply to arbitrary lists of morphemes which might not have anything in common phonologically, putting us on the slippery slope to phonology becoming merely a list of morpheme-specific instructions in which no generalizations are expressed.

A different diacritic analysis of *h-aspiré* was advocated by Kiparsky (1973). He responded to Selkirk and Vergnaud's (1973) absolute-neutralization analysis by noting the tendency of *h-aspiré* to be lost under morphological derivation (CHAPTER 88: DERIVED ENVIRONMENT EFFECTS): *le Hitler* [lɔ.it.lɛʁ] 'the Hitler' but *l'hitlérisme* [lit.lɛ.ʁizm] 'the hitlerism'. Disappearance of an initial consonant only in suffixed words would be decidedly odd and certainly does not otherwise exist in French. Morphological idiosyncrasies, however, as Kiparsky (1973) notes, are well known to be able to disappear under derivation. In English, for instance, the root *stand* in isolation takes the irregular past tense *stood*, but the compound verb *grandstand* takes the regular past tense: *grandstanded*, **grandstood*. (However, see Burzio 2000 and Yu 2000 for frameworks in which marked structures can be lost exclusively in affixed words.)

which devoice in coda position, and morphemes containing stops of the first class as being exceptions to devoicing, they propose to mark the distinction between the three kinds of stops in underlying representations alone. Non-alternating voiced stops would be underlyingly [+voice], and non-alternating voiceless stops underlyingly [−voice]. The alternating stops would be underlyingly [0voice], with [−voice] being filled in on them in coda position and [+voice] in onset position (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION).

A problem for this approach is pointed out by Kager (2009), and had previously been noted by Kiparsky (1993a), who argues for an underspecification account of derived environment effects. This is that underspecification cannot work for exceptional alternations, where the alternating phonological property is represented privatively. For example, in the case of the Dutch "minor rule" of open syllable lengthening, an underspecification account would need to assign distinct underlying representations to the three kinds of vowels: non-alternating long vowels, non-alternating short vowels, and alternating vowels. The first would presumably be bimoraic and the latter monomoraic. But what representation are we then to assign to the alternating vowels? Privative representations allow only a two-way contrast, but when contrasts are neutralized only in an exceptional class of morphemes, the underspecification approach needs a three-way underlying contrast. (In the case of the Dutch example, this objection will not go through if the apparent length contrast is really a tense/lax contrast, as van Oostendorp 2000 argues.) One final issue with the prespecification approach, as with all representational approaches to exceptions, is how to account for the loss of exceptionality under derivation.

For additional critiques of the input-underspecification approach, see Becker (2009: §2.6), which deals specifically with Turkish, and Albright (2002: ch. 5), which discusses a similar pattern in Lakhota.

3.4 *Exceptionality via listing*

Besides the underspecification model, there are two other approaches in OT which locate exceptionality in underlying forms. The first is the one discussed earlier in which "minor rules" always involve selection of lexically listed allomorphs.

The other is proposed by Zuraw (2000). Her concern is with the following paradox. Often in a language a phonological process will either fixedly apply or not apply to any given existing word. At the same time, if we examine the lexicon as a whole, there will be identifiable statistical trends regarding where the rule applies and where it does not. These statistical trends often can be shown to be psychologically real to native speakers; for instance they affect speakers' willingness to apply the process in nonce words. We thus need a model of grammar which incorporates both statistical trends as well as fixed pronunciations for existing words.

Zuraw (2000) focuses on the example of Tagalog nasal substitution. Prefix-final nasals in Tagalog sometimes simply assimilate to the place of articulation of a following root-initial consonant (CHAPTER 81: LOCAL ASSIMILATION); other times, the nasal and following consonant coalesce into a nasal segment homorganic with the underlying root-initial consonant. Nasal substitution is lexically irregular, in that some polymorphemic words undergo it and some don't. Indeed, simply marking morphemes as to whether they trigger or undergo nasal substitution is were found of rules which were undeniably post-lexical (because they applied between words) but which showed properties of lexical rules. One example (Kaisse 1990): in English the Rhythm Rule (Liberman and Prince 1977) retracts word-level primary stresses to avoid clash with a primary stress in a following word:

(25) *ˌthir'teen ~ 'thirˌteen 'men* (**ˌthir'teen 'men*)

There are several words, for instance *discrete,* which fail to undergo this retraction, implying that they are exceptions to the Rhythm Rule:

(26) *dis'crete ~ dis'crete 'math* (**ˌdiscrete 'math*)

Another standard view in rule-based Lexical Phonology is that only post-lexical rules can introduce non-contrastive features (Mohanan 1982; Kiparsky 1985). Here too, examples have been proposed of rules that would qualify as post-lexical on this basis, but which nevertheless have exceptions – see for instance Chomsky (1962), Withgott (1982), and Steriade (2000) on North American English flapping (though cf. Davis 2005; Bermúdez-Otero, forthcoming: §2.4.5.2). The many documented cases of word-specific sub-contrastive phonetics which have helped to motivate exemplar-based theories of phonology (see e.g. Pierrehumbert 2001 and references therein) are also relevant to this question.

Stratal OT (Kiparsky 2000; Bermúdez-Otero, forthcoming, among many others) is a version of Lexical Phonology in which the phonology of each stratum is an OT grammar rather than a rule-based grammar. In this theory it is predicted that any process which misapplies cyclically must be able to have lexical exceptions; see Bermúdez-Otero (2007b, 2008, as well as references therein), who calls this "Chung's Generalization" in recognition of Chung (1983: 63). The idea is this: for a rule to misapply cyclically, it must apply in a surface-true manner at stratum *n,* and the result of applying it at stratum *n* must be protected by faithfulness at stratum *n*+1 (which may be identical to stratum *n,* in the case of cyclicity at the stem level). This entails that stratum *n*+1 will also be faithful to exceptional inputs which are supplied to it directly by the rich base rather than by an earlier stratum. For some possible counterexamples to this prediction, see Benua (1997: ch. 3) and Albright (2008). Benua's example is from Sundanese, where nasal harmony operates cyclically, even though vowel nasality is not contrastive anywhere in the language, including in affixes that are added after an earlier cycle's application of nasal harmony.

5 Conclusion

The study of exceptions is all about the nature of linguistic generalizations, and how and whether speakers can internalize generalizations that are imperfect. As such, exceptionality is an issue of central importance to any theory of language.

There are two ways in which an item can be arbitrarily marked to behave exceptionally. The first is to assign an abstract diacritic which indicates that parts of the grammar (rules, constraints) or whole grammars (co-phonologies) are or are not applicable to that item. The second is to give the item a special underlying

Anttila, Arto. 2002. Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20. 1–42.

Atkins, Beryl T., Alain Duval, Rosemary C. Milne, Pierre-Henri Cousin, Hélène M. A. Lewis, Lorna A. Sinclair, Renée O. Birks & Marie-Noëlle Lamy. 1998. *Collins Robert unabridged French–English English–French Dictionary/Le Robert & Collins senior dictionnaire français–anglais anglais–français.* 5th edn. Glasgow & Paris: Collins & Dictionnaires Le Robert.

Bailey, Charles-James N. 1973. Variation resulting from different rule orderings in English phonology. In Charles-James N. Bailey & Roger Shuy (eds.) *New ways of analyzing variation in English*, 211–252. Washington: Georgetown University Press.

Becker, Michael. 2009. Phonological trends in the lexicon: The role of constraints. Ph.D. dissertation, University of Massachusetts, Amherst.

Becker, Michael, Nihan Ketrez & Andrew Nevins. 2008. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish devoicing neutralization. Unpublished ms., Reed College, Yale University & Harvard University (ROA-1001).

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Ph.D. dissertation, University of Massachusetts, Amherst (ROA-259).

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.

Bermúdez-Otero, Ricardo. 2007a. Diachronic phonology. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 497–517. Cambridge: Cambridge University Press.

Bermúdez-Otero, Ricardo. 2007b. On the nature of the cycle. Paper presented at the 15th Manchester Phonology Meeting. Handout available (June 2010) at www.bermudez-otero.com/15mfm.pdf.

Bermúdez-Otero, Ricardo. 2008. [ðə swɪˈŋɒmɪtə ˈtɜːnd əˈgenst sə ˈmɪŋɪs ˈkæmbəl]: Evidence for Chung's generalization. Paper presented at the 16th Manchester Phonology Meeting. Handout available (June 2010) at www.bermudez-otero.com/16mfm.pdf.

Bermúdez-Otero, Ricardo. Forthcoming. *Stratal Optimality Theory*. Oxford: Oxford University Press.

Boersma, Paul. 2007. Some listener-oriented accounts of *h*-aspiré in French. *Lingua* 117. 1989–2054.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45–86.

Brame, Michael K. 1972. On the abstractness of phonology: Maltese ʕ. In Michael K. Brame (ed.) *Contributions to generative phonology*, 22–61. Austin: University of Texas Press.

Burzio, Luigi. 2000. Cycles, non-derived-environment-blocking, and correspondence. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: Phonology, syntax, and acquisition*, 47–87. Oxford: Oxford University Press.

Bybee, Joan. 2001. *Phonology and language use.* Cambridge: Cambridge University Press.

Carlson, Barry F. 1969. Unmarked order and lexical exceptions. *Hawaii Working Papers in Linguistics* 1. 205–212.

Chomsky, Noam. 1962. A transformational approach to syntax. In Archibald A. Hill (ed.) *Proceedings of the 3rd Texas Conference on Problems of Linguistic Analysis in English*, 124–158. Austin: University of Texas Press.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Chung, Sandra. 1983. Transderivational relationships in Chamorro phonology. *Language* 59. 35–66.

Coats, Herbert S. 1970. Rule environment features in phonology. *Papers in Linguistics* 2. 110–140.

Coats, Herbert S. 1974. On the alternation *j/v* in Russian. In Demetrius J. Koubourlis (ed.) *Topics in Slavic phonology*, 29–42. Cambridge, MA: Slavica.

Coats, Herbert S. & Theodore M. Lightner. 1975. Transitive softening in Russian conjugation. *Language* 51. 338–341.

Coetzee, Andries W. 2009. Learning lexical indexation. *Phonology* 26. 109–145.
Cooper, Adam. Forthcoming. Challenging locality in constraint indexation: Epenthesis in the Vedic Perfect. *Papers from the Annual Meeting of the North East Linguistic Society* 40.
Cressey, William W. 1972. Irregular verbs in Spanish. In Jean Casagrande & Bohdan Saciuk (eds.) *Generative studies in Romance languages*, 236–246. Rowley, MA: Newbury House.
Davis, Stuart. 2005. Capitalistic v. militaristic: The paradigm uniformity effect reconsidered. In Laura J. Downing, T. A. Hall & Renate Raffelsiefen (eds.) *Paradigms in phonological theory*, 107–121. Oxford: Oxford University Press.
de Lacy, Paul. 2004. Maximal words and the Māori passive. In John J. McCarthy (ed.) *Optimality Theory in phonology: A reader*, 499–512. Oxford: Blackwell.
Dell, François. 1973. Two cases of exceptional rule ordering. In F. Kiefer & N. Ruwet (eds.) *Generative grammar in Europe*, 141–153. Dordrecht: Reidel.
Fidelholtz, James L. 1975. Word frequency and vowel reduction in English. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 11. 200–213.
Finley, Sara. 2010. Exceptions in vowel harmony are local. *Lingua* 120. 1549–1566.
Fukazawa, Haruka. 1999. Theoretical implications of OCP effects on features in Optimality Theory. Ph.D. dissertation, University of Maryland at College Park (ROA-307).
Hale, Kenneth. 1973. Deep-surface canonical disparities in relation to analysis and change: An Australian example. In Thomas A. Sebeok (ed.) *Current trends in linguistics*, vol. 11: *Diachronic, areal, and typological linguistics*, 401–458. The Hague: Mouton.
Halle, Morris & Andrew Nevins. 2009. Rule application in phonology. In Eric Raimy & Charles Cairns (eds.) *Contemporary views on architecture and representations in phonology*, 355–382. Cambridge, MA: MTT Press.
Harms, Robert T. 1968. *Introduction to phonological theory*. Englewood Cliffs, NJ: Prentice-Hall.
Hayes, Bruce & Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104.
Hudson, Grover. 1974. The representation of non-productive alternation. In John M. Anderson & Charles Jones (eds.) *Historical linguistics*, vol. 2, 203–229. Amsterdam: North-Holland.
Inkelas, Sharon, Cemil Orhan Orgun & Cheryl Zoll. 1997. The implications of lexical exceptions for the nature of grammar. In Iggy Roca (ed.) *Derivations and constraints in phonology*, 393–418. Oxford: Clarendon Press.
Itô, Junko & Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.) *The handbook of Japanese linguistics*, 62–100. Malden, MA & Oxford: Blackwell.
Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51. 639–671.
Kager, René. 2009. Lexical irregularity and the typology of contrast. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Studies in honor of Paul Kiparsky*, 397–432. Cambridge, MA: MIT Press.
Kaisse, Ellen M. 1990. Toward a typology of postlexical rules. In Sharon Inkelas & Draga Zec (eds.) *The phonology–syntax connection*, 127–143. Chicago: University of Chicago Press.
Karttunen, Frances. 1970. Problems in Finnish phonology. Ph.D. dissertation, Indiana University, Bloomington.
Kenstowicz, Michael. 1970. Lithuanian third person future. In Jerrold M. Sadock & Anthony L. Vanek (eds.) *Studies presented to Robert B. Lees by his students*, 95–108. Edmonton: Linguistic Research.
Kiparsky, Paul. 1968. How abstract is phonology? Distributed by Indiana University Linguistics Club. Reprinted in Paul Kiparsky (1982). *Explanation in phonology*, 119–163. Dordrecht: Foris.
Kiparsky, Paul. 1973. Phonological representations: Abstractness, opacity, and global rules. In Osamu Fujimura (ed.) *Three dimensions of linguistic theory*, 57–86. Tokyo: Taikusha.

Kiparsky, Paul. 1982. From Cyclic Phonology to Lexical Phonology. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*. Part I, 131–175. Dordrecht: Foris.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Kiparsky, Paul. 1993a. Blocking in nonderived environments. In Sharon Hargus & Ellen M. Kaisse (eds.) *Studies in Lexical Phonology*, 277–313. San Diego: Academic Press.

Kiparsky, Paul. 1993b. Variable rules. Handout of paper presented at Rutgers Optimality Workshop 1, Rutgers University.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–365.

Kisseberth, Charles W. 1970. The treatment of exceptions. *Papers in Linguistics* 2. 44–58.

Klausenburger, Jürgen. 1977. A non-rule of French: h-aspiré. *Linguistics* 192. 45–52.

Kraska-Szlenk, Iwona. 1997. Exceptions in phonological theory. In Bernard Caron (ed.) *Proceedings of the 16th International Congress of Linguists*. CD-ROM: Paper No. 0173. Oxford: Pergamon.

Kraska-Szlenk, Iwona. 1999. Syllable structure constraints in exceptions. In John R. Rennison & Klaus Kühnhammer (eds.) *Phonologica 1996: Syllables!?*, 113–131. The Hague: Thesus.

Lakoff, George. 1970. *Irregularity in syntax*. New York: Holt, Rinehart & Winston.

Levy, Mary & James L. Fidelholtz. 1971. Arabic broken plurals, rule features, and lexical features. *Glossa* 5. 57–70.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.

Lightner, Theodore M. 1968. On the use of minor rules in Russian phonology. *Journal of Linguistics* 4. 69–72.

Lightner, Theodore M. 1972. *Problems in the theory of phonology: Russian phonology and Turkish phonology*. Edmonton: Linguistic Research Inc.

Mahanta, Shakuntala. 2007. Directionality and locality in vowel harmony. Ph.D. dissertation, University of Utrecht.

Marcus, Gary F., Ursula Brinkmann, Harald Clashen, Richard Wiese & Steven Pinker. 1995. German inflection: The exception that proves the rule. *Cognitive Psychology* 29. 189–256.

Mascaró, Joan. 2007. External allomorphy and lexical representation. *Linguistic Inquiry* 38. 715–735.

Matteson, Esther. 1965. *The Piro (Arawakan) language*. Berkeley: University of California Press.

McCarthy, John J. 1981. The role of the evaluation metric in the acquisition of phonology. In C. L. Baker & John J. McCarthy (eds.) *The logical problem of language acquisition*, 218–248. Cambridge, MA: MIT Press.

Mohanan, K. P. 1982. Lexical Phonology. Ph.D. dissertation, MIT. Distributed by Indiana University Linguistics Club.

Oostendorp, Marc van. 2000. *Phonological projection: A theory of feature content and prosodic structure*. Berlin & New York: Mouton de Gruyter.

Pater, Joe. 2009. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In Steve Parker (ed.) *Phonological argumentation: Essays on evidence and motivation*, 123–154. London: Equinox.

Pesetsky, David. 1979. Russian morphology and lexical theory. Unpublished ms., MIT.

Phillips, Betty S. 1984. Word frequency and the actuation of sound change. *Language* 60. 320–342.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam & Philadelphia: John Benjamins.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Reynolds, William. 1994. Variation and phonological theory. Ph.D. dissertation, University of Pennsylvania.

Robinson, Orrin W. 1977. Rule reordering and lexical diffusion. In William S.-Y. Wang (ed.) *The lexicon in phonological change*, 69–85. The Hague: Mouton.

Selkirk, Elisabeth. 1972. The phrase phonology of English and French. Ph.D. dissertation, MIT.

Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Selkirk, Elisabeth & Jean-Roger Vergnaud. 1973. How abstract is French phonology? *Foundations of Language* 10. 249–254.

Steriade, Donca. 2000. Paradigm uniformity and the phonetics–phonology boundary. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, 313–334. Cambridge: Cambridge University Press.

Swadesh, Morris & C. F. Voegelin. 1939. A problem in phonological alternation. *Language* 15. 1–10.

Tessier, Anne-Michelle. Forthcoming. UseListedError: A grammatical account of lexical exceptions in phonological acquisition. *Papers from the Annual Meeting of the North East Linguistic Society* 39 (ROA-1015).

Tranel, Bernard. 1981. *Concreteness in generative phonology: Evidence from French.* Berkeley: University of California Press.

Wang, William S.-Y. 1969. Competing changes as a cause of residue. *Language* 45. 9–25.

Withgott, Mary M. 1982. Segmental evidence for phonological constituents. Ph.D. dissertation, University of Texas, Austin.

Yu, Alan C. L. 2000. Stress assignment in Tohono O'odham. *Phonology* 17. 117–135.

Zonneveld, Wim. 1978. *A formal theory of exceptions in generative phonology.* Dordrecht: Foris.

Zuraw, Kie. 2000. Patterned exceptions in phonology. Ph.D. dissertation, University of California, Los Angeles (ROA-788).

Volume I: General Issues and Segmental Phonology

Volume II: Suprasegmental and Prosodic Phonology

Volume III: Phonological Processes

Volume IV: Phonological Interfaces

Volume V: Phonology across Languages

The Blackwell Companion to Phonology is a major reference work drawing together 124 new contributions from leading scholars in the field. Led by a team of renowned international scholars, the *Companion* represents a diverse range of approaches and methodologies to key phenomena in phonological research. In contrast to other handbooks and reference works currently available for phonology, these volumes focus on phenomena and case studies that highlight past and ongoing debates in the field. Interdisciplinary connections, such as those with the social and computational sciences, are covered, as well as statistical and experimental techniques. The *Companion* will be a touchstone for future phonological theorists, giving an overview of data and insights which any theory of phonology should be able to cover.

Simultaneously offering broad coverage and a high level of detail, *The Blackwell Companion to Phonology* is a landmark work available in both print and online editions that will be indispensable to students and researchers in the field for years to come.

www.companiontophonology.com

Cover image: Theo van Doesburg, *Color scheme for Café Aubette, Strasbourg, France* 1928. New York, Museum of Modern Art (MoMA). Ink and gouache on paper, 27.3 x 62.9 cm. Gift of Lily Auchincloss, Celeste Bartos, and Marshall Cogan 390.1982 © 2010. Digital image, The Museum of Modern Art, New York / Scala, Florence.

Cover design by Nicki Averill

WILEY-BLACKWELL
www.wiley.com/wiley-blackwell

VOLUME V

Phonology across Languages

Edited by van Oostendorp, Ewen, Hume, and Rice

THE BLACKWELL COMPANION TO Phonology

WILEY-BLACKWELL

VOLUME V

Phonology across Languages

Edited by

Marc van Oostendorp

Colin J. Ewen

Elizabeth Hume

Keren Rice

THE BLACKWELL COMPANION TO

Phonology

WILEY-BLACKWELL

Brief Contents

Volume I

Volume II

Volume III

Volume IV

Volume V

Full Table of Contents

Volume I

Volume II

Suprasegmental and Prosodic Phonology

Volume III

Phonological Processes

Volume IV

Phonological Interfaces

Volume V

Phonology across Languages

107 Chinese Tone Sandhi

Bao Zhiming

1 Introduction

It has long been recognized that tone in Chinese (and other mainland Asian tone languages) is different from tone in African languages. Two differences are noteworthy (see also CHAPTER 45: THE REPRESENTATION OF TONE). First, while African tone languages have small inventories of tones which can be characterized in terms of pitch levels of H (high), M (mid), and L (low), Chinese has a larger number of contrastive tones, a fact recognized in the numbers 1 (low) through 5 (high) widely used in Chinese linguistics to transcribe tones. Second, Chinese tonal inventories contain more tones with complex contours than their African counterparts. African tone has played a major role in the development of generative phonology, with Asian tone languages contributing empirical data to support or refine theoretical claims based on African tonal phenomena. Many of the tone sandhi processes that are common in African languages, such as spreading, are also attested in Chinese. Naturally, tone sandhis which involve a contour are either unique to or more common in Chinese and other Asian languages with large tonal inventories. The complexity and range of tone sandhi in Chinese have contributed to our understanding of tonal phenomena in general. The difference in the size of tonal inventory between Chinese and African languages has received scant attention from phonologists, and its theoretical significance is not fully understood.

Chinese tones are typically described in terms of the four Middle Chinese (ca. 900 CE) categories: *ping* 'even', *shang* 'rising', *qu* 'departing', and *ru* 'entering'. Each of the four categories is realized in two registers, *yin* and *yang*, giving rise to a total of eight tones. These are illustrated in (1) with the tonal inventory of Songjiang, a Wu dialect spoken in what is now metropolitan Shanghai (Bao 1990, 1999a):

(1)	I	*ping*	a.	53	t'i	'ladder'
					ti	'low'
			b.	31	di	'lift'
	II	*shang*	a.	44	t'i	'body'
					ti	'bottom'
			b.	22	di	'brother'

III	*qu*	a.	35	t'i	'tear'
				ti	'emperor'
		b.	13	di	'field'
IV	*ru*	a.	$\underline{5}$	p'aʔ	'tap'
				paʔ	'hundred'
		b.	$\underline{3}$	baʔ	'white'

Throughout the chapter, Roman numerals refer to the tonal categories, and underlining marks the *ru* tones, which are realized on syllables ending in voiceless or glottal stops. As the Songjiang inventory shows, the *yin* tones (a) occur with voiceless onsets and are higher in pitch, and the *yang* tones (b) occur with voiced onsets and are lower in pitch.

2 Common tone sandhis in Chinese

When tones occur in a compound or phrase, they may undergo change. I use the term *citation tone* to refer to tones that are lexically specified for each morpheme, which for our purpose coincides with a single character and a single syllable. I use *sandhi tone* to refer to tones derived from citation tones in specific context. In the Chinese linguistics literature, a tone is commonly described in terms of two dimensions: pitch height (or register) and pitch movement (or contour) over the duration of the syllable, which is the presumed tone-bearing unit in Chinese. We can approach tone sandhi from the perspective of an individual tone or from the perspective of an entire tonal inventory. The term *tone sandhi* is used in these two senses. Where the difference in perspective is important, I will use the term *tone sandhi system* or simply *system* to refer to the sandhi processes that affect the entire tonal inventory of a dialect.

Tone sandhi patterns vary enormously across Chinese dialects, and Pike's (1948) contour-based Africanist–Asianist characterization is, as we shall see, simplistic. We can recognize four broad types of tone sandhi in Chinese. The first type includes sandhis which target register, contour, or both (the whole tone) and are sensitive to neighboring tones. I call this type *contextual tone sandhi*. The second type refers to tone sandhis which target tones in a particular position, typically the left or right edge of a polysyllabic string. In this type, tonal context does not play a role. I call this type *positional tone sandhi*. The third type is what I call *templatic tone sandhi*, where the sandhi patterns cannot be derived from the lexically specified citation tones without referring to some prespecified tonal template. The fourth type is *tone spread*, which is commonly attested in African tone languages and familiar in the generative phonological literature. To the extent possible, I will cite data from the original sources, with the original tonal and segmental transcriptions. Often tones from the original data are given different featural interpretations in the theoretically oriented secondary literature. This is especially true for contour tones with a slight curvature, such as 21, 45, or 214. By using the tonal transcriptions from the sources, I hope not only to give credit to the field linguists, but also to highlight the interpretive possibilities of numerically transcribed tones.

2.1 *Contextual tone sandhi*

The tonal phonology of Tianjin, 150 km to the east of Beijing, offers us a good case study of sandhi processes targeting the register and contour separately. The Tianjin sandhi system, which involves all citation tones in all disyllabic combinations, has been studied intensely in generative phonology; see M. Chen (1986, 2000) for a summary. Tianjin has four citation tones, 21, 45, 213, and 53. Compared with the dialect of nearby Beijing, Tianjin's tone sandhi is richer. The relevant data are displayed in (2) (unless otherwise stated, all Tianjin data are cited from Li and Liu 1985).

(2) a. 21-21 → 213-21
b. 213-213 → 45-213
c. 53-53 → 21-53
d. 53-21 → 45-21

Other combinations do not exhibit sandhi. The analytical details often depend on the interpretation of the numerically transcribed citation tones. If we consider 213 to be low rise, 21 to be low fall, and 45 to be high rise, the four Tianjin tone sandhis can be reduced to two processes, informally stated as follows:

(3)	a.	low rise	→	high rise	cf.	(2b)
		high fall	→	low fall		(2c)
	b.	fall	→	rise		(2a, d)

The rules in (3a) target the register by raising it to derive 45 or lowering it to derive 21. Rule (3b) targets the contour, turning fall to rise. It is worth noting that (3a) does not affect the contour, while (3b) does not affect the register. This is significant, since it suggests a formal separation between register and contour. We will return to this issue in §3.1.

Not all tone sandhi phenomena can be analyzed with rules that manipulate either the register or the contour. There are sandhis which involve the lowering or raising of the pitch, as well as a change in contour. The Beijing Mandarin tone sandhi is one such example. Since Beijing Mandarin is designated as standard Chinese, its phonology is well known, and has been analyzed extensively in the scholarly literature (Chao 1968; Yip 1980, 1995, 2002; Z. Wu 1984; M. Chen 2000; Duanmu 2000). The account here follows M. Chen (2000). Like Tianjin, Beijing has four tones, 55 (T1), 35 (T2), 214 (T3), and 51 (T4). In disyllabic phrases, the third tone 214 becomes 35 when it is followed by another 214. This is the T3 Sandhi of Beijing Mandarin (M. Chen 2000: 102):

(4) 214 → 35 / __ 214

To derive 35, we need to change the contour from concave to rising and raise the pitch height at the same time. The T3 Sandhi is a composite process.

More specimens of tone sandhis are displayed in (5).

(5) a. *Sandhi involving register*
Luoyang (Mandarin; He 1996)
53-53 → 31-53
Pinggu (Mandarin; S. Chen 1998)
13-13 → 35-13
Pingyao (Mandarin; Hou 1980)
35-13 → 13-13
13-53 → 35-423

b. *Sandhi involving contour*
Mingshui (Mandarin; Gao 2000)
55 55 → 53-55
Pingyao (Mandarin; Hou 1980)
53-53 → 35-423
Pucheng (Min; Norman 1987)
35-24/<u>12</u> → 53-24/<u>12</u>
24-24/<u>12</u> → 31-24/<u>12</u>

c. *Sandhi involving register and contour*
Gaomi (Mandarin; Li 2004)
53-31 → 24-31
Pingyao (Mandarin; Hou 1980)
35-35 → 31-35
Yongkang (Wu; Yuan 1989)
35-T → 11-T
Raoping (Kejia; Zhan and Liu 2004)
<u>21</u>-T → <u>55</u>-T

In (5), T represents any tone. For each dialect, the major group to which it belongs is indicated, together with the source. Context-sensitive tone sandhi is attested in all major dialect groups, especially among Mandarin dialects.

2.2 Positional tone sandhi

Besides the context-sensitive tone sandhis that we see in Beijing, Tianjin, and some of the dialects in (5), many dialects exhibit sandhi conditioned by the position of the tone undergoing the sandhi. When we examine the sandhi behavior of the entire inventory of citation tones of a given dialect, we see two types of tonal systems that exhibit positional tone sandhi. In the first type, all citation tones in the inventory undergo sandhi by virtue of position. In the second type, some citation tones undergo positional tone sandhi, but other citation tones in the inventory either do not undergo sandhi or undergo sandhi conditioned by tonal context. We will describe these two subtypes in turn.

The Southern Min dialect of Xiamen has been analyzed extensively in generative phonology, and its tonology is well understood (see Yip 1980; M. Chen 1987, 2000; Duanmu 1995; Barrie 2006; among others). Xiamen has a total of seven citation tones, including two checked tones. The five non-checked tones are listed below (the Xiamen data are cited from M. Chen 2000):

(6)

Ia	44	Ib	24
II	53		
IIIa	21	IIIb	22

All five citation tones undergo sandhi in phrase-initial position, regardless of the tone that follows. The sandhi rules are illustrated below:

(7) a. 44 → 22 p'ang 44 'fragrance'
b. 24 → 22 we 24 'shoes'
c. 22 → 21 pī22 'sick'
d. 21 → 53 ts'u 21 'house'
e. 53 → 44 hai 53 'ocean'

These rules can be arranged in a circular fashion, as follows:

(8) 24 → 22 → 21 → 53 → 44 → 22

This is the so-called Min Circle (cf. M. Chen 1987). The checked tones 32 and 4 also exhibit positional tone sandhi, but do not participate in the Min Circle.

While all Xiamen citation tones change in phrase-initial position, only some of the tones in the Kejia dialect of Raoping exhibit this sort of sandhi behavior. Raoping has six citation tones, as shown below (the Raoping data are cited from Zhan and Liu 2004):

(9)

Ia	11	Ib	55
II	53		
III	35		
IVa	21	IVb	55

In disyllabic phrases, all six tones, except 11 and 55, undergo sandhi in phrase-initial position. The sandhis are summarized in (10).

(10) a. 55 → 31 / __ 11, 53, 21
55 → 33 / __ 55, 35, 55
b. 53 → 33 / __ 11, 35
53 → 11 / __ 55, 53, 21, 55
c. 35 → 33 / __ T
d. 21 → 55 / __ T

As can be seen, 55 and 53 undergo sandhi conditioned by context (10a, 10b), but 35 and 21 undergo positional sandhi (10c, 10d).

2.3 *Templatic tone sandhi*

In addition to contextual and positional tone sandhis, polysyllabic compounds or phrases in some dialects exhibit tonal patterns that cannot be derived in a straightforward manner from the lexically specified tones of the component syllables. Such tone patterns can be most profitably analyzed on the basis of a template with prespecified tones. The tone patterns of reduplicated doublets in two Mandarin dialects, Harbin and Wanrong, are typical examples of templatic tone sandhi. We will examine the two dialects in turn.

Monosyllabic nouns in Chinese reduplicate to express universal quantification, *ren* 'person' *vs. ren ren* 'each/every person'. In Wanrong, such reduplicated doublets surface with two tone patterns, illustrated below:

(15)		*base*	*reduplicant*		
	a.	51	24-51	tɕia tɕia	'every home'
	b.	24	24-51	p'ei p'ei	'every plate'
	c.	55	55-51	k'uei k'uei	'every bundle'
	d.	33	33-51	tɕy tɕy	'every sentence'

As can be observed in (15), the second copy has 51 regardless of the citation tone of the base. Furthermore, 51 reduplicates as 24-51. The patterns in (15) can be derived through a fixed template with prespecified tone 51, to which rule (14b) applies to derive pattern (15a). The template is shown below:

```
(16)  σ - σ
          |
          51
```

Unlike the template for disyllabic compounds (14a), the tonal template of nominal reduplication applies to all four citation tones in the inventory of Wanrong.

Templatic tone sandhi is not restricted to reduplicated doublets. Liujia, a Chinese dialect spoken in the linguistically mixed area of northern Guangxi, has a single robust disyllabic sandhi template for all citation tones that undergo sandhi. Liujia has a total of ten citation tones, including four checked tones. According to Zeng and Niu (2006), the dialect affiliation of Liujia is not easy to establish, perhaps due to contact with the Miao and Yao languages spoken in the region. Since the checked tones exhibit similar sandhi behavior to the non-checked tones, we will exclude the checked tones from our consideration. The six non-checked tones are 53, 343, 33, 35, 41, and 22. The falling tone 41 does not undergo sandhi, nor does 33, except when it is followed by 22. In phrase-initial position, 53 and 343 become 11. The sandhi behavior of 35 and 22 is more complex. In the vast majority of disyllabic phrases involving the two tones they surface as 11 in phrase-initial position, but in a small number of phrases they remain unchanged or surface as 22, due to a complex and idiosyncratic interplay between the syllable-initial consonant and the phrase-final tone. By Zeng and Niu's (2006) calculation, out of a total of 100 disyllabic tonal combinations (including the four checked tones), 60 combinations exhibit tone sandhi, out of which 51 combinations have 11 as the sandhi tone. Some of the disyllabic tone patterns are exemplified below:

(17)	a.	53-53 → 11-53	kou tei	'height'
		53-35 → 11-35	fɔŋ y	'wind and rain'
	b.	343-53 → 11-53	lɪaŋ sem	'conscience'
		343-35 → 11-35	nan min	'unavoidable'
	c.	35-53 → 11-53	ma tɕ'ie	'horse wagon'
		35-35 → 11-35	lou ma	'old horse'
	d.	22-53 → 11-53	lu təŋ	'street lamp'
		22-35 → 11-35	liu li	'manage'

The data support the template for disyllabic tone sandhi displayed in (18).

(18) σ - σ
     |
     11

According to Zeng and Niu (2006), the sandhi is triggered by iambic rhythm, where a citation tone is reduced to 11 in the metrically weak position. In view of M. Chen's (2000) accentual analysis of New Chongming (see §3.4 below), Zeng and Niu's (2006) characterization may be given an accentual interpretation: the prominent syllable retains its citation tone, and the other syllable assumes the default tone 11. What is clear from the data is that the robust sandhi template (18) has a strong neutralizing effect on disyllabic tonal contrasts.

2.4 Tone spread

In both contextual and templatic tone sandhis, the citation tone and its sandhi counterpart are confined to the lexically specified tone-bearing syllable. The sandhi tones do not spread to neighboring syllables of a polysyllabic string. While tone spreading is characteristic of African tone languages (cf. Odden 1995), it is also attested among Wu dialects. The Wu tonology has been discussed widely in the literature (see Zee and Maddieson 1979; Ballard 1980; Xu *et al.* 1981, 1982, 1983; Selkirk and Tong 1990; Qian 1992; Zhu 1999; M. Chen 2000; Rose 2000; Cao 2002; among many others). The Shanghai tone sandhi is representative of what I will call *limited tone spread* characteristic of the Wu dialects. Since it is well discussed in the generative literature, I will use it as an example. Unless otherwise stated, the data are cited from Xu *et al.*'s (1981, 1982, 1983) detailed work on Shanghai tone sandhi.

Shanghai has a total of five citation tones, 53, 34, 13, 55, and 12. The tone patterns of disyllabic, trisyllabic, and quadrisyllabic compounds are largely determined by the initial tones. The tone patterns with the initial tones 53, 34, and 13 are displayed in (19).

(19)

base	disyllabic	trisyllabic	quadrisyllabic
53	55-31	55-33-31	55-33-33-31
34	33-44	33-55-31	33-55-33-31
13	22-44	22-55-31	22-55-33-31

The checked tones exhibit the same sandhi behavior. Three properties of the Shanghai tonology are worth noting. First, the lexically specified tone spreads from the initial syllable to the neighboring syllable only, but not beyond. In other words, the spreading is local. Second, the morphosyntactic structure of the polysyllabic compound does not play a role in tone sandhi, as shown in the trisyllabic compounds below:

(20) 53-σ-σ → 55-33-31

a. tɕ'iɤ 53 [bi 13-tɕ'i 34] 'bad temper'
   bad temper

b. [ʔi 53-tsɿ 34] ti 34 'tobacco shop'
tobacco shop

The two compounds have different morphosyntactic structure, but exhibit the same tone pattern. Third, for trisyllabic strings, the remaining syllable is prespecified with the default tone 31, and for the quadrisyllabic strings, the default tones are 33-31. The default tones cannot be derived from the initial tone through spreading or some other means; Shanghai tonology retains the templatic tone sandhi that we have seen in other Chinese dialect families. Details aside, the Shanghai tone pattern can be summarized as follows (T = citation tone):

(21) *Tone sandhi of Shanghai compounds*

σ σ (σ σ)
T 33 31

The Shanghai tone patterns can be easily analyzed in generative phonology (see also CHAPTER 14: AUTOSEGMENTS). The precise analysis depends on what we assume to be the internal structure of the tones, especially contour tones (see §3.1). Here, we sketch the analysis of Selkirk and Shen (1990), which takes contour tones to be composed of level tones. Selkirk and Shen (1990) propose a three-step analysis: delete all non-initial citation tones, spread the initial tone, and insert the default tones. The derivation of (22) illustrates this analyis:

(22) Underlying tɕʻiɤ bi tɕʻi 'bad temper'
53 13 34

Tone deletion tɕʻiɤ bi tɕʻi
53

Tone spread tɕʻiɤ bi tɕʻi
5 3

Default tone tɕʻiɤ bi tɕʻi
53 33 31

Shanghai-style limited tone spread is widely attested among Wu dialects; see Qian (1992) and Cao (2002).

Tone spreading is freed from the Shanghai-type restriction in Danyang, a Wu dialect spoken in a region that borders on the Mandarin-speaking area. Danyang tone sandhi has attracted a lot of attention from generative phonologists; see Chan (1991), Yip (1989), Duanmu (1994), M. Chen (2000), and Bao (2004). All these works are based on the detailed analysis by Lü (1980). Danyang has a total of six of what Lü (1980: 88) calls Type A word melodies, listed in (23).

3 Theoretical issues in Chinese tonology

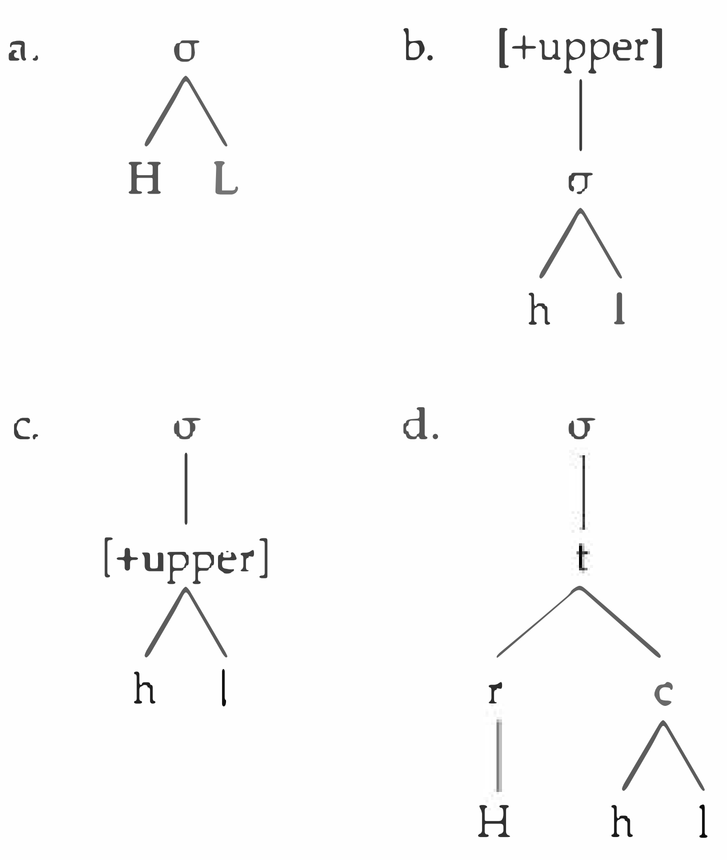
The tonal phenomena of Chinese have attracted a great deal of attention from linguists working in formal linguistics, especially in generative phonology. M. Chen (2000) summarizes the theoretical work on Chinese tone in terms of four 'leitmotifs': the internal structure of tone, the mechanism of tone sandhi, the relationship between tone and accent/stress, and the tone sandhi domain. Here, we will briefly summarize the results in these four areas of research on Chinese tone. The last section summarizes recent OT analyses of some of the same tonal phenomena.

3.1 *Internal structure of tone*

Among the world's tone languages, it has long been recognized that Chinese has contour systems, and African or native American languages have, by contrast, level (or even) systems; see Pike (1948). Both systems have tones that are phonetically level in pitch, such as H or L, and tones that are falling or rising in pitch movement. The crucial difference between the two systems is the treatment of contour tone. For Pike (1948) and early generative phonologists, contour tone is a basic unit, so a falling or rising tone is not phonemically equivalent to a combination of H and L, even though phonetically a presumptive unitary falling tone is indistinguishable from a sequence of H-L. Chinese dialectologists typically treat contour tones as basic units.

The far richer representational architecture of non-linear phonology (CHAPTER 45: THE REPRESENTATION OF TONE; CHAPTER 14: AUTOSEGMENTS; CHAPTER 27: THE ORGANIZATION OF FEATURES) allows different approaches to the structure of tone, especially contour tone. Using the high falling tone 53 as an example, we will examine four models that have been proposed by scholars working in generative phonology. They are shown in (26).

(26)



In (26), σ is the tone-bearing unit, and the tonal elements H, L, h, and l are defined in terms of the features [upper] and [raised], as follows (Yip 1980).

(27) H: [+upper]
L: [−upper]
h: [+raised]
l: [−raised]

Returning to the models shown in (26), (26a) is the Africanist model, in which H and L are associated with a single tone-bearing unit. This model is consistent with Woo's (1969) analysis, which attempts to reduce Chinese contour tones to combinations of level tones, in the same way that contour tones in African languages are typically analyzed. Here, σ is a tone-bearing unit which is sufficiently long to accommodate two tones with different pitch levels. In Woo's formulation, the two tones (H and L) are arranged sequentially within the feature matrix of the tone-bearing unit. The Africanist model is further argued for in Duanmu (1994).

The structures in (26b, 26c) are non-linear representations. In (26b), proposed by Yip (1980), the register feature [upper] and the pitch level feature [raised] are independent autosegments associated with a common tone-bearing unit. In (26c), proposed by Yip (1989), the register is analyzed as the tone root, to which h [+raised] and l [−raised] attach to specify the falling contour. The register feature is then associated with the tone-bearing unit. In (26d), proposed by Bao (1990, 1999a), tone consists of register and contour as sister nodes. The model in (26a) does not recognize the registral difference – the *yin–yang* opposition – in Chinese tones. The *yin–yang* opposition is expressed in (26b)–(26d), with register being expressed by the feature [upper] (*yin* [+upper], *yang* [−upper]). The formal difference among (26b)–(26d) is twofold. First, contour tone is treated as a unit in (26c, 26d), but not in (26b). Second, contour is dependent on register in (26c), not in (26b, 26d), where register and contour are independent structural elements. (26d) is the closest to the traditional analysis of tone in Chinese linguistics.

There is ample empirical evidence in support of the separation of register and contour. We have seen individual sandhis in (5); here we will pay attention to tone sandhi that involves an entire tone inventory of a given dialect. We will discuss two types of evidence. First, tones with the same contour may undergo the same tone sandhi process; in other words, contour can be used to define natural classes. Second, tones may harmonize their register without affecting the contour. We will look at contour-related sandhi phenomena first.

Yiyang, a Xiang dialect, has an interesting tone sandhi pattern in disyllabic compounds. Yiyang has two rising tones, 34 and 13, and two falling tones, 41 and 21. (It also has the checked tone <u>45</u>, which does not concern us here.) In disyllabic compounds, the tones undergo sandhi, as shown below (Xu 2001: 11–13):

(28) a. 34 → 33
i 34 sã 34 (33) 'doctor'
tɕiau 34 (33) fa 34 'to water flowers'
b. 13 → 33
fa 34 pu 13 (33) 'flowery cloth'
tɕi 13 (33) ma 41 'to ride horses'
c. 41 → 11
sã 34 tin 41 (11) 'hilltop'
ta 41 (11) li 13 'to thunder'

d. 21 → 11
ts'ə 34 tsã 21 (11) 'bus stop'
lo 21 (11) ts'ə 34 'to ride buses'

The sandhi tones are given in parentheses.

Yiyang exhibits a peculiar form of positional sandhi. In disyllabic compounds of the form verb–object, the first tone undergoes sandhi; in compounds with other morphosyntactic structures, the second tone undergoes sandhi. The sandhi process, regardless of position, is the same: the rising tones become 33 (28a, 28b), and the falling tones become 11 (28c, 28d). Tonal contour is used to define natural classes.

The same type of contour-driven tone sandhi is reported in disyllabic compounds in Yongfu, a Min dialect. Yongfu has a total of seven citation tones, shown below (Z. Zhang 1992: 23–25):

(29) Falling: 31, 53, <u>53</u>
Rising: 24
Level: 11, <u>55</u>, 21

In addition to the seven tones, Yongfu has the mid-level tone 33 in sandhi context only. The treatment of 21 is crucial for our analysis of the Yongfu tone sandhi. Since Yongfu already has a low falling tone 31, we will treat 21 as a lower-mid-level tone, despite the slight declination in pitch. In any given dialect, 21, 22, or 11 could very well be allotones; see Rose's (1990) detailed acoustic analysis of the Zhenhai tonology. The featural interpretation of the numerically transcribed tones needs to pay attention not only to the phonetic pitch that the numbers represent, but also to the tonal oppositions within the tonal inventory of a given dialect. For this reason we treat 21 as a lower-mid-level tone, identical to 22. We will, however, continue to follow Z. Zhang's (1992) transcription.

In Yongfu disyllabic compounds, the tones 24, 11, 21, and 55 become 33, and 31, 53, and <u>53</u> become 21. Some relevant compounds follow:

(30) a. kim 24 (33) tsam 24 'golden needle'
sioŋ 11 (33) loŋ 53 'pine nut'
p'ĩ 21 (33) au 53 'behind back'
siok <u>55</u> (33) go 11 'big goose'
b. gia 31 (21) gei 53 'skill'
ts'ie 53 (21) tsiŋ 24 '(purified) water'
tsuat <u>53</u> (21) ts'iaŋ 21 'extreme'

(30a) is the dominant pattern. According to Zhang (1992), these four tones become 55 when followed by 31 or <u>21</u>; we will ignore this detail here. It is clear from the data in (30) that Yongfu exhibits positional sandhi. Since we interpret 21 as the lower-mid tone 22, in opposition with the low tone 11 and the mid tone 33, we can state the Yongfu tone sandhi informally as follows:

(31) a. falling tones → 33
b. non-falling tones → 21 (= 22)

The rules apply to tones in initial position.

Zhangzhou, another Min dialect, has the same type of tone sandhi, which affects the stem tone when the stem is affixed with the diminutive suffix /a/ 53. Some examples are shown below.

(32) ts'ia 34 (33) a 53 'car'
hi 13 (33) a 53 'fish'
haŋ 22 (33) a 53 'lane'
kau 53 (55) a 53 'dog'
ts'i 21 (55) a 53 'house'

The data are cited from Yang (2006), where they are given in Chinese characters. Here, IPA transcriptions are used. The analysis crucially depends on the treatment of 21 in the Zhangzhou inventory. Assuming that 21 is the low falling tone in opposition to the high falling tone 53 in the Zhangzhou inventory, the tone sandhi can be formulated informally as follows:

(33) a. falling tones → 55
b. non-falling tones → 33

Like their counterparts in Yongfu, the two rules apply to tones in initial position. The Yongfu and Zhangzhou tone sandhi demonstrates the independent role that contour plays in defining natural classes, supporting the structure of tone shown in (26e), which encodes contour as sister to register.

Contour is not only used to define natural classes, it also participates in tone sandhi independently of register. This is the tone sandhi found in the dialect of Pucheng, which is located in northern Fujian. Pucheng has a total of seven citation tones: 35, 24, 55, 54, 12, 11, and 43. According to Norman (1987: 331), 12 has a "very tense glottal quality throughout." We can analyze 12 and 24 as allotones of the low rising tone. In Pucheng, as we saw in (5b), the rising tones 35 and 24 become 53 and 31, respectively, before the low rising tones 24 and 12; all other disyllabic combinations maintain their citation tones. The Pucheng sandhi system can be described with a simple rule, informally given in (34):

(34) rising → falling / __ low rising

Changes in contour do not affect the register: high rising 35 becomes high falling 53, and low rising 24 becomes low falling 31. Contour and register are independent.

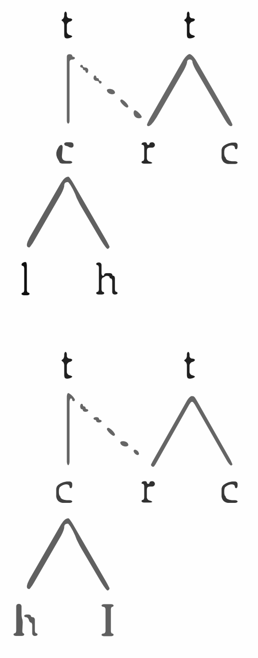
Register also plays a crucial role independently of contour. Here, we will examine the register harmony in the Min dialect of Chaozhou, based on the dialect of Jieyang (Cai 1991; Bao 1999b). The citation and sandhi tones are displayed in (35).

(35)

	citation	*sandhi*	*environment*
a.	53	35	53, 55, 5
		24	33, 213, 11, 2, 35
b.	213	53	53, 55, 5
		42	33, 213, 11, 2, 35
c.	33	33	all tones
d.	55	11	all tones
e.	35	21	all tones
f.	11	11	all tones

Chaozhou also has two checked tones, 2 and 5, which exhibit the same sandhi behavior as 213 and 55, respectively (35b, 35d). The sandhi affects tones in the initial position of disyllabic compounds. Like other Min dialects we have seen, Chaozhou exhibits a strong tendency toward positional sandhi. Of interest to us are the sandhis in (35a, 35b). Here, the citation tone 53 becomes rising, but the register is determined by the following tone. Similarly, the citation tone 213 becomes falling, with the register determined by the following tone. Using a tone model that allows the formal separation between register and contour, we can express the sandhis in (35a, 35b) as in (36):

(36) a. b.



While the register harmony is strong in Chaozhou's tone sandhi, the behavior of 35 presents a serious challenge to the analysis. In Yip's (1980, 1989) models, and in Bao's (1990, 1999a) model assumed in the analysis presented in (36), 35 must be a high-register tone, in opposition to the low-register 24. We would expect 35 to behave like other high-register tones, namely 53, 55, 5, but it behaves like a low-register tone. The behavior of 35 can be understood if it is not treated as a simple tone, but as a sequence of two level tones, 3 (L) and 5 (H), consistent with the Africanist model (26a). Although the Chaozhou data show robust register harmony, they leave a "wrinkle" that opens up the possibility of an Africanist analysis of contour tone.

3.2 Tone sandhi domain

By *tone sandhi domain* we mean the string of syllables in which tone sandhi rules apply. Most descriptive field reports and theoretical works on tone sandhi in Chinese linguistics focus on polysyllabic compounds, which serve as the default tone sandhi domain. Our survey of the typology of Chinese tone sandhi reveals significant variation among Chinese dialects in the way sandhi rules apply within such tone sandhi domains. This is, however, not the only way in which tone sandhi domains are formed. There are tone sandhi domains which are not morphosyntactic constituents. M. Chen (2000) calls them phonological words, which are bounded by edges of major phrases (CHAPTER 51: THE PHONOLOGICAL WORD). The tone sandhi processes are the same.

Xiamen and Shanghai offer interesting case studies of how phonological words are formed. We look first at Xiamen. We have seen the tonal inventory of Xiamen in (6) and the positional tone sandhi encapsulated in the Min Circle (8). The formation of the phonological word in Xiamen is illustrated in (37) (M. Chen 2000: 439):

(37) a. [ji [sia k'a kin]$_{VP}$]$_S$ 'he writes faster'
he write more fast
(44 53 53 53) → (22 44 44 53)

b. [ji [sia]$_{VP}$ #]$_S$ k'a kin 'it would be faster for him to write'
(44 53) (53 53) → (22 53)(44 53)

The brackets indicate syntactic structure, and the parentheses tone sandhi domains. The string is ambiguous: in (37a), *k'a kin* 'more fast' functions as an adjunct, modifying the verb *sia* 'write'. In (37b), *k'a kin* functions as the predicate, and the sentence *ji sia* 'he write' is the subject. The presence of a major phrasal boundary (VP) in (37b) is crucial; it separates the string into two tone sandhi domains. In (37a), by contrast, there is no string-internal major phrasal boundary, and the whole string is a single tone sandhi domain. M. Chen (2000: 441) formulates the rule as follows:

(38) {Right, X^{max}}, X^{max} not an adjunct

This rule places the boundary # to the right of VP in (37b), creating two tone sandhi domains. The rules in the Min Circle apply to derive the two surface tone patterns.

Shanghai offers a near mirror image of Xiamen. We have seen the Shanghai sandhi in §2.4, which is based on polysyllabic compounds. In phrases, the same local spreading rule applies. The key to understanding the tonology of sentences is phrasing – the formation of tone sandhi domains, which is illustrated in (39) (Selkirk and Shen 1990: 322).

(39) a. zɿ [ləʔ [zɑ̃ hɛ]]
(13 <u>12</u>) (13 34) → (22 44) (22 44)
live at Shanghai
'live in Shanghai'

b. ma [[liɑ̃ bən] [sɿ]]
(13 13 34) (53) → (22 55 31)(53)
buy some CL book
'buy some books'

For consistency, I use the segmental and tonal transcriptions of Xu *et al.* (1981, 1982, 1983), with [ɛ] for their [E]. As in the Xiamen data (37), the square brackets mark syntactic structure, and the parentheses tone sandhi domains. As can be seen, there is a mismatch between syntactic and phonological structures. According to Selkirk and Shen (1990), the crucial information is the left edge of the maximal projection of a lexical item, here the nominal phrases /zɑ̃ hɛ/ 'Shanghai' and /sɿ/ 'book'. The tone sandhi domains are determined by the following parameter (Selkirk and Shen 1990: 332):

(40) {Left, Lex^{max}}

With the tone sandhi domains set by (40), the sandhi rule informally shown in (21) applies to yield the observed tone patterns.

3.3 *Cyclicity and directionality*

In §2 we established the typology of tone sandhi among Chinese dialects. In positional and spreading sandhis, the tone sandhi domain plays a crucial role. Once the tone sandhi domain is formed, tone sandhi proceeds in a straightforward manner. As we have seen in §3.2, in dialects such as Xiamen and Shanghai, the domain-internal morphosyntactic structure, if it exists, does not influence the way tone sandhi rules apply. In many Mandarin dialects, tone sandhi is cyclic, where tone sandhi rules are regulated by morphosyntactic structure. Beijing T3 Sandhi (4) offers a classic example of cyclic application (M. Chen 2000: 102):

(41) a. [lao-hu] dan 'brave'
tiger gall
214-214 214 → 35 35 214
b. zhi [lao-hu] 'coward'
paper tiger
214 214-214 → 214 35 214

T3 Sandhi applies cyclically, first to *lao-hu*, destroying the context in (41b) in the second cycle. Such cyclic application of tone sandhi rules appears to be found in context-sensitive and non-spreading tone sandhi systems. In the spreading (Wu) and positional (Min) types, tone sandhi is not cyclic.

The matter is more complex. Even among Mandarin dialects, cyclicity cannot be taken for granted. In Tianjin, which we first saw in §2.1, the sandhi patterns of trisyllabic compounds do not exhibit the effect of cyclicity. The disyllabic sandhi rules shown in (2) are repeated below:

(42) a. 21-21 → 213-21
b. 213-213 → 45-213
c. 53-53 → 21-53
d. 53-21 → 45-21

Most trisyllabic patterns can be derived through the left-to-right application of these basic sandhi rules. This is illustrated in (43) (Li and Liu 1985: 79).

(43) 213-213-213 → 45-45-213

a. tʂ'aŋ [taŋ uei]
factory party committee
'factory party committee'
b. [ɕi lian] ʂuei
wash face water
'water for washing faces'

The two compounds have different morphosyntactic structure, but the same tone pattern. The syntactic bracketing plays no role in the sandhi process. Treating the whole compound as a single tone sandhi domain, we can derive the pattern by applying rule (42b) left to right iteratively:

(44) (213-213)-213 → 45-(213-213) → 45-45-213

But the left-to-right application of the sandhi rules does not work for all cases. Li and Liu (1985: 80) point out the "exceptional" sandhi behavior of compounds composed of three 53 tones, shown below:

(45) 53-53-53 → 53-21-53

The left-to-right iterative application of (42c) yields the wrong pattern 21-21-53, or 213-21-53 with the additional application of (42a); see (46a). Here the pattern can be derived through the right-to-left application of (42c), bleeding the environment for the second iteration of the rule; see (46b).

(46) a. *Left-to-right application*
(53-53)-53 → 21-(53-53) → (21-21)-53 → 213-21-53
b. *Right-to-left application*
53-(53-53) → 53-21-53

Given the understanding of rule application in generative phonology, one would expect cyclicity or directionality to be consistent in regulating tone sandhi in any given system. The Tianjin tone sandhi facts are unexpected. What is clear is that the trisyllabic tone patterns avoid tonal sequences that would violate any of the four rules in (42). The directionality-related phenomena resist straightforward analysis.

3.4 Tone, accent, and stress

Tone, accent, and stress are notoriously slippery notions in the Chinese tonological literature. Whereas tone is part of the lexical specification of morphemes, it is not clear how it relates to accent and stress in a given tone sandhi system, if indeed the latter two notions can be made precise (see CHAPTER 42: PITCH ACCENT SYSTEMS). Here I will follow McCawley (1968) in characterizing the distinctions as in (47):

(47) a. Tone: (σ)(σ)(σ)(σ)
b. Stress: (σσ)(σσ)
c. Accent: (σσσσ)

The three notions are differentiated in terms of the size of units in which tonal contrasts are maintained. In the tonal system, tonal contrast is maintained in the syllable – in a polysyllabic string each syllable is associated with a tone. In the metrical (stress) system, it is maintained in the foot, and in the accentual system, in a unitary domain. We will see that not all Chinese dialects are tonal; some have developed into an accentual system, as argued by M. Chen (2000).

We can approach the relationship of tone with accent and/or stress from two perspectives: tonal reduction and the prominence relations that underpin tone sandhi. Generally speaking, the former is the focus of research in the Chinese linguistic circle, and the latter the focus within the general framework of generative phonology. We will discuss tonal reduction first.

Across Chinese dialects, there is a widespread phenomenon known as the "light tone." The light tone can be found on system words or affixes, which may be proceeds not by position but by tone, in accordance with the Middle Chinese (ca. 900 CE) classification. Chongming has all the Middle Chinese tones, shown below:

(50) *Citation tones of old and new varieties of Chongming*

	Old	New		Old	New
Ia	55	H	Ib	24	L
IIa	424	HM	IIb	242	ML
IIIa	33	MH	IIIb	313	LM
IVa	<u>55</u>	<u>H</u>	IVb	<u>22</u>	<u>L</u>

The Old Chongming tones are cited from H. Zhang (1979, 1980), and the New Chongming tones from M. Chen (1997, 2000). The checked tones <u>55</u> and <u>22</u> exhibit the same sandhi behavior as 55 and 24, respectively. The citation tones of New Chongming have undergone slight change from their Old Chongming counterparts, especially IIIa. M. Chen (1997: 199) treats LM and L as allotonic variants of Ia; the former variant is equivalent to 24 in Old Chongming. Other than the slight difference in the phonetic realization of IIIa, there is little difference between the old and new varieties of Chongming.

According to Chen (1997), the disyllabic tone sandhi in New Chongming divides the eight tones into two natural classes: the even tones I and IV (H/<u>H</u> and L/<u>L</u>) and the oblique tones II and III, all of which exhibit contour in New Chongming. The robust sandhi patterns involve I/IV, which surface invariably as H in phrase-final position, and retain their citation pitch values in phrase-initial position. We summarize the patterns below (O = oblique tone; E = even tone; σ = syllable with default tone; adapted from M. Chen 1997: 194):

(51) a. O-E → σ-H
b. E-O → E-σ
c. E-E → E-H

When the final tone is E (H/<u>H</u>, L/<u>L</u>), it surfaces as H. When the first tone is E, it remains unchanged, yielding two broad patterns: H-σ (< Ia/IVa-σ) and L-σ (< Ib/IVb-σ). E-E doublets surface as either H-H (< Ib/IVb-σ) or L-H (< Ib/IVb-σ). The syllable σ with an oblique tone loses its lexically specified citation tone, and acquires the default tone. The sandhi behavior of O-O is similar to the sandhis shown in (51), but involves more complex pitch realizations. To avoid complications, we will not include the O-O sandhis in our discussion; for a complete OT-inspired analysis, see M. Chen (1997, 2000).

The sandhi patterns summarized in (51) are also attested in Old Chongming, with negligible differences in the pitch values of the default tone, which is 31 in initial position, and 33 or 3 in final position (H. Zhang 1979). Incidentally, by Zhang's (1979) description and analysis, we cannot use contour to classify the Chongming tones as a natural class, since Ia is 55 and Ib is 24; see (50). The tonal categories I and IV play a crucial role in determining the observed surface tone patterns of doublets.

The patterns in (51) lend themselves to a two-step analysis: (1) mark accentual prominence on the even tones (I, IV), which surface as H in final position, and remain unchanged in initial position; (2) the unaccented syllable loses its lexical and output 13s to match their slope, and prevent 23s from matching theirs. Furthermore, the local left-to-right spread of Shanghai (see (21)) is due to MATCH-SLOPE, instead of the usual OT treatment of spreading in terms of alignment.

4 Conclusion

In the preceding sections we have surveyed the whole range of tone sandhi across Chinese dialects, and some of the central issues in Chinese tone that have attracted the attention of theoretical phonologists. Through the collective effort of field linguists and theoretical phonologists, we now have a good understanding of the sandhi phenomena in Chinese dialects. While Pike's (1948) Asianist–Africanist dichotomy has ample empirical support from Chinese dialects, it presents Chinese tonology as a monolithic block that obscures a varied landscape of the tone sandhi typology. This typology is summarized below:

(54) a. Contextual sandhi, e.g. Tianjin
b. Positional sandhi, e.g. Xiamen
c. Templatic sandhi, e.g. Liujia
d. Tone spread, e.g. Shanghai

Our survey shows a strong clustering tendency that coincides with major dialect boundaries. Contextual sandhi is most commonly found in Mandarin, positional sandhi in Min, and tone spread in Wu. All three types may contain tonal templates.

The complex landscape of tone sandhi presents interesting empirical data for phonological theory. The tone sandhi data have contributed to our understanding of the theoretical issues such as the structure of tone, tone sandhi domain, cyclicity, and the interaction of tone with accentual and/or metrical prominence. For the past half century, Chinese tone sandhi has attracted the attention of phonologists working mainly in the evolving theories of generative phonology, from the rule-based approach of classical generative phonology to the rule- and constraint-based approach of non-linear phonology, and now to the constraint-based approach of Optimality Theory. Due to the complexity of the sandhi phenomena, many of the theoretical issues remain unresolved.

REFERENCES

Ballard, William L. 1980. On some aspects of Wu tone sandhi. *Journal of Asian and African Studies* 19. 83–163.
Bao, Zhiming. 1990. On the nature of tone. Ph.D. dissertation, MIT.
Bao, Zhiming. 1999a. *The structure of tone*. New York & Oxford: Oxford University Press.
Bao, Zhiming. 1999b. Tonal contour and register harmony in Chaozhou. *Linguistic Inquiry* 30. 485–493.
Bao, Zhiming. 2003. Tone, accent, and stress in Chinese. *Journal of Linguistics* 39. 147–166.
Bao, Zhiming. 2004. Accentualism in Chinese. *Language and Linguistics* 5. 863–889.
Bao, Zhiming. 2005. Fangyan jiechu dui biandiao yufa de yingxiang [The impact of dialect contact on the grammar of tone sandhi]. *Yuyanxue Luncong* 31. 22–47.
Barrie, Michael. 2006. Tone circles and contrast preservation. *Linguistic Inquiry* 37. 131–141.

Odden, David. 1995. Tone: African languages. In Goldsmith (1995), 444–475.
Pike, Kenneth L. 1948. *Tone languages.* Ann Arbor: University of Michigan Press.
Qian, Nairong. 1992. *Dandai Wuyu yanjiu* [*Studies in modern Wu*]. Shanghai: Shanghai Education Press.
Rose, Phil. 1990. Acoustics and phonology of complex tone sandhi. *Phonetica* 47. 1–35.
Rose, Phil. 2000. Independent depressor and register effects in Wu dialect tonology. *Journal of Chinese Linguistics* 30. 39–81.
Selkirk, Elizabeth & Tong Shen. 1990. Prosodic domains in Shanghai Chinese. In Sharon Inkelas & Draga Zec (eds.) *The phonology–syntax connection*, 313–337. Chicago: University of Chicago Press.
Woo, Nancy. 1969. Prosody and phonology. Ph.D. dissertation, MIT.
Wu, Jiansheng. 2008. Wanrong fangyan liangzi zu liandu biandiao he qingsheng [Tone sandhi in doublets and light tone in the Wanrong dialect]. *Fangyan* 2008. 303–311.
Wu, Zongji. 1984. Putonghua sanzizu biandiao guilü [Principles of trisyllabic tone sandhi in Mandarin Chinese]. *Zhongguo yuyan xuebao* 1984. 70–92.
Xu, Baohua, Tang Zhenzhu & Qian Nairong. 1981. Xingpai Shanghai fangyan de liandu biandiao [Tone sandhi in New Shanghai] Part 1. *Fangyan* 1981. 145–155.
Xu, Baohua, Tang Zhenzhu & Qian Nairong. 1982. Xingpai Shanghai fangyan de liandu biandiao [Tone sandhi in New Shanghai] Part 2. *Fangyan* 1982. 115–128.
Xu, Baohua, Tang Zhenzhu & Qian Nairong. 1983. Xingpai Shanghai fangyan de liandu biandiao [Tone sandhi in New Shanghai] Part 3. *Fangyan* 1983. 197–201.
Xu, Hui. 2001. *Yiyang fangyan yufa yanjiu* [*Studies of the grammar of Yiyang dialect*]. Changsha: Hunan Education Press.
Yang, Xiuming. 2006. Zhangzhou fangyan "a" zuici de liangzi zu liandu biandiao [Tone sandhi in disyllabic phrases with the suffix "a" in the Zhangzhou dialect]. *Fangyan* 2006. 50–55.
Yip, Moira. 1980. The tonal phonology of Chinese. Ph.D. dissertation, MIT.
Yip, Moira. 1989. Contour tones. *Phonology* 6. 149–174.
Yip, Moira. 1995. Tone in East Asian languages. In Goldsmith (1995), 476–494.
Yip, Moira. 2002. *Tone.* Cambridge: Cambridge University Press.
Yuan, Jiahua. 1989. *Hanyu fangyan gaiyao* [*An introduction to Chinese dialects*]. 2nd edn. Beijing: Wenzi Gaige Chubanshe.
Zee, Eric & Ian Maddieson. 1979. Tones and tone sandhi in Shanghai: Phonetic evidence and phonological analysis. *UCLA Working Papers in Phonetics* 45. 93–129.
Zeng, Xiaoyu & Niu Shunxin. 2006. Liujia hua liangzi zu liandu de yunlü biandiao jiqi yuanyin chutan [Preliminary investigation into the rhythmic tone sandhi in disyllabic phrases in the Liujia dialect and its causes]. *Fangyan* 2006. 290–308.
Zhan, Bohui & Liu Zhenfa. 2004. Guangdong Raoping Shangrao Kejia hua de liangzi liandu biandiao [Disyllabic tone sandhi in the Kejia dialect of Shangrao, Raoping, Guangdong]. *Fangyan* 2004. 195–199.
Zhang, Huiying. 1979. Chongming fangyan de liandu biandiao [Chongming tone sandhi]. *Fangyan* 1979. 284–302.
Zhang, Huiying. 1980. Chongming fangyan de sanzizu liandu biandiao [Trisyllabic tone sandhi in Chongming]. *Fangyan* 1980. 15–34.
Zhang, Jie. 2007. Directional asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics* 16. 259–302.
Zhang, Zhengxing. 1992. *Zhangping fangyan yanjiu* [*Studies on the Zhangping dialect*]. Beijing: Zhongguo Shehui Kexue.
Zhong, Qi. 2003. Changsha hua de qingsheng [Light tones in the Changsha dialect]. *Fangyan* 2003. 255–264.
Zhu, Xiaonong. 1999. *Shanghai tonetics.* Munich: Lincom Europa.

108 Semitic Templates

Outi Bat-El

1 Introduction

Semitic morphology is characterized by phonological restrictions on the shape of the words, allowing only a limited set of prosodic templates and vocalic patterns. The prosodic templates, which set the size restrictions on words and display the permissible syllable structure, host a limited set of vocalic patterns, and in some cases also affixes. The stem consonants fit into the consonantal positions provided by the prosodic templates, as do the vowels of the vocalic pattern.

To clarify these notions, consider the derivational paradigm in Table 108.1. The words in Table 108.1 are structurally related on both the vertical and horizontal axes of the paradigm. On the horizontal axis, they share the stem consonants. On the vertical axis, they share a prosodic template, a vocalic pattern, and, in the two rightmost columns, an affix. The properties on the vertical axis provide words with their phonological structure. When these properties are combined, i.e. CaCaC, hiCCiC, CCiCa, they form what is known as "Semitic templates" (McCarthy 1981). Throughout the chapter, I use the term "configuration" for this

Table 108.1 Derivational paradigm in Modern Hebrew

Prosodic template:	CVCVC		CVCCVC		CCVCV		*Stem consonant*
Vocalic pattern:	{a a}		{i i}		{i}		
Affix:			*h-*		*-a*		
	ga'dal	'to grow'	hig'dil	'to enlarge'	gdi'la	'growing'	{g d l}
	sa'gar	'to close'	his'gir	'to extradite'	sgi'ra	'closing'	{s g r}
	ʃa'tak	'to keep quiet'	hiʃ'tik	'to quieten'	ʃti'ka	'silence'	{ʃ t k}
	ka'lat	'to absorb'	hik'lit	'to record'	kli'ta	'absorption'	{k l t}
	za'rak	'to throw'	hiz'rik	'to inject'	zri'ka	'throwing, injection'	{z r k}

combination of properties, while the term "template" is used exclusively for prosodic templates.

The system of configuration is found in all Semitic languages in various paradigmatic relations, although to various degrees. This is demonstrated in §2 with examples from several languages. As the configurations consist of both segmental and prosodic elements, their structure is non-linear (see CHAPTER 105: TIER SEGREGATION). The structure of the configurations, and of Semitic words in general, is introduced in §3, with emphasis on theoretical developments in the representation of the prosodic template. The two ensuing sections illustrate the mapping of the configurations, where §4 is devoted to a procedural approach and §5 to a constraint-based approach. Within each of these two sections, two types of input are considered: the consonantal root and the word/stem. The concluding remarks in §6 draw attention to the status of the configurations within a cross-linguistic perspective.

2 The nature of Semitic morphology

In Semitic morphology, words are organized into classes, identified by their configuration. The class system in verbs (§2.1) is more prominent and restricted than in nouns (§2.2); nevertheless, the same structural generalizations hold, regardless of the lexical category.

2.1 *Verb classes (binyanim)*

Verbs belonging to the same class have an identical configuration, as exemplified in (1). The designator of a verb class is the 3rd masculine singular perfect, which is free of inflectional affixes.[1]

(1) *Verb classes in Palestinian Arabic* (Elihay 2004)

CiCeC		tCaCCaC		staCCaC	
'misek	'to grasp'	'tʕallam	'to study'	'staʕmal	'to use'
'fihem	'to understand'	'tdarraʕ	'to burp'	'staqbal	'to welcome'
'nizel	'to go down'	'tbaddal	'to be replaced'	'stawʕab	'to take in'
'ħimeq	'to lose temper'	'tbaxxar	'to evaporate'	'staslam	'to submit'
'liħeq	'to catch'	'tdawwaʃ	'to shower'	'staʃhad	'to quote'

The restricted structure of the configurations is evident in (1), where all the configurations are disyllabic, with a final CVC syllable. They differ in the first syllable, CV in Ci.CeC and CCVC in staC.CaC and tCaC.CaC. The latter two configurations are distinguished by their prefixes, where consonant positions (C-slots) not occupied by an affix are left for the stem consonants. In staC.CaC the prefix

[1] Throughout the chapter, I do not consider irregular verbs, which for phonological reasons (often only historically motivated) deviate from the regular configuration. I also ignore the epenthetic [ʔi] in Arabic, which rescues word-initial clusters. Epenthesis is obligatory in Standard Arabic (e.g. [ʔis'taʕmal] 'to use'), but optional in spoken dialects, unless the word appears phrase-initially or after a consonant-final word (e.g. ['baːsil is'taqbalo] *vs.* ['faːdi (i)s'taqbalo] 'Basil/Fadi welcomed him').

occupies the two positions of the initial complex onset, thus leaving three slots for the stem consonants, while in tCaC.CaC the prefix occupies only one position in the complex onset, thus leaving four slots for the consonants. When a configuration provides four slots for stem consonants but the stem has only three different consonants, one consonant occupies two slots (e.g. ['tʕallam] 'to study' *vs.* ['tħarkaʃ] 'to provoke').

Semitic languages vary with regard to prosodic "plasticity." Hebrew verbs accommodate as many stem consonants as possible (i.e. respecting the OCP and the Sonority Sequencing Generalization; CHAPTER 49: SONORITY), as long as the verb does not exceed the disyllabic maximal size (Bat-El 1994a, 2003a). In contrast, Amharic adjusts the number of syllables in the template according to the number of consonants (McCarthy 1985; Rose 2003). That is, Hebrew expands its syllabic inventory beyond CV and CVC, keeping the disyllabic template, while Amharic expands its syllabic template, keeping a restricted syllabic inventory (see Bender and Fulass 1978 for a study of Amharic verbs).

(2) *Templatic plasticity (Amharic)* vs. *syllable plasticity (Hebrew)*

	Amharic		*Hebrew*	
3 Cs	səbbər-ə	'to break'	siper	'to tell'
4 Cs	məsəkkər-ə	'to testify'	tirgem	'to translate'
5 Cs	tə-ngədaggəd-ə	'to walk in a zigzag'	flirtet	'to flirt'
6 Cs	aqənatt'ər-ə	'through violently'	trinsfer	'to transfer'

The vocalic pattern of the configuration is an arbitrary subset of possible combinations of vowels. The vocalic patterns in Tigrinya verbs, for example, employ only four ([i ɨ ə a]) out of the seven vowels ([i ɨ u e ə o a]) in the language (Buckley 2003).

(3) *Vocalic patterns in Tigrinya* (Buckley 2003)

		'offend'		'bless'
simple perfective	əə	bəddəl-ə	aə	barək-ə
simple gerundive	əi	bəddil-u	ai	barik-u
causative jussive	əɨ	jə-bəddɨl	aɨ	jə-barɨk

In Standard Arabic, however, which has only three vowels in its inventory, the vocalic patterns in the verb inflectional paradigm are partially predicted, given a base with a specified vowel (Guerssel and Lowenstamm 1996).

(4) *Predictable vocalic patterns in Standard Arabic (binyan I)*

perfective			*imperfective*		
Ø	'dˤarØb	→	i	'jadˤrib	'to hit'
i	'labis	→	a	'jalbas	'to dress'
a	'katab	→	u	'jaktub	'to write'
u	'kabur	→	u	'jakbur	'to be great'

Verb configurations are often associated with thematic-syntactic properties, such that verbs in different configurations are derivationally related if they share the same stem consonants.

(5) *Derivational relations in Modern Hebrew verbs*

CaCaC		CiCeC		hitCaCeC	
ga'dal	'to grow'	gi'del	'to raise'	hitga'del	'to aggrandize'
xa'ʃav	'to think'	xi'ʃev	'to calculate'	hitxa'ʃev	'to consider'
ka'dam	'to precede'	ki'dem	'to promote'	hitka'dem	'to progress'
pa'rak	'to unload'	pe'rek	'to dismantle'	hitpa'rek	'to disintegrate'
ka'ʃar	'to bind'	ki'ʃer	'to connect'	hitka'ʃer	'to get in touch'

The thematic-syntactic properties of the configurations are relational rather than absolute, such that the property assigned by a configuration is largely contingent upon the base of the derived verb (Berman 1978; Horvath 1981; Doron 2003; Laks 2007). For example, the Hebrew configuration hitCaCeC assigns deaccusative in [hir'giz] 'to make someone angry' → [hitra'gez] 'to become angry', but reciprocal in [xi'bek] 'to hug' → [hitxa'bek] 'to hug each other'. Moreover, [hitʔa'lel] 'to torture' is neither deaccusative nor reflexive, as it is not derived from another verb. Similarly in Arabic, ʔaCCaC assigns causative in ['ʤalas] 'to sit down' → ['ʔaʤlas] 'to bid one to sit down', but ['ʔarsal] 'to send' is not causative, since it does not have a base verb (Wright 1962). Verbs sharing a configuration may also share a semantic property, as is the case with Arabic (t)CajCaC verbs. Watson (2006: 192) reports that in most dialects these verbs refer to "physical state with pejorative overtones of pretence," but in San'ani Arabic they denote "harmless childish naughtiness."

The grammatical function of the configuration is apparent in various Semitic languages, but not in all. In Modern Aramaic, which has only two verb classes (as opposed to five in Hebrew, 11 in Maltese, and 14 in Standard Arabic), the configurations have only structural properties, to the extent that there are hardly any related verbs from the two classes (Hoberman 1992).

A verb class defines the inflectional paradigm of the verb, such that verbs belonging to the same class have the same configuration in every form in the paradigm (Aronoff 1994; see also CHAPTER 83: PARADIGMS).

(6) *Inflectional paradigms in Palestinian Arabic verbs* (Elihay 2004)[2]

perfect	*imperfect*		*perfect*	*imperfect*	
CiCeC	-iCCaC		CaCCaC	-iCaCCeC	
'misek	-'imsak	'to grasp'	'massak	-'massek	'to let hold'
'fihem	-'ifham	'to understand'	'fahham	-'fahhem	'to explain'
'nizel	-'inzal	'to go down'	'nazzal	-'nazzel	'to bring down'
'ħimeq	-'iħmaq	'to lose temper'	'ħammaq	-'ħammeq	'to make angry'
'liħeq	-'ilħaq	'to catch'	'laħħaq	-'laħħeq	'to manage'

Since the inflectional paradigm of a verb is contingent upon its configuration, every new verb must belong to one of the verb classes. This is manifested by the native configuration found in loan verbs, such as [til'fen] 'to phone', [si'mes] 'to send

[2] Person/number/gender features are indicated by suffixes in the perfect (e.g. ['fihm-u] 'they understood', [fah'ham-ti] 'you (FEM SG) explained') and prefixes + suffixes (in some forms) in the imperfect (e.g. ['t-ifham-u] 'you (PL) will understand', [n-'fahhem] 'we will explain'). As all imperfect forms take a prefix, the imperfect stems are preceded by a dash.

However, a rich configuration system in the nominal category is found in the singular/plural paradigm of several Semitic languages (see Ratcliffe 1998b for a comparative study), such as Arabic (Hammond 1988; McCarthy and Prince 1990; Ratcliffe 1997, 1998a; McCarthy 2000; Watson 2002, 2006), Tigre (Palmer 1962; Raz 1983), and Tigrinya (Palmer 1955; Buckley 1990). In Arabic, for example, most underived nouns and lexicalized derived nouns (Abd-Rabo 1990; Boudelaa and Gaskell 2002) are pluralized in a configuration system called "broken plural," which contrasts with the suffixation mode of pluralization called "sound plural."

(8) *Broken plurals*

a. *Standard Arabic* (Wright 1962)

	plural	*singular*	
ʔaCCa:C	ʔaħ'ka:m	ħukm	'judgment'
	ʔaq'da:m	'qadam	'footstep'
	ʔaj'ma:n	ja'mi:n	'oath'
ʔaCCuC	'ʔabħur	baħr	'sea'
	'ʔazmun	'zaman	'time'
	'ʔalsun	li'sa:n	'tongue'
CuCuC	'suquf	saqf	'roof'
	'ʔusud	'ʔasad	'lion'
	'surur	sa'vri:r	'throne'
CiCa:C	ri'ma:ħ	rumħ	'spear'
	xi'ba:l	'xabal	'hill'
	li'ʔa:m	la'ʔi:m	'base'

b. *Tigre* (Palmer 1962)

	plural	*singular*	
ʔəCCiC	ʔəkbid	kəbid	'belly'
	ʔəbʔis	biʔis	'husband'
	ʔəbħɨr	biħər	'sea'
ʔəCCuC	ʔəqlub	qəlɨb	'root'
	ʔəmtud	mɨtid	'stake'
	ʔədhub	dihəb	'gold'
CəCəCCi	ħənəddi	ħɨndi	'hoof'
	kətərri	katra	'pigeon'
	kədəbbi	kədbet	'floor'
CəCaCit	məsanit	məsni	'friend'
	wərazit	wəreza	'bachelor'
	ʔərawit	ʔərwe	'serpent'

While in the verb system, every class has a fixed confiiguration for each tense/aspect form, allowing a predictable system of one-to-one correspondence, in the noun system there is one-to-many correspondence (Bateson 2003). As shown in (8) above, a singular configuration may correspond to several plural configurations (e.g. Arabic CVCC in [ħukm] – [ʔaħ'ka:m] *vs.* [rumħ] – [ri'ma:ħ]). In addition, a singular noun may have two or three alternative corresponding plural forms, more so in the spoken dialects (e.g. Arabic ['ʕanzi] – [ʕi'na:z] ~ [ʕu'nu:z] ~ ['ʔaʕnuz] 'goat(s)', ['qafil] – [ʔaq'fa:l] ~ ['ʔaqful] 'lock(s)'). However, as discussed in §4.2 below, there are some tendencies for nouns with certain configurations to select particular plural configurations.

The system of configurations is not limited to nouns and verbs. As shown below, Amharic argot (Leslau 1964) and Arabic hypocoristics (Davis and Zawaydeh 2001) take specific configurations (Caj(C)CiəCi and CaCCuːC, respectively) regardless of the shape of the base. Similarly, Arabic adjectives and superlatives (Wright 1962) each take a consistent configuration.

(9) *Other configurations*

a. *Arabic adjectives*

CaCiːC		ʔaCCaC	
ka'biːr	'big'	'ʔakbar	'bigger'
sˤa'ɣiːr	'small'	'ʔasˤɣar	'smaller'
tˤa'wiːl	'long'	'ʔatˤwal	'longer'
ba'ʕiːd	'far'	'ʔabʕad	'farther'

b. *Arabic hypocoristics*		c. *Amharic argot*			
CaCCuːC		Caj(C)C$_i$əC$_i$			
bas'suːm	'basma		bajtət	bet	'house'
sal'luːm	sa'liːm		gajbəb	gəbba	'enter'
jas'suːr	'jaːsir		zajfnən	zəffənə	'sing'
sam'muːr	sa'miːra		wajrk'ək'	wərk'	'gold'

The preference for disyllabic forms, also exhibited in (9), is overwhelming, although, as noted above, some languages display a limited flexibility. This preference is found also in Hebrew acronym words (Bat-El 1994b; Zadok 2002), which are mostly disyllabic, regardless of the number of words in the input. For example, both the two-word base [mat'bea xuc] 'foreign currency' and the three-word base [mer'kaz texno'logia xinu'xit] 'educational technology center' surface as the acronym word [ma'tax]. Moreover, four-word bases also give rise to disyllabic acronym words, as in [mam'ram], whose base is [mer'kaz maxʃe'vim (ve-)ri'ʃum memu'kan] 'automated computer center'.

3 The structure of Semitic configurations

The term configuration correlates with the traditional Semitic template/pattern, represented as fully specified words, such as [qaːtˤal, qitˤtˤel, hiqtˤil], etc. (also [paːʕal] or [paːqad] in Hebrew and [faʕal] in Arabic).[4] As the stem consonants ({qtˤl}, {pʕl}, {pqd}, {fʕl}) are not part of the configuration, replacing them with C-slots gives us the type of structure used in the previous sections, i.e. CaːCaC, CiCCeC, hiCCiC, etc.

Configurations play a central role in the older grammarian studies (see §4.2) of Semitic morphology, mostly with reference to class membership and relations among words. In the absence of a theoretical model of non-linear phonological structure, these studies do not consider the internal structure of the configuration, and refer to relations among words in terms of phonological alternations, such as vowel change/ablaut and gemination.

4 The traditional terminology associated with the configurations is *wazn* (plural *ʔawzaːn*) in Arabic, and *binyan* (plural *binyanim*) for verbs and *miškal* (plural *miškalim*) for nouns in Hebrew.

McCarthy (1981) points out the restrictive nature of the templates in (11). Every template in (11a) has a counterpart with an initial complex onset in (11b), and another with an additional CV syllable in (11c). The absence of CVCVCVC in the first row in (11c) is due to a constraint prohibiting a sequence of two light syllables. These templates can be expressed with an archi-template, which generalizes all and only the possible templates in (11), with the addition of a vowel deletion rule that resolves the prohibited sequence of two light syllables.

(12) *Archi-template of the prosodic templates in Standard Arabic verbs* (McCarthy 1981)

$$\left(\begin{Bmatrix} C \\ CV \end{Bmatrix}\right) \; CV \; \left(\begin{Bmatrix} C \\ V \end{Bmatrix}\right) \; CVC \qquad (V \rightarrow \text{Ø} \; / \; CVC \_\_ CVC)$$

As shown in Table 108.2, there are more configurations than CV templates, where the difference among configurations sharing a template is in the affixes and their position, as well as in the distribution of the stem consonants.

Although the templates consist of CV-slots, reference to the syllable is inevitable, as seen in McCarthy's (1981) generalizations, such as "no binyan which begins with a consonant cluster is three or more syllables long overall" (1981: 386). This statement refers directly to the restriction on the number of syllables in the template, which is not explicitly expressed in the CV template. There is definitely a disyllabic core template, which can be minimally expanded with either CV or C (see Kiparsky 2003 for C as a demi-syllable). Moreover, as shown in Table 108.2 (c), the CV expansion is always a derivational prefix. In addition, Lowenstamm and Kaye (1986), in their study of compensatory lengthening in Tiberian Hebrew, demonstrate the essential role of the syllable in the configurations.[5]

3.2.2 *Syllabic templates*

The templates in (a) and (b) of Table 108.2 are disyllabic. The two trisyllabic templates in (c) are disyllabic on the stem level, since the initial CV, as noted above, is occupied by a derivational prefix (e.g. [ta-'baddal] 'to be replaced', [ta-'dˤa:jaq] 'to be disturbed'). That is, verbs in Arabic are disyllabic either on the stem level, the word level, or both. The same is true for Hebrew verbs, which can be disyllabic with or without a derivational prefix (e.g. [hik'dim] 'to come first', [ki'dem] 'to promote'), or trisyllabic, but only with a prefix (e.g. [hitka'dem] 'to progress').

As argued in McCarthy and Prince (1986, 1993a, 1995), the disyllabicity restriction found in Semitic morphology reflects a universal preference for a binary foot (see also CHAPTER 40: THE FOOT; CHAPTER 44: THE IAMBIC–TROCHAIC LAW). Thus, the most general template of Arabic verb is a binary syllabic foot. The advantage of the syllabic template has been supported with data from Modern Hebrew (McCarthy 1984), where verbs from the same class have different CV templates but an identical syllabic template, consisting of two syllables, i.e. a foot (see, however, Amharic templatic plasticity in (2)).

5 Note that also the theory of Government Phonology refers to syllables, but the only possible syllable is CV (Lowenstamm 1996), in which case reference to syllables seems to be redundant. Within this theory, Arabic template consists of CV-CVCVCV, where the initial CV is the derivational head.

Table 108.2 CV templates and verb configurations in Standard Arabic[a]

	CV template				Configuration		Verb	
a.	1.		CV	CVC[b]	Ia	CaCaC	'fataħ	'to open'
					Ib	CaCiC	'ħasib	'to think'
					Ic	CaCuC	'qabuħ	'to be ugly'
	2.		CVC	CVC	II	CaC$_i$C$_i$aC	'kassar	'to break'
						CaCCaC	vʃamʕal	'to scatter'
					IV	ʔaCCaC	'ʔakram	'to honor'
	3.		CVV	CVC	III	Ca:CaC	'sa:baq	'to run a race'
b.	1.	C	CV	CVC	VII	nCaCaC	'nbasatˤ	'to be pleased'
					VIII	CtaCaC	'qtabal	'to receive'
					IX	CCaC$_i$aC$_i$	'ħwalal	'to squint'
	2.	C	CVC	CVC	X	staCCaC	'staʕmal	'to use'
					XII	CCawC$_i$aC	'hdawdab	'to be curved'
					XIII	CCawwaC	'ʕlawwad	'to be heavy'
					XIV	CCanC$_i$aC$_i$	'ʕʃanʤaʤ	'to go quickly'
						CCanCaC	'branʃaq	'to bloom'
					VI	CCanCay	'ʕlandaj	'to be strong'
	3.	C	CVV	CVC	XI	CCa:C$_i$aC$_i$	'swa:dad	'to be black'
c.	2.	CV	CVC	CVC	V	taCaC$_i$C$_i$aC	ta'farraq	'to be dispersed'
						taCaCCaC	ta'ʕafrat	'to act like a devil'
	3.	CV	CVV	CVC	VI	taCa:CaC	ta'ka:lam	'to converse'

[a] As Watson (2002) notes, most dialects of Arabic do not retain configurations above X. In addition, merger in the prosodic template (but not the configuration) is found in several dialects, including Palestinian Arabic, where (c) merged with (b) via the deletion of the vowel in the first syllable. That is, Standard Arabic [ta'farraq] (c2) corresponds to Palestinian Arabic ['tfarraq] (b2).
[b] The CVCVC template is considered one configuration (due to a unified inflectional paradigm), although it has several sub-configurations that differ in the quality of the second vowel (see (4) above).

(13) *Modern Hebrew verb configurations*

Syllabic configuration		*CV configuration*		
I	$\sigma_a\ \sigma_a$	CaCaC	pa'tax	'to open'
II	$^{n}\sigma_i\ \sigma_a$	niCCaC	nix'nas	'to enter'
III	$^{h}\sigma_i\ \sigma_i$	hiCCiC	hig'dil	'to enlarge'
		hiCCCiC	hiʃ'prits	'to squirt'
IV	$\sigma_i\ \sigma_e$	CiCeC	gi'del	'to raise'
		CiCCeC	tir'gem	'to translate'
		CCiCCCeC	trins'fer	'to transfer'
V	$^{hit}\sigma_a\ \sigma_e$	hitCaCeC	hitla'beʃ	'to get dressed'
		hitCaCCeC	hitbar'gen	'to become a bourgeois'

In Arabic, unlike in Hebrew, syllable structure plays a major role in distinguishing among configurations, as is evident from the distinction among CaCaC (I), CaCCaC (II), and Ca:CVC (III). The first configuration is distinguished from

Empirical support for this view is provided by the transfer phenomena in broken plurals (§2.1.2), where properties of the singular form that cannot be encoded in the consonants are transferred to the plural form. Arabic broken plurals (17a) exhibit vowel-quantity transfer in trisyllabic plurals: a short vowel in the plural corresponds to a short vowel in the singular and a long vowel in the plural corresponds to a long vowel in the singular (McCarthy and Prince 1990; McCarthy 2000). Broken plurals in Tigre (17b), as well as in Tigrinya (Palmer 1955), exhibit vowel-quality transfer in trisyllabic plurals: a high front vowel in the plural corresponds to a front vowel in the singular, a high back vowel in the plural corresponds to a back vowel in the singular, and a central vowel in the plural corresponds to a central vowel in the singular (Palmer 1962).

(17) *Vowel quantity transfer*

a.	*Yemeni Arabic* (Qafisheh 1992)			
		singular	*plural*	
	short vowel	ˈdarzan	daˈraːzin	‘dozen’
		ˈmaktab	maˈkaːtib	‘office’
	long vowel	finˈʤaːn	faˈnaːʤiːn	‘cup’
		makˈtuːb	maˈkaːtiːb	‘letter’
	Standard Arabic (Ratcliffe 1998a)			
		singular	*plural*	
	short vowel	ˈxaːtam	xaˈwaːtim	‘signet ring’
		ˈʕaqrab	ʕaˈqaːrib	‘scorpion’
	long vowel	ʤaːˈmuːs	ʤawaːˈmuːs	‘buffalo’
		mifˈtaːħ	mafaːˈtiːħ	‘key’
b.	*Tigre* (Palmer 1962)			
		singular	*plural*	
	central vowel	misgid	məsagid	‘mosque’
		ʤənʤər	ʤənaʤir	‘chain’
	front vowel	bərmil	bəramil	‘barrel’
		bistˤan	bəsatˤin	‘garden’
	back vowel	kitkut	kətakut	‘young bird’
		maskot	məsakut	‘window’

Similarly, Ratcliffe (1998a) mentions cases of vowel polarity in Arabic CVCC nouns: when the vowel in the singular is low the vowel in the plural tends to be high ([qalb] – [quˈluːb] ‘heart(s)’), and when the vowel in the singular is high the vowel in plural tends to be low ([qufl] – [ʔaqˈfaːl] ‘lock(s)’).

Other cases of transfer, which cannot be attributed to the consonantal root or the configuration, are drawn from the formation of Hebrew denominative verbs (Bolozky 1978; Bat-El 1989, 1994a, 2003a, 2003b; Gafos 1998; Ussishkin 1999, 2000). A verb derived from a noun with an affix may include the affix consonant as part of its stem (e.g. [merˈkaz] ‘center’ → [mirˈkez] ‘to center’; cf. [riˈkez] ‘to concentrate (INTRANS)’; [parˈʃan] → [pirˈʃan] ‘to commentate’; cf. [peˈreʃ] ‘to interpret’). Also, when the base consists of consonant clusters, these clusters are preserved in the derived verb (e.g. [guʃˈpanka] ‘approval’ → [giʃˈpenk] ‘to approve’; *[gʃiˈpenk], *[giʃˈpnek]).

in the formation of [ħu'ru:b] 'wars' from [ħarb], the circumscribed material <ħar> is mapped into the template $[\sigma_\mu\sigma_{\mu\mu}]_F$, yielding $[\hbar araa]_F$. Melodic overwriting assigns the appropriate vocalic pattern of the plural form and the residual base segments ([b] in this case) are adjoined. In longer singular forms, however, the residue is an entire syllable, in which case the basic iambic template is expanded. This is the case in ['madxal] – [ma'da:xil] 'entrance(s)', where the material within the iambic template is $[\text{mada:}]_F$, while [xil] forms an additional syllable. As the final syllable is outside the template, it preserves its original structure from the singular form, reflecting faithfulness to the base (McCarthy 2000). This explains the transfer phenomena in (18) above, which appear, as predicted, only in the final syllable.

5 Mapping a configuration: A constraint-based approach

As noted in §3.2.3, the prosodic structure of the configuration is assigned by the independently motivated constraints in (14): FtBin, which sets the lower bound at two syllables, and Align(σ, PrWd), which sets the upper bound at two syllables. The segmental elements of the configuration, i.e. the vocalic pattern and affix (if any) are considered an affix, and thus provided in the input. The position of the affix within the word is determined by independent universal constraints.[6] The type of constraints involved depends on the assumption regarding the input, i.e. whether it is a consonantal root (§5.1) or a fully specified word/stem (§5.2). Throughout this section, I assume the effect of the templatic constraints in (14), and thus do not consider candidates that exceed the disyllabic template.

5.1 *Root + configuration*

Under the root-based approach, the input consists of a consonantal root and the segmental elements of the configuration, which are considered an affix. When the affix consists of vowels only, markedness constraints are responsible for syllable structure, and thus for the linear order of the vowels and the consonants.

(19) **Complex, FinalC,[7] Onset >> *Coda (Hebrew [xi'ber] 'to connect')*

{xbr}+{ie}	*Complex	FinalC	Onset	*Coda
a. xbier	*!			*
b. xibre		*!		*
c. ixber			*!	**
☞ d. xiber				*

[6] Within the standard OT approach to morphology, affixes, just like bases, are introduced as lexical items in the input, as well as in alignment constraints. Russell (1995, 1999) eliminates this duality, arguing that affixes should be introduced as constraints only (see also Yip 1998; Adam and Bat-El 2008). An analysis of Hebrew configurations within this approach is provided in Bat-El (2003b).

[7] FinalC (McCarthy 1993), formally stated as AlignR(PrWd, C), requires a word to end in a consonant.

In general, whether the root consists of three consonants or four (e.g. [tir'gem] 'to translate'), or whether the affix consists of vowels only or vowels plus a prefix (e.g. [hig'dil] 'to enlarge'), the templatic constraints in (14) determine the prosodic template and the markedness constraints in (19) are responsible for the sequential order of the segments.

Note that the root consonants (with the exception of glides) must be surface-true, which can be attributed to an undominated constraint MAX(Root). Amharic and Hebrew both respect this constraint, but they differ with regard to the violability of other constraints. As exemplified in (2), Amharic verbs can be trisyllabic in order to accommodate all root consonants without violating *COMPLEX (e.g. [məsəkkər-ə] 'to testify'). Therefore, the templatic constraints are violated. Hebrew verbs always respect the templatic constraints, but *COMPLEX can be violated (e.g. [trins'fer] 'to transfer') in order to accommodate all consonants.

5.2 *Word + configuration*

Under the word-based approach, the input consists of a base word or stem and the segmental elements of the configuration (i.e. vocalic pattern and affix). The constraint ranking FAITHAFFIX >> FAITHSTEM (Ussishkin 2000) is responsible for melodic overwriting (18b), ensuring that all affix segments are surface-true. As in the root-based approach (§5.1), markedness constraints provide the syllable structure.

However, the markedness constraints derive the correct output only when the input consists of CV and CVC syllables (e.g. Hebrew [ga'dal] 'to grow' → [gi'del] 'to raise', ['telefon] 'phone' → [til'fen] 'to phone'). Hebrew denominative verbs derived from bases with more than four consonants violate the markedness constraints. For example, [prig'res] 'to progress' has an initial complex onset and [kim'pleks] 'to make complex' a medial complex onset, while [ib'strekt] 'to make abstract' has no initial onset at all, although a glottal stop is inserted in phrase-initial position. The crucial observation is that consonants adjacent in the base are also adjacent in the derived verb. For such cases, it is necessary to adhere to the faithfulness constraint CONTIGUITY, which preserves adjacency between consonants in the input. Of course, the phenomenon justifying CONTIGUITY cannot be accounted for within the root-based approach.

(20) [ab'strakti] 'abstract' → [ib'strekt] 'to make abstract'

abstrakti+{i e}	CONTIGUITY	ONSET
a. bistrekt	*! (b s)	
☞ b. ibstrekt		*

In Arabic verbs, CONTIGUITY is low-ranked, as seen in denominative verbs such as in [baql] 'herbage' → ['ʔabqal] 'to produce herbage', [sˤajf] 'summer' → ['tsˤajjaf] 'to go on a summer vacation' (cf. Hebrew [faks] 'fax' → [fik'ses] 'to send a fax'), [frans] 'France' → ['tfarnas] 'become French, act li e a Frenchman' (cf. Hebrew [flirt] 'flirt' → [flir'tet] 'to flirt').[8]

As for configurations with affixes, Arabic is challenging, as it employs both prefixes and infixes. Arabic configurations display two types of contrast: contrast

[8] Note that Arabic loan nouns do preserve the source clusters, as in [blastik] 'plastic' and [ʔekspres] 'express' (Thornburg 1980).

in the prosodic template – medial *vs.* initial CC (ʔaCCaC (IV) *vs.* nCaCaC (VII)) – and contrast in affixation: prefix *vs.* infix (nCaCaC (VII) *vs.* CtaCaC (VIII)). All these configurations respect the templatic constraints in (14), which set the minimal and maximal bound of two syllables. According to Wright (1962), verbs in these configurations are all derived from the basic configuration CaCaC (1).

Regardless of their position in the word, the affixes are attached via an alignment constraint ALIGN-L(Aff, PrWd), which requires the affix to be aligned with the left edge of the prosodic word (McCarthy and Prince 1993b). Each affix has its alignment constraint, competing with ALIGN-L(Stem, PrWd), which requires the left edge of the stem to align with the left edge of the prosodic word. The ranking ALIGN-L(Aff, PrWd) >> ALIGN-L(Stem, PrWd) yields a prefix, and the opposite ranking yields an infix.

(21) *Arabic prefix* [ʔ, n] vs. *infix* [t]

ALIGN-L(ʔ, PrWd), ALIGN-L(*n*, PrWd) >> ALIGN-L(Stem, PrWd) >>
ALIGN-L-(*t*, PrWd)

Assuming gradient alignment, where the further the relevant element is from the edge the more violation marks it gets, the constraints above account for nCaCaC and CtaCaC only; ʔaCCaC is worse than *ʔCaCaC with respect to ALIGN-L(Stem, PrWd), and thus predicted not to be selected as the optimal candidate. However, with the constraint *ʔC, which prohibits a glottal followed by a consonant, all the three configurations are derived.[9]

(22) a. nCaCaC

{n}, {CaCaC}	*ʔC	ALIGN-L (ʔ, PrWd)	ALIGN-L (*n*, PrWd)	ALIGN-L (Stem, PrWd)	ALIGN-L (*t*, PrWd)
a. naCCaC				**!	
☞ b. nCaCaC				*	
c. CanCaC			*!*		
d. CnaCaC			*!		

b. CtaCaC

{t}, {CaCaC}	*ʔC	ALIGN-L (ʔ, PrWd)	ALIGN-L (*n*, PrWd)	ALIGN-L (Stem, PrWd)	ALIGN-L (*t*, PrWd)
a. taCCaC				*!*	
b. tCaCaC				*	
c. CatCaC					**!
☞ d. CtaCaC					*

[9] A glottal stop is perceptually weak, and a preconsonantal position is also weak, and thus this constraint is highly motivated. Other effects of this constraint can be seen in the deletion of stem-initial ʔ (plus compensatory lengthening) in /ʔaʔθar / → [ˈʔaːθar] 'to prefer' (cf. [ˈakram] 'to honor') and /ʔuʔmul/ → [ˈʔuːmul] 'work!' ([ˈʔuktub] 'write!') (Wright 1962).

c. ʔaCCaC

{ʔ}, {CaCaC}	*ʔC	Align-L (ʔ, PrWd)	Align-L (n, PrWd)	Align-L (Stem, PrWd)	Align-L (t, PrWd)
☞ a. ʔaCCaC				**	
b. ʔCaCaC	*!			*	
c. CaʔCaC		*!*			
d. CʔaCaC		*!			

6 Concluding remarks

At the very beginning of her 1982 monograph *The syntax of words*, Selkirk notes the delimitations of her model: "The W-syntax of the Semitic languages, then, includes two components, only one of which is of the sort I am attempting to characterize here" (1982: 2–3). The two components are affixation and configuration, where the latter one is excluded from Selkirk's model.

The system of configurations is characteristic of Semitic morphology and is often considered unique, and therefore excluded from various theoretic models. However, a universal perspective of the system of configurations proves the contrary (Bat-El 2003a). A configuration consists of a prosodic template and vocalic pattern, and thus relation among words exhibit alternations in these structural properties. With reference to the vocalic pattern, the structural relation between [pi'ter] 'to dismiss' and [pu'tar] 'to be dismissed' in Hebrew is like that between *sing* and *sang* in English. With reference to the prosodic structure, the structural relation between [ni-mi'ʃax] 'to last (PAST)' and [ji-ma'ʃex] 'to last (FUT)' in Hebrew is like that between [hiwt-inaj] 'to walk-GERUND' and [hiwiit-al] 'to walk-DUB' in Yawelmani (Kisseberth 1969; Archangeli 1984). Moreover, broken plurals are found also in Hausa, a non-Semitic language (Rosenthall 1999).

That Semitic languages seem to be unique is due to combination and pervasiveness (Bat-El 2003a). While English exhibits only ablaut and Yawelmani exhibits only prosodic alternation, Semitic languages combine both within the same paradigms (combination). While English exhibits ablaut in a subclass of paradigms and Yawelmani exhibits prosodic alternation in stems associated with specific suffixes (template suffixes), Semitic languages exhibit ablaut and prosodic alternation in most paradigms (pervasiveness). See, however, Maltese in §2.1.1.

Viewing the configuration as a combination of independent structures allows the analysis of Semitic morphology within the same theoretical models proposed for other natural languages.

ACKNOWLEDGMENTS

I would like to thank Anbessa Teferra and Tareq Muhalwas for their help with the Amharic and Arabic data, respectively. I appreciate the comments on this chapter provided by Lior Laks, the editors Keren Rice and Marc van Oostendorp, and two anonymous reviewers. All disclaimers apply.

Ussishkin, Adam. 1999. The inadequacy of the consonantal root: Modern Hebrew denominal verbs and output–output correspondence. *Phonology* 16. 401–442.

Ussishkin, Adam. 2000. The emergence of fixed prosody. Ph.D. dissertation, University of California, Santa Cruz.

Ussishkin, Adam & Andrew Wedel. 2003. Gestural motor programs and the nature of phonotactic restrictions: Evidence from loanword phonology. *Proceedings of the West Coast Conference on Formal Linguistics* 22. 505–518.

Watson, Janet C. E. 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Watson, Janet C. E. 2006. Arabic morphology: Diminutive verbs and diminutive nouns in San'ani Arabic. *Morphology* 16. 189–204.

Wright, William. 1962. *A grammar of the Arabic language*. 1st edn. 1859. Cambridge: Cambridge University Press.

Yip, Moira. 1988. Template morphology and the direction of association. *Natural Language and Linguistic Theory* 6. 551–577.

Yip, Moira. 1998. Identity avoidance in phonology and morphology. In Steven Lapointe, Diane Brentari & Patrick Farrell (eds.) *Morphology and its relation to phonology and syntax*, 216–246. Stanford: CSLI.

Zadok, Gila. 2002. Abbreviations: A unified analysis of acronym words, clippings, clipped compounds, and hypocoristics. M.A. thesis, Tel-Aviv University.

109 Polish Syllable Structure

Christina Y. Bethin

The syllable as a phonological prosodic constituent serves to organize segment sequences in terms of sonority, associating sonority peaks/syllabic segments with syllable nuclei and non-syllabic segments with syllable margins as onsets and codas (Blevins 1995; Zec 2007). The sequencing of segments within onsets and codas is also said to be governed by sonority, in that sonority rises toward the syllable peak and falls after it (Sievers 1881; Jespersen 1904); this has been formalized in the generative phonological literature as a sonority sequencing principle/generalization (Selkirk 1982, 1984; Clements 1990; chapter 49: sonority). Because syllable structure is not contrastive, it is most likely created by syllabification processes as part of a phonological derivation or selected by a set of syllable structure constraints in an output-oriented framework. In either case, certain organizational principles hold on segment sequencing within a syllable, and the question is how these properties of syllables are implemented. Polish has contributed to discussions on syllable structure primarily by virtue of its extraordinarily complex consonant clusters, which appear to violate the sonority sequencing generalization. It also has a set of nasal vowels that bear on the structure of the syllable nucleus, and vowel–zero alternations that interact with syllabification.

In derivational approaches where syllables are constructed on the basis of the segmental string in some principled way, surface sonority sequencing violations have raised questions about the structure of syllables, the existence of unsyllabified (extrametrical) segments and their incorporation into prosodic levels, and about syllabification and resyllabification at various derivational levels. In constraint-based grammars such as Optimality Theory (Prince and Smolensky 2004), the complexity of the Polish syllable puts it in the category of languages in which faithfulness constraints on segment deletion and insertion appear to be ranked above markedness constraints on syllable margins, so one question is whether and when syllable structure constraints have visible effects on output forms. In non-derivational Government Phonology or CV Phonology, the syllable level is independent of the melodic or segmental tier, and consists of a series of onsets and rhymes/nuclei, some of which may be empty or have unassociated melodies, so consonant clusters that violate sonority sequencing in pronunciation do not violate it in the phonology when they are treated as onsets followed by empty nuclei. The fact that syllable structure seems to be implicated in certain aspects of Polish morphology presents a challenge to all of the above, and to any model of phonology–morphology interaction.

The chapter is organized as follows: §1 deals with the syllable margins: sonority sequencing, syllabification, and extrasyllabicity (§1.1), resyllabification (§1.2), and the phonological *vs.* the phonetic syllable (§1.3). §2 concerns the syllable nucleus, primarily the description of nasal vowels as potential complex nuclei (§2.1) and the treatment of glides and high vowels (§2.2). Because it is unusual to find that a language which tolerates extraordinarily complex sequences of consonants sometimes does not, as seems to be the case in some aspects of Polish allomorphy, there is a discussion of syllable structure and morphology in §3. Concluding remarks are in §4.

1 The syllable margins

Polish has complex clusters of various types both in onset and coda position within the syllable, exemplified by words such as *rtęć* [rt] 'mercury', *teatr* [tr] 'theater', *ptak* [pt] 'bird', *szept* [pt] 'whisper', *bzdura* [bzd] 'nonsense', *wzbronić* [vzbr] 'to forbid', *siostr* [str] 'sister (GEN PL)', *z pstrągiem* [spstr] 'with a trout', *przestępstw* [mpstf] 'transgression (GEN PL)'.[1] Complex consonant clusters in Polish are in large part the result of historical change: (i) the loss of short high vowels (yers) in weak position (word-final or in the syllable before one with a non-yer or strong yer vowel) as in Common Slavic *mĭgla > *mgła* 'mist' (see also CHAPTER 122: SLAVIC YERS); (ii) the obstruentization of palatalized /rʲ/ and /w/ to /ʒ/ and /v/, respectively; and (iii) the decomposition of nasal vowels into sequences of vowel plus nasal consonant, as in *prěstǫpĭstvŭ > [pʃestempstf] 'transgression (GEN PL)'. Prefixation by /z/ (/s/), /v/ (/f/), and /vz/ (/fs/) contributes to the large number of complex clusters across morphemes (e.g. *skrwawić* [s+krf-] 'to inflict bloody wounds'), and word-medial clusters are enhanced by prepositions and prefixes such as *ob-*, *-od-*, *przed-*, *pod-*, etc. Loanwords further extend the possibilities in codas, e.g. *tekst* 'text', *filtr* 'filter', *asumpt* 'inducement'.

1.1 *Sonority sequencing and syllabification*

A fundamental question is whether Polish consonant clusters are organized in any particular way in the syllable, and specifically, whether syllable onsets and codas obey the principle of rising sonority toward the syllable peak and decreasing sonority after the syllable peak. If allowance is made for *s/z*+C and *f/v*+C sequences, then most Polish clusters in onsets and codas do exhibit sonority sequencing: *strach* 'fear', *dno* 'bottom', *kart* 'cards (GEN)', *sens* 'sense', *kar-ty* 'cards', *sios-tra* 'sister'. But sonority violations in both onsets (1) and codas (2) and in word-medial clusters (3), both morpheme-internally as well as across morphemes, have

[1] A complete typology of Polish consonant clusters is given in Bargiełówna (1950), where morpheme boundaries within clusters and clusters of foreign origin are noted. Sawicka (1974: 67–75) provides an inventory of word-initial and word-final clusters (ignoring voicing assimilation and final devoicing); see also Rowicka (1999a: 309–344). In Mikoś (1977) and Rochoń (2000: 283–301) palatalized labials decomposed into labial + /j/ sequences are counted as onset clusters, and final devoicing is represented in codas.

(3) *Sonority sequencing violations in medial clusters/onsets/codas*

/ʑdʑʨ/	*najeźdźcze*	'invasive'
/bwk/	*jabłko*	'apple'
/brn/	*srebrny*	'silver'
/snk/	*czosnku*	'garlic (GEN SG)'
/trk/	*wiatrka*	'wind (GEN SG)'
/ndrk/	*mędrkować*	'to act like a smart alec'

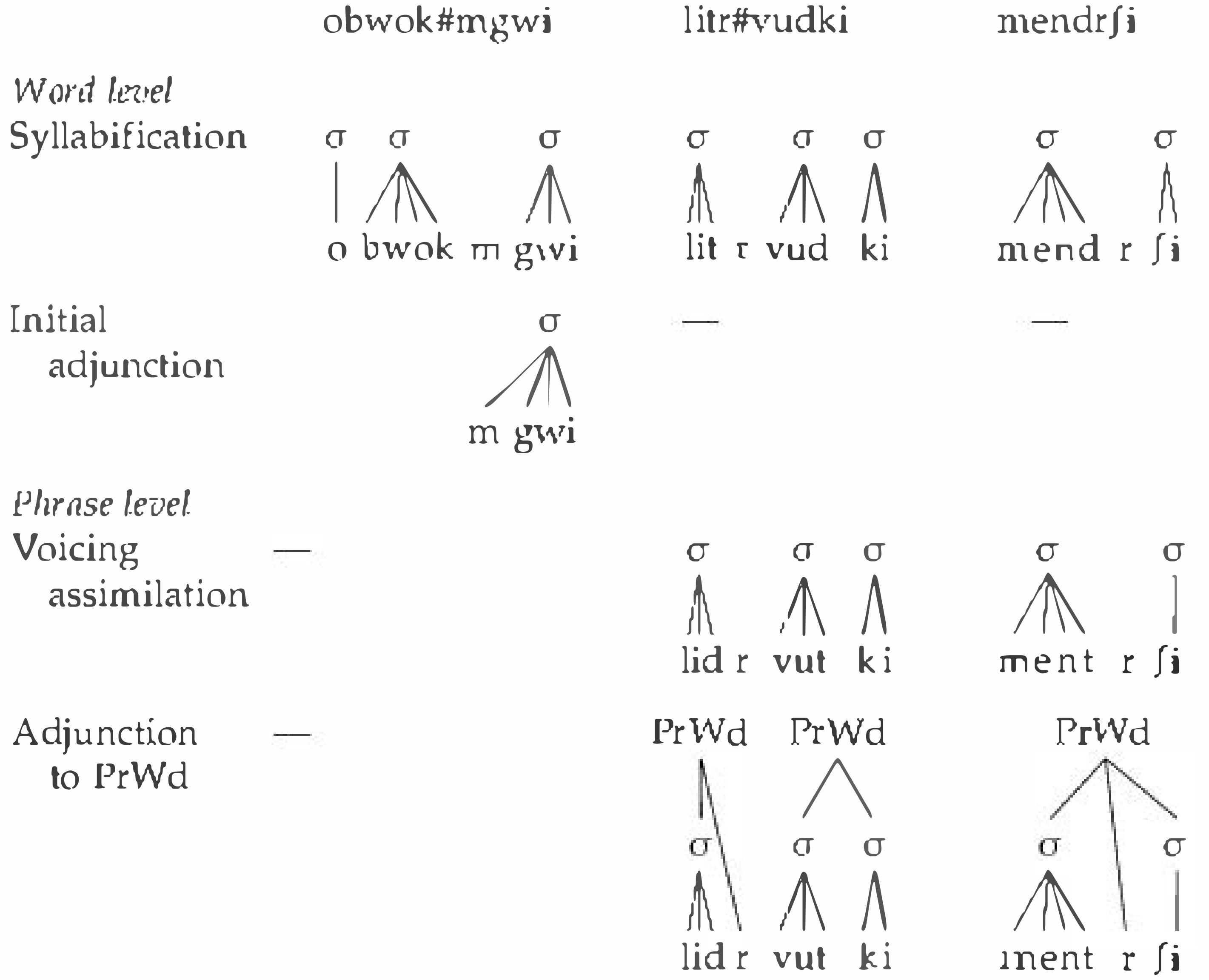
It was Kurylowicz (1952) who first related the typology of consonant clusters to syllable structure by proposing a type of syllable template with branching onsets and codas and a classification of sounds as belonging either to groups or elements. In addition to onset groups with *s/z* followed by an obstruent and sonorant (e.g. *str, skr, zdr*), he identified sets of two-member groups consisting of obstruent + sonorant, *s/z* + obstruent, and *s/z* + sonorant. (For historical reasons and in their alternations /v/ (/f/) and /ʒ/ (/ʃ/) sometimes pattern with sonorants if they derive from *w and *rʲ, and they are often treated as sonorants in synchronic phonology [Rubach 1984; Gladney 2004; Gussmann 2007: 208–210, 307–312].) Syllable onsets are said to be maximally binary branching, with terminal nodes which can be filled by a combination of groups and/or elements, e.g. /v-str/ consists of an element followed by a group, /skr-f/ is a group + an element. Coda clusters are taken to be the mirror image (sonorant + obstruent (RC)) of onset (obstruent + sonorant (CR)) types, with the additional important proviso that onset groups (CR) may also occur in codas. Thus, codas may contain two elements (*k-t*), a coda group (*rb*), a coda group + element (*rt-f*), an element followed by an onset group (*m-st*), a coda group followed by an onset group (*mp-sk*), or simply an onset group (*dr*), as in *kadr* 'frame (in film)'. This approach does not distinguish among consonant types beyond obstruent and sonorant and thus implies that there are no systematic gaps in onset and coda sequences. Furthermore, morphology and phonology do not necessarily coincide: *zmrozić* 'to freeze' [z+mroʑitɕ] is said to have an onset of /ʑm-r/.

The basic insights of Kuryłowicz (1952) have been developed in versions of Government Phonology or CV phonology, where phonetic consonant clusters are said to contain empty nuclei in the phonological representation and thus are not really complex clusters (Gussmann 1991, 1992, 2007: 180–247; Cyran and Gussmann 1999; Rowicka 1999a; Cyran 2003: 161–321; Scheer 2004; and elsewhere). For example, the onset in *krnąbrny* 'unruly, restive' is said to have a branching onset (O = *kr*) before an empty nucleus (N = Ø), followed by a simplex onset (O = *n*) (Gussmann 2007: 215), and there are no sonority violations to speak of within onsets as such. Phonetic word-final codas are said to be onsets before an empty nucleus: *tkwić* 'to pierce, stick' [tkfʲitɕ] is /t-N-kw-i-tɕ-N/. Government Phonology does predict certain gaps in cluster types due to governing relations, but as there is no issue of syllabification, this approach will not be discussed in detail here.

In derivational generative phonology syllable structure is built up by a series of ordered syllabification rules, first constructing the syllable nucleus and then building the syllable onset (Steriade 1982; Clements and Keyser 1983; Levin 1985; and others). The ordering of the onset rule before coda formation is a way of accounting for the well-known asymmetry that no language prohibits onsets or requires codas (in Optimality Theory this is implicit in the formulation of the constraints

As these differences are fairly robust, a distinction has to be made between initial and final/medial sonority reversals. The R in #RC sequences is prosodified by initial adjunction, either to the phonological word (Rubach and Booij 1990b), or to the initial syllable as a complex onset (Bethin 1992: 41, 171–174; Gussmann 1992). The prosodification of #RC sequences takes place early in the derivation, at the word or lexical level, while that of sonorants in CR# or CRC contexts is done by adjunction to the phonological word at the phrase or post-lexical level. The differences with respect to transparency in voice assimilation are said to be due to the interaction of prosody and voicing rules, with voicing processes applying to syllabified or prosodified segments in word-initial position but before adjunction takes place elsewhere, as exemplified in (5).

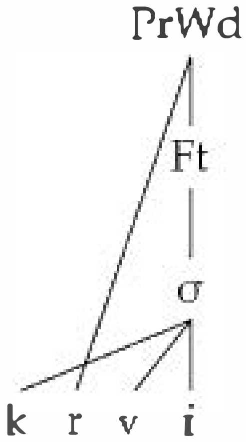
(5) *Prosodification of sonorants in* #RC, #CR, *and* CRC *sequences*



In non-derivational frameworks such as Optimality Theory, this is translated to differences in representation. Sonorants in #RC sequences are said to be associated to the syllable or foot (Rubach 1997; Rochoń 2000: 135–144), where they are specified for voice and not transparent to voicing assimilation. Final and medial sonorants are associated to the prosodic word and are thus transparent to voicing processes. Evidence supporting proposals for different prosodification of sonorants in Polish comes from variation in word medial position in how speakers syllabify the sonorant, e.g. *piosn-ka* or *pios-nka* 'song', and from the fact that sonorants in CRC and CR# clusters may sometimes be lost, e.g. *jabłko* 'apple' [japko], *mógł* 'he could' [mukw̥], [muk], but initial sonorants are not deleted.

Other problematic cases are sonorants trapped within onsets (#CRC-), because these sonorants lose voice in the environment of voiceless obstruents, e.g. *płci* 'sex (GEN SG)' [pw̥ytɕi], and they are transparent with respect to progressive devoicing of fricatives, e.g. *krwi* /krvi/ 'blood (PREP/DAT SG)' [kr̥fi]; *trwać* /trvatɕ/ 'to last' [tr̥fatɕ] (Booij and Rubach 1987; Bethin 1992: 175–178; Gussmann 1992, 2007: 311–312; Rubach 1997; Rochoń 2000: 143; on the relation between sonorants and voice, see also CHAPTER 8: SONORANTS). If the initial obstruent is syllabified as an onset and if sonorants which are transparent to voicing processes are not syllabified but adjoined at a higher prosodic level, then the adjunction of trapped sonorants to the Prosodic Word not only violates the Strict Layer Hypotheses (prosodic structure is said to be strictly hierarchical in organization, cf. Selkirk 1978 and CHAPTER 51: THE PHONOLOGICAL WORD), as do all cases of adjunction above, but here adjunction to the PrWd also forces the crossing of association lines, as shown in /krvi/ (6).

(6)



This case has been a challenge and it served as an argument in support of Derivational Optimality Theory (DOT) in Rubach (1997), where the SSG dominates PARSESEGMENT (segments must be associated to prosodic structure) at one level, leaving the /r/ not incorporated into the syllable onset, and Progressive Voicing selects the [k<r>fi] candidate over [k<r>vi]. This output then becomes the input to level two, where constraint ranking may be different; here the ranking of PARSESEGMENT over SSG chooses [krfi] over [k<r>fi], with the sonorant incorporated into the syllable and pronounced. In Gussmann (1992) this case served as an argument for postulating delinking from voice specification and resyllabification rules in order to meet well-formedness conditions on syllable structure. A different view of so-called trapped sonorants and syllabic consonants is given in Rowicka (1999a: 258–265), and there is an extensive discussion of these and related problems in Scheer (2004: 165–193).

The connection of voicing processes in Polish to prosodic structure (Bethin 1984) was further explored with respect to the licensing of voice only in syllable onset position and the assignment of sonorant voice by default only to prosodified sonorants (Rubach and Booij 1990b; Bethin 1992; Gussmann 1992; Lombardi 1995a, 1995b). Rubach (1996) argues against a syllabic analysis of voice assimilation and for binary specification of voice with no reference to syllable structure, though he retains the notion that sonorant voicing is subject to prosodification, which in Rubach (1997: 562) is formulated as SONORANT DEFAULT: all and only syllabified sonorants are [+voice]. Given that there is variation in syllabification in actual implementation where voiced segments are parsed into codas (Szpyra-Kozlowska 1998), voicing in Polish is probably segmentally conditioned and not a syllable-based process (Steriade 1997; Blevins 2003).

1.2 Resyllabification

When syllable structure is built up by a series of syllable-building rules in a derivational framework, it is in principle possible for these rules to interact with other phonological rules of the grammar, so there has been some discussion about whether syllable structure rules are ordered with respect to other rules; whether they reapply if the application of an earlier rule disturbs syllable structure, when they are triggered to reapply; and what happens to segments that lose their association to syllable structure in the course of a derivation. In non-derivational frameworks it is a question of what determines whether a potential syllable nucleus becomes activated and how this relates to prosodic well-formedness.

Polish exhibits a well-known set of vowel–zero alternations that bear on syllable structure in various ways. If the vowel is said to be present in the lexical representation and deleted under certain conditions, then one has to assess its viability as a syllable peak and potentially the loss of a syllable nucleus in vowel deletion. If the vowel is for some reason exempt from syllabification, what are its properties and what happens to adjacent consonants in the absence of a potential syllable nucleus? If the vowel gains prosodic status during a derivation, what are the consequences for syllable structure? If, on the other hand, the alternation is taken to be the result of epenthesis under certain conditions, what properties of syllable structure trigger epenthesis?

The vowel–zero alternations are found in inflection and derivation in roots, stems, prefixes, and prepositions, and primarily involve the vowel /e/ (which sometimes alternates with /o/), and also [i] or [ɨ] in derived imperfectives or adjectives, as shown in (7).

(7) *Vowel–zero alternations* ([e ɨ] indicate alternating vowels)

kr[e]*w* (NOM SG)	*krwi* (GEN SG)	'blood'
ł[e]*b* (NOM SG)	*łba* (GEN SG)	'forehead'
wiad[e]*r* (GEN PL)	*wiadro* (NOM SG)	'pail'
świat[e]*ł* (GEN PL)	*światło* (NOM SG)	'light'
des[e]*k* (GEN PL)	*deska* (NOM SG)	'board'
cuki[e]*r* (NOM SG)	*cukru* (GEN SG)	'sugar'
cukier[e]*k* (NOM SG)	*cukierka* (GEN SG)	'candy'
cukierecz[e]*k* (NOM SG)	*cukiereczka* (GEN SG)	'candy (DIM)'
wini[e]*n* (MASC SG)	*winna* (FEM SG)	'guilty'
umrzeć (PERF)	*umi*[e]*rać* (IMPERF)	'to die'
pod[e]*brać* (PERF)	*podbi*[e]*rać* (IMPERF)	'to select'
od[e]*pchnąć* (PERF)	*odp*[ɨ]*chać* (IMPERF)	'to push away'
lekcja (N)	*lekc*[ɨ]*jny* (ADJ)	'lecture'

The alternating vowels are known as yers (to distinguish them from the historical "jers") since many derive from the historical change lowering (or vocalizing) or deleting the Common Slavic high lax vowels ("jers") in strong and weak position, respectively (Lightner 1963, 1972; for an overview, see Rochoń 2000: 46–86). The alternation has been generalized to forms in which it is not etymological and to loanwords, e.g. *sweter* 'sweater', *swetra* (GEN SG). (For more detail about yers in Slavic, see CHAPTER 122: SLAVIC YERS.)

Earlier proposals that the vowel–zero alternation is the result of vowel epenthesis (CHAPTER 67: VOWEL EPENTHESIS), primarily motivated by the need to syllabify non-syllabic roots or complex consonant clusters (Czaykowska-Higgins 1988; Gorecka 1988; Piotrowski 1992), faced problems. The presence of vowel epenthesis, as in *kalka* 'carbon paper' ~ *kalek* (GEN PL) and *serek* 'cheese (NOM SG, DIM)' ~ *serka* (GEN SG), in an otherwise fine coda sequence (*lk* or *rk*), the lack of epenthesis in many clusters violating sonority sequencing, e.g. *jabłko* 'apple', *krnąbrny* 'restive, unruly', and the existence of many minimal or near-minimal pairs, e.g. *miotła* 'broom', *miotel* (GEN PL), but *zamiotł* 'he swept', *barka* 'barge', *barek* (GEN PL), and *bark* 'shoulder', *barka* (GEN SG), *oset* 'thistle', *osty* (NOM PL), and *most* 'bridge', *mosty* (NOM PL), made the prediction of epenthesis questionable. The vowel in some form is present in the lexical representation.

Most analyses postulate some type of underlying yer vowels which are distinguished from non-alternating vowels, either by a diacritic/feature or by the (partial) absence of feature specification, by no association to the skeletal tier (floating matrices) or the mora (mora-less vowels), by being an empty position on the skeletal tier, or by being a special type of empty nucleus in the lexicon (Lightner 1963, 1972; Laskowski 1975; Gussmann 1980a, 2007; Rubach 1984, 1986; Spencer 1986; Szpyra 1989, 1992a, 1995: 90–135; Rubach and Booij 1990a, 1990b; Bethin 1992; Zoll 1993; Yearley 1995; Scheer 1997, 2004; Rowicka 1999a; Cyran 2003: 161–202). In this view, the presence and location of the alternation are given and it appears as a full vowel under certain set conditions.

Analyses that postulate underlying yers with no representation on a level at which they could be prosodified (no skeletal slot, no mora) incorporate procedures for the full vocalization of yers at some point in the derivation when they become potential syllable nuclei. The vocalization and deletion of yers is usually controlled by the following vowel: if it is a yer, the preceding yer will vocalize; if it is a non-yer vowel, the preceding yer will delete. Typically, this alternation is found in the last syllable of a stem. In generative analyses it is implemented by a rule of Lower (Lightner 1972; Gussmann 1980a: 39–40) vocalizing a yer before a yer in the following syllable and deleting yers elsewhere; this requires the postulation of yers (represented here as V) at the end of words that never vocalize in that position: /desVkV/ 'board (GEN PL)' becomes [desek].

It has long been known that yer behavior in derivational morphology is different from that in inflectional morphology (Isačenko 1970), in that consecutive yers all tend to vocalize except for the last one in derivation: *cukier* 'sugar' ~ *cukru* (GEN SG), *cukierek* 'candy' ~ *cukierka* (GEN SG), *cukiereczek* 'candy (DIM)' ~ *cukiereczka* (GEN SG).[3] Here recourse to the phonological cycle was proposed as one solution; this raises questions about when the syllable-building rules apply and how much of the original syllable structure, if any, is preserved on subsequent cycles when resyllabification takes place. Rubach and Booij (1990a) argue that syllabification is cyclic and continuous (Hayes 1980; Itô 1986), but because onset and nucleus

[3] These problems have been extensively discussed in the literature (Rubach 1984, 1986; Nykiel-Herbert 1985; Szpyra 1989: 200–225, 1992b, 1995: 90–135; Rubach and Booij 1990a; Rowicka 1999a: 267–298, 1999b; Scheer 1997, 2004: 167–172; Gussmann 2007: 234–245, 301–307; Kraska-Szlenk 2007: 54–100). In the Government Phonology or CVCV framework, yer vocalization is treated as association to the melody if the yer is not immediately followed by a filled nucleus (Gussmann 2007) or as the association of alternating melodies that are not subject to Proper Government (Scheer 2004), and iterativity is not an issue per se.

structure from an earlier stage is preserved at later stages, resyllabification essentially involves the reassignment of codas as onsets when a nucleus becomes available. An abbreviated derivational example is given in (8), where consonant palatalization is ignored. An underlying yer cannot be syllabified until it vocalizes; unvocalized yers are simply deleted. Codas are said to resyllabify as onsets on subsequent cycles when yer vocalization provides a potential syllable nucleus, but otherwise CV structure is preserved.

(8) *Yer lowering, resyllabification, and yer deletion in* cukierek 'candy'

[[[tsukVr] + Vk] + V]

Cycle 1
Year Lower
Syllabification σ
tsuk Vr

Cycle 2
Year Lower e
Syllabification σ σ
tsu ker Vk

Cycle 3
Year Lower e
Syllabification σ σ σ
tsu ke rek V

Postcyclic
Year Deletion Ø

[tsu-k'je-rek]

In non-derivational analyses where syllabification processes do not apply and reapply, the interaction of yers and syllable structure is implemented as constraint interaction with respect to giving yers full vowel status in some form and constraints on complex margins or prosodic alignment. These approaches rely more on syllable structure constraints, in addition to other prosodic and morphological requirements. Analyses that operate with a yer in the input representation have an advantage over purely epenthetic accounts in designating the position of the alternation. Yearley (1995) proposes that yers (in Russian) are mora-less vowels that vocalize by acquiring a mora as a result of syllable structure requirements (a prohibition on complex codas and various prosodic alignment constraints disfavoring extrasyllabic segments). Sequential yers are vocalized by recursive evaluation. Rochoń (2000: 101–108) takes syllabification to be driven by the interaction of sonority sequencing, permitting clusters of equal sonority and a set of syllable structure constraints (ONSET, *CODA, *σ[CC, *σ[CCC, *σ[CCCC, CC]σ; 2000:

123–130). Jarosz (2006) suggests that yer vocalization is controlled by constraints on syllable structure specific to codas. The idea is that principles of syllable well-formedness come into play in creating a potential syllable nucleus.

In terms of syllabification within phonological words, the prefix/preposition boundary is somewhat of an issue. Unlike inflectional and derivational boundaries, the prefix boundary is maintained with respect to syllable structure, even if this produces surface violations of the core CV syllable: the prefixes /od-/, /nad-/ in *odosobnić* 'to isolate', *nadużyć* 'to abuse', and prepositions such as /nad-/ in *nad oknem* 'above the window' are syllabified [od-o-], [nad-u], [nad-o] (Kuryłowicz 1952; Booij and Rubach 1984, 1987). This leads Booij and Rubach (1984) to conclude that prosodic words may be different from morphological words, and that syllable structure is built up within a prosodic word domain. Non-syllabic prefixes/prepositions cannot constitute syllables on their own, and are necessarily syllabified with the stem across the boundary, e.g. *znieść* {z+ɲeɕtɕ} 'take, carry down'. The problematic cases are prefixes before glide-initial stems/roots, as in {rozV+jVm+ov+ɨ} 'truce (ADJ)', syllabified across the prefix boundary as [ro-zej-mo-vɨ], and sonority reversals which permit variation: {o+mdletɕ} 'to faint' is [o-mdletɕ] or [om-dletɕ], {pʒɨ+mknoɲtɕ} 'to close, shut' is [pʒɨ-mknoɲtɕ] or [pʒɨm-knoɲtɕ], with a strong preference for the second variant, which resyllabifies across the prefix boundary (Rubach and Booij 1990b). Here adjunction plays a role. These syllabification restrictions as well as exceptional vowel–zero behavior support the strength of the prefix boundary as a special case in Polish (see also Szpyra 1989, 1992b; Rowicka 1999a: 267–298, 1999b; Gussmann 2007: 234–245; and references therein).

1.3 *Phonological* vs. *phonetic syllables*

Most phonologists agree that sonority sequencing in some form is operative in the phonological syllable of Polish, while sonority sequencing violations abound in the phonetic syllable; the latter led some linguists to reject any role for principles of sonority sequencing in Polish (Sawicka 1974, 1995; Dukiewicz 1995; and others). Phonetic syllables are not always functioning phonological ones in terms of sonority peaks: the two sonority peaks in *trwa* 'it lasts' make it phonetically disyllabic [tr-va] (Dukiewicz 1995: 76), though for purposes of penultimate stress assignment in Polish and when pronounced at normal speech rate as [tr̥fa], the word is monosyllabic. In careful speech, word-final Cw sequences may be fully voiced/syllabic, e.g. *mógł* 'he could' is [mugw], [mugu], but commonly [mukw̥], [muk] (Sawicka 1995: 76, 139; Rochoń 2000: 250–255; Gladney 2004).

The phonetic syllable may have sonority sequencing violations not admitted into the phonological syllable, yet frequency data show that Polish syllable structure tends to favor the less marked type postulated in the phonology. Complex clusters are more common in onset than in coda position both in inventory and in textual frequency. Bargiełówna (1950) lists 231 two-member onsets but only 100 two-member codas. Initial clusters have a text frequency of 86.7% and coda clusters that of 13.3% (Sawicka 1974: 117–121). Simple clusters are more common than more complex ones: Sawicka (1974: 119) notes that CC clusters constitute 88.9% of occurring clusters in texts and CCC only 9.3%, in spite of a more equal distribution in the inventory where CC clusters comprise 53.6% and CCC 42.7% of all clusters. Two-member codas are much more frequent than three-member codas, constituting 98.7% *vs.* 1.3% of tokens in written texts compared to inventory, where the ratio is 74.4% to 19.6%. There are some systematic gaps in clusters, such that /j/ occurs only adjacent to a vowel, no /rl/, /lr/ occur in onsets, and no /lr/ in coda position (Sawicka 1995: 169), and there are no sequences of three or four sonorants in general.

There are also significant differences in terms of occurring cluster types in Polish: Dukiewicz (1985) documents that consonant (C) + sonorant (R) clusters are more frequent than are RC types in word-initial position: CR = 10,423, CCR = 2025, RR = 1435, CRR = 135, CCCR = 120, CRC = 80, RC = 10, RCR = 6, CRCR = 4, CCRR = 3, CCRC = 0 per 100,000 tokens in journalistic texts. Sawicka (1974: 122–123) reports on onset cluster types in written texts: sonorant + sonorant = 7%, sonorant + fricative = 0%, sonorant + stop = 0%, fricative + sonorant = 17.6%, stop + sonorant = 22.0%, fricative + stop = 21%, stop + fricative = 16.5%, fricative + fricative = 14.2%, stop + stop = 8.0%, where the high incidence of fricative + stop and fricative + fricative onsets is in large part due to prefixes and prepositions, as is also the case for spoken data (Dunaj 1985).

The most frequent coda types are fricative + stop = 53.3%, and sonorant + stop = 26.0%. The high frequency of fricative + stop (Sawicka 1974) is due to the high textual frequency of the verb *jest* 'is'; in spoken texts 35.5% of Dunaj's (1985) fricative + stop tokens were due to *jest*. Among the sonorant + stop clusters, nasals were most common, and this is most likely the contribution of nasal vowels which are pronounced as nasal consonants before stops.

Native speaker intuitions sometimes differ from structures proposed for the phonological syllable and speakers accept some variation in syllabification at the phonetic level, especially in longer medial sequences of four or more obstruents as in [skompstfo] 'avarice' or in RCR clusters as in *pośmiertny* [rtn] 'posthumous': subject data in Rubach and Booij (1990a: 126–127) shows [lʲis-tfa] (40 tokens), [lʲist-fa] (11 tokens), but [lʲi-stfa] (8 tokens) for 'board', [skomp-stfo] (28 tokens), [skom-pstfo] (19 tokens), [skomps-tfo] (2 tokens), but no tokens of [skompst-fo] or [sko-mpstfo] for 'stinginess'.[4] Variation is found within morphemes as well as across morpheme boundaries. Szpyra-Kozlowska (1998: 70) reports that subjects accept variation between complex onsets and coda-onset syllabification even for CC and CR sequences, but sonority/syllable contact principles were respected in RC, which was overwhelmingly R-C, e.g. *ma-tka* 'mother' (9 tokens), and *mat-ka* (36 tokens), *mo-dry* 'blue' (22 tokens) and *mod-ry* (23 tokens), but *ko-tlet* 'cutlet' (6 tokens) and *kot-let* (39 tokens), also *ła-dny* 'pretty' (7 tokens) and *ład-ny* (38 tokens).[5] How variation is to be represented in phonological grammars is still under investigation.

[4] The test consisted of a list of words dictated to students who were asked to write them with syllable divisions. It is not clear how many students took this test initially, but ten students took the test three times at one-week intervals (60 tokens) to determine intra-speaker variation, which for *patrzeć* 'look' was 3:10. Intra-speaker variation in CRC sequences such as *karmnik* 'feeder' was tested with 19 subjects, and here variation was of the order of 7:19 and 9:19 (Rubach and Booij 1990a: 438). It is difficult to say how much an effect orthography and orthoepy played in the task of hyphenating written items.

[5] Szpyra-Kozlowska (1998) used 300 word stimuli of varying length and cluster types (as well as stimuli with one intervocalic consonant) dictated (twice) to the same 45 subjects (native speaker students aged 19–23) in two sessions of 150 words per session. Items included 175 CC words, 65 CCC, 25 CCCC, 5 CCCCC, and 30 with C (C = obstruents and sonorants, e.g. *warta* 'guard' and *watra* 'guard fire'), but results are given only for CC syllabification (1998: 68–69). Subjects were asked to write the words with hyphens at syllable breaks. In a follow-up experiment 20 of the subjects syllabified selected items orally, with results almost identical to the written test, so the written test was taken to be reliable.

selection takes place and that the /-ij-/ surfaces as a syllable nucleus in order to syllabify the stem-final consonant (Bethin 1987, 1992: 189–221; Szpyra 1989: 131–158; Rubach and Booij 1990, 2001). Thus 'kick!' /kop<n> + ii/ is fully syllabified as [kop-nʲij] and 'defend' /broɲ + ii/ is fully syllabified as [broɲ]. However, there are examples with RR or CR structures (*karm* 'feed!', *módl* [dl] 'pray!', *rozświetlij* and *rozświetl* 'illuminate!'), or syllabifiable sequences which still take *-ij* (*uwielbij* or *uwielb* 'adore!'). Variation in the imperative forms is exemplified by some verb stems in (11) (see also Gussmann 2007: 167–178). Final devoicing is indicated in the forms below.

(11) *Variation in the imperative*

mądrzyj	[mondʒij]	~	*mądrz*	[montʃ]	'mouth off'
naświetlij	[naɕfʲjetlʲij]	~	*naświetl*	[naɕfʲjetl̥]	'illuminate'
zaczernij	[zatʃerɲij]	~	*zaczerń*	[zatʃerɲ]	'blacken'
karmij	[karmʲij]	~	*karm*	[karm]	'feed'
wątpij	[vontpʲij]	~	*wątp*	[vontp]	'doubt'

Not all variant forms show sonority reversals in stem-final clusters and some verbs with sonority reversals occur without the syllabic suffix. One possibility is that adjunction may take place before allomorphy selection for some speakers, another is that the availability of a syllabic allomorph may make it possible to choose less marked structures with simple onsets and no complex codas, or the status of any cluster-final segment may be ambiguous and thus produce variation. To some extent, the explanation depends on the analytical framework, but the consensus in the literature is that imperative allomorphy for the most part is sensitive to syllable structure in some form, either to unsyllabifiable consonants (Rubach 1986; Rubach and Booij 1990a, 1990b, 2001; Bethin 1992: 204–218) or to the presence of underlying yers or empty nuclei (Gussmann 1980b, 2007: 178).

3.2 Comparative allomorphy

The comparative suffix is *-sz-* [ʃ] or *-ejsz-* [ejʃ], and it also appears to be sensitive to syllabification: the syllabic suffix is favored when there is an unsyllabified stem-final consonant, usually a sonorant in a CR sequence (Bethin 1987, 1992: 189–218; Rubach and Booij 1990a, 1990b), as shown in (12), where voicing assimilation and progressive devoicing are given in the phonetic representation.

(12) *Comparative allomorphy*

nowy	[novi]	*nowszy*	[nofʃɨ]	'newer'
stary	[starɨ]	*starszy*	[starʃɨ]	'older'
trudny	[trudni]	*trudniejszy*	[trudnʲejʃi]	'more difficult'
łatwy	[watfi]	*łatwiejszy*	[watfʲjejʃi]	'easier'

The extrasyllabicity of the stem-final segment is resolved when it can syllabify as an onset with the syllabic allomorph: *wierny* 'loyal' is /vʲjer<n>/ and becomes /vʲjer-nʲej-ʃi/ in the comparative in preference to the /vʲjer-<n>-ʃi/ option. Rubach and Booij (1990a) use comparative allomorphy as an argument that syllabification must take place in the cyclic component and when the cluster-final segment, usually yers, the prosodic independence of prefixes and prepositions, the representation of nasal vowels as distinct from vowel + nasal sequences, and syllable effects in morphology have stimulated work fundamental to phonological theory and they will continue to be a challenge for the field.

ACKNOWLEDGMENTS

Thanks to Marc van Oostendorp and Keren Rice for invaluable comments and suggestions, and to two anonymous referees for assistance in improving the manuscript.

REFERENCES

Bargiełówna, M. 1950. Grupy fonemów spółgłoskowych współczesnej polszczyzny kulturalnej. *Biuletyn Polskiego Towarzystwa Językoznawczego* 10. 1–25.

Benni, Tytus. 1964. *Fonetyka opisowa języka polskiego*. Wrocław: Ossolineum.

Bethin, Christina Y. 1984. Voicing assimilation in Polish. *International Journal of Slavic Linguistics and Poetics* 29. 17–32.

Bethin, Christina Y. 1987. Syllable structure and the Polish imperative desinence. *Slavic and East European Journal* 31. 76–89.

Bethin, Christina Y. 1988. Polish nasal vowels. *International Journal of Slavic Linguistics and Poetics* 38. 1–39.

Bethin, Christina Y. 1992. *Polish syllables: The role of prosody in phonology and morphology.* Columbus: Slavica.

Biedrzycki, Leszek. 1963. Fonologiczna interpretacja polskich głosek nosowych. *Biuletyn Polskiego Towarzystwa Językoznawczego* 22. 25–45.

Blevins, Juliette. 1995. The syllable in phonological theory. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 206–244. Cambridge, MA & Oxford: Blackwell.

Blevins, Juliette. 2003. The independent nature of phonotactic constraints: An alternative to syllable-based approaches. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*, 375–403. Cambridge: Cambridge University Press.

Booij, Geert & Jerzy Rubach. 1984. Morphological and prosodic domains in Lexical Phonology. *Phonology Yearbook* 1. 1–27.

Booij, Geert & Jerzy Rubach. 1987. Postcyclic versus postlexical rules in Lexical Phonology. *Linguistic Inquiry* 18. 1–44.

Borowsky, Toni. 1986. Topics in the lexical phonology of English. Ph.D. dissertation, University of Massachusetts, Amherst.

Brooks, Maria Zagorska. 1968. *Nasal vowels in Contemporary Standard Polish: An acoustic-phonetic analysis.* The Hague: Mouton.

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 283–333. Cambridge: Cambridge University Press.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable.* Cambridge, MA: MIT Press.

Cyran, Eugeniusz. 2003. *Complexity scales and licensing strength in phonology*. Lublin: Wydawnictwo Katolickiego Uniwersytetu Lubelskiego.

Cyran, Eugeniusz & Edmund Gussmann. 1999. Consonantal clusters and governing relations: Polish initial consonant sequences. In Harry van der Hulst & Nancy Ritter (eds.) *The syllable: Views and facts*, 219–247. Berlin & New York: Mouton de Gruyter.

Czaykowska-Higgins, Ewa. 1988. Investigations into Polish morphology and phonology. Ph.D. dissertation, MIT.

Dukiewicz, Leokadia. 1968. The acoustic-phonetic correlates of ą, ę in present-day Polish. In Wiktor Jassem (ed.) *Speech analysis and synthesis*, 53–68. Warsaw: Państwowe Wydawnictwo Naukowe.

Dukiewicz, Leokadia. 1985. Nagłosowe grupy spółgłosek w polskich tekstach popularno naukowych i prasowych. *Studia Gramatyczne* 6. 17–34.

Dukiewicz, Leokadia. 1995. Fonetyka. In Wróbel (1995), 7–103.

Dunaj, Bogusław. 1985. *Grupy spółgłoskowe współczesnej polszczyzny mówionej w języku mieszkańców Krakowa*. Kraków: Uniwersytet Jagielloński.

Feldstein, Ronald. 1983. Positional variation of nasals in Contemporary Standard Polish. *International Journal of Slavic Linguistics and Poetics* 28. 7–23.

Fisiak, Jacek & Stanisław Puppel (eds.) 1992. *Phonological investigations*. Amsterdam: John Benjamins.

Gladney, Frank Y. 2004. On sonorants and syllabicity in Polish. *Biuletyn Polskiego Towarzystwa Językoznawczego* 60. 117–132.

Gorecka, Alicja. 1988. Epenthesis and the coda constraint in Polish. Unpublished ms., MIT.

Gussmann, Edmund. 1978. *Contrastive Polish–English consonantal phonology*. Warsaw: Państwowe Wydawnictwo Naukowe.

Gussmann, Edmund. 1980a. *Studies in abstract phonology*. Cambridge, MA: MIT Press.

Gussmann, Edmund. 1980b. The phonological structure of the Polish imperative. *Studia Gramatyczne* 3. 33–46.

Gussmann, Edmund. 1981. Glide shifts in Polish. In Wolfgang U. Dressler, Oskar E. Pfeiffer & John R. Rennison (eds.) *Phonologica 1980*, 169–178. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Gussmann, Edmund. 1991. Polish syllable structure: A hypothesis and its problems. In Maciej Grochowski & Daniel Weiss (eds.) *Words are physicians for an ailing mind*, 207–213. Munich: Sagner.

Gussmann, Edmund. 1992. Resyllabification and delinking: The case of Polish voicing. *Linguistic Inquiry* 23. 29–56.

Gussmann, Edmund. 2007. *The phonology of Polish*. Oxford: Oxford University Press.

Hayes, Bruce. 1980. A metrical theory of stress rules. Ph.D. dissertation, MIT.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hyman, Larry M. 1985. *A theory of phonological weight*. Dordrecht: Foris.

Isačenko, Aleksander V. 1970. East Slavic morphophonemics and the treatment of the jers in Russian: A revision of Havlík's Law. *International Journal of Slavic Linguistics and Poetics* 13. 73–124.

Itô, Junko. 1986. Syllable theory in prosodic phonology. Ph.D. dissertation, University of Massachusetts, Amherst.

Jarosz, Gaja. 2006. Polish yers and the finer structure of output–output correspondence. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 31. 181–192.

Jassem, Wiktor. 1973. *Podstawy fonetyki akustycznej*. Warsaw: Państwowe Wydawnictwo Naukowe.

Jespersen, Otto. 1904. *Lehrbuch der Phonetik*. Leipzig & Berlin: Teubner.

Kraska-Szlenk, Iwona. 2007. *Analogy: The relation between lexicon and grammar*. Munich: Lincom Europa.

Kuryłowicz, Jerzy. 1952. Uwagi o polskich grupach spółgłoskowych. *Biuletyn Polskiego Towarzystwa Językoznawczego* 11. 54–69. Reprinted (1960) in *Esquisses linguistiques*, 221–232. Wrocław: Ossolineum.

Laskowski, Roman. 1975. *Studia nad morfonologią współczesnego języka polskiego*. Wrocław: Wydawnictwo Polskiej Akademii Nauk.

Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.

110 Metaphony in Romance

Andrea Calabrese

The term metaphony, which was first used in French (*metaphonie*) as a translation of the German word *Umlaut*, is traditionally used in Romance linguistics to refer to a process in which a vowel assimilates partially or totally to the height of a following vowel. Metaphony is one of the most characteristic phonological processes characterizing the Romance languages, and in particular Italian varieties (although not Tuscan and, therefore, standard Italian, which is based on Tuscan). For example, in many Italian dialects, high mid vowels are raised to high before a high vowel. In the same context, low mid vowels can be raised to high mid, or diphthongized. This has resulted in numerous alternations, as in the examples /'verde ~ 'virdi/ 'green (SG~PL)', /'pɛde ~ 'pedi/ or /'pɛde ~ 'pjedi/ 'foot (SG~PL)', where the plural form shows metaphonic effects. Less commonly, a stressed low vowel may also be affected by metaphonic raising, and other changes may take place as well (cf. Maiden 1991; Calabrese 1985, 1998). The conditioning factor in the metaphonic alternations has been obscured in many dialects by the neutralization or deletion of final vowels. This has produced vowel alternations in nominal and verbal paradigms for which there is no overt trigger.

As observed by Anderson (1980: 43), the formal mechanism characterizing metaphony is not distinct from that underlying other sorts of vowel harmony (CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS; CHAPTER 118: TURKISH VOWEL HARMONY; CHAPTER 123: HUNGARIAN VOWEL HARMONY): it involves the spreading of a vowel feature from a given vowel to other vowels. As this term is used in Romance linguistics, however, the fundamental difference lies firstly in the feature that is spread – a height feature ([high], [low], or [ATR]; CHAPTER 21: VOWEL HEIGHT) in the case of metaphony – and secondly in the restriction on the target of the process – a stressed vowel. In "normal" vowel harmony, all vowels in a word can be targets of the harmonic process. In §2.1.3, I will consider Walker's (2005) analysis of metaphony, which accounts for the local properties of this phenomenon – i.e. for the fact that the target is a stressed vowel and the trigger a post-tonic high vowel – in a functional perceptual perspective. I will later (§4.1) briefly discuss an analysis of metaphony in terms of domain of application.

Other processes with the same restriction on a stressed target, but spreading non-height features, are found in other languages, starting from Germanic Umlaut. In this chapter I will restrict my attention to the phenomena that are traditionally referred to by the term metaphony in the Romance languages. This is an arbitrary choice from a theoretical point of view, but this delimitation will allow a tighter and more adequate exposition of facts and analysis.

1 Description of Italian metaphony

1.1 Types of vowel systems

In order to describe Italian metaphony adequately, a brief description of the vowel systems found in Italian dialects is needed. In stressed syllables, the typical system of Standard Italian (Tuscan) and many Italian dialects is the seven-vowel system in (1), which is given in (2) in terms of distinctive features (CHAPTER 17: DISTINCTIVE FEATURES; CHAPTER 19: VOWEL PLACE; CHAPTER 21: VOWEL HEIGHT).

(1)

i		u
e		o
ɛ		ɔ
	a	

(2)

	i	e	ɛ	a	ɔ	o	u
high	+	–	–	–	–	–	+
low	–	–	–	+	–	–	–
back	–	–	–	+	+	+	+
round	–	–	–	–	+	+	+
ATR	+	+	–	–	–	+	+

An important issue concerning (2) involves which feature accounts for the contrast between high mid and low mid vowels. Chomsky and Halle (1968), following Bloch and Trager (1942) and Jakobson *et al.* (1952), use the feature [tense] to account for this contrast. Problems with the feature [tense] were pointed out by Catford (1977) (see also Ladefoged and Maddieson 1996). In recent years, Romance phonologists have started to employ the feature [Advanced Tongue Root] ([ATR]), which was first used to account for vocalic contrast in West African languages (Ladefoged 1964; Lindau 1975; see also Halle and Stevens 1969), and I adopt this feature in (2). Ladefoged and Maddieson (1996) argue against this trend and propose that height contrasts in mid – and high – vowels in Germanic and Romance languages are simply due to different dislocations of the tongue body. Calabrese (2008) instead argues that these contrasts should indeed be accounted for in terms of advancement/non-advancement of the tongue root and that the apparent dislocation of the tongue body is due to the fact that the tongue is an elastic mass: thus, an advancement of the tongue root results in a slight raising and fronting of the tongue body. Evidence for this may be found in a recent ultrasound study of vowel articulations in the southern Salentino dialect of Tricase by Grimaldi *et al.* (2010), where it is shown that the difference found in this dialect between high mid /e/ and low mid /ɛ/, where the former is the result of metaphonic tensing before /i/, is the result of tongue root advancement as shown in Figure 110.1:

1.2 *Italian metaphony*

In seven-vowel systems, the typical targets of metaphony are the mid [+ATR] vowels /e o/, which raise to /i u/ when followed by a high vowel.[2] There is dialectal variation in the case of the mid [−ATR] vowels, where we essentially find three groups of dialects. In one group of dialects, these vowels diphthongize. In another group of dialects, they are raised to mid [+ATR] vowels. And finally, in still another group of dialects, they do not change in a metaphonic context. Typical metaphonic systems are exemplified by the dialects in (4), (5), and (6).

(4) *The dialect of Calvello* (Gioscïo 1985)[3]

Metaphonic alternations: /e o/ → [i u]; /ɛ ɔ/ → [je wo]

a. [+ATR] /e o/

i. Class I/II adjectives

	singular	*plural*	
masc	'sulu	'suli	'alone'
fem	'sola	'sole	
masc	'niru	'niri	'black'
fem	'nera	'nere	

ii. Class III adjectives and nouns

singular	*plural*	
ka'vrone	ka'vruni	'charcoal'
'mese	'misi	'month'
'verde	'virdi	'green'

iii. Metaphonic alternations in the present singular of verbs

'mitti	'kurri	(2nd person)
'mette	'korre	(3rd person)
'put'	'run'	

b. [−ATR] /ɛ ɔ/

i. Class I/II adjectives

	singular	*plural*	
masc	'ɣrwossu	'ɣrwossi	'big'
fem	'ɣrɔssa	'ɣrɔsse	
masc	'vjekkju	'vjekkji	'old'
fem	'vɛkkja	'vɛkkje	

ii. Class III adjectives and nouns

singular	*plural*	
'pɛre	'pjeri	'foot'
'dɛnte	'djenti	'tooth'
'fɔrte	'fworti	'strong'

[2] In some dialects only /i/ triggers the change. This issue will be discussed later.

[3] The examples from this dialect are given in intermediate forms. There is a low-level process that neutralizes unstressed vowels to [ə] in the final position of a phonological word so that there are alternations such as the following: /sandə/ 'saint' vs. /sandu pjetrə/ 'St Peter'; /kwirə/ 'that' vs. /kwiru kanə/ 'that dog' (see Gioscio 1985; Kaze 1991).

vowel system in (3). The neutralization in [ATR] values rendered the old metaphony system opaque, as in (9), where the same set of phonetically mid [−ATR] vowels has two different phonological behaviors according to the metaphonic context. The additional raising of mid vowels in unstressed syllables (['sɛntu ~ sin'timu] 'feel-PRES (1SG~1PL)', [ka'nɔsku ~ kanuʃ'ʃimu] 'know-PRES (1SG~1PL)' complicates the situation further (see Calabrese 1985 for an analysis).

(9) *Northern Salentino*

singular	*plural*	
'mɛsi	'misi	'month'
'nɔʧi	'nuʧi	'nut'
'pɛti	'pjɛti	'foot'
'kɔri	'kwɛri	'heart'

2 Analyses of metaphony

The major theoretical problem in characterizing Italian metaphony lies in the different phonological behavior of the mid [+ATR] and mid [−ATR] vowels. Recall that whereas mid [+ATR] vowels are consistently raised to high vowels across dialects in a metaphonic context (i.e. before high vowels), mid [−ATR] vowels may be diphthongized (4), raised to mid [+ATR] (5) or high [+ATR] (7), or even fail to undergo any change (6) in the same context. As we will see below, there is substantial agreement in the literature with respect to the basic nature of the process affecting mid [+ATR] vowels: it essentially involves height assimilation between a high vowel and the preceding stressed vowel. In contrast, there is no such agreement with regard to the nature of the process(es) affecting the mid [−ATR] vowels in a metaphonic context. On the one side, there are those who aim to achieve a unified account for metaphony and therefore propose that the same process applies to mid [+ATR] and mid [−ATR] vowels. For these scholars, the different outcomes observed in the case of the mid [−ATR] vowels are due to further phonological operations that apply to the outputs of height assimilation. On the other side, there are those who propose that an adequate account of the different outcomes observed in the case of the mid [−ATR] vowels requires the postulation of two independent processes: one of height assimilation applying to mid [+ATR] vowels, and a different one applying to mid [−ATR] vowels. In this section I will discuss the different proposals addressing the problem of the different phonological behaviors of the mid [+ATR] and mid [−ATR] vowels in a metaphonic context. A brief critical assessment of the different analyses of metaphony discussed in §2.1 and §2.2 is given in §2.3.

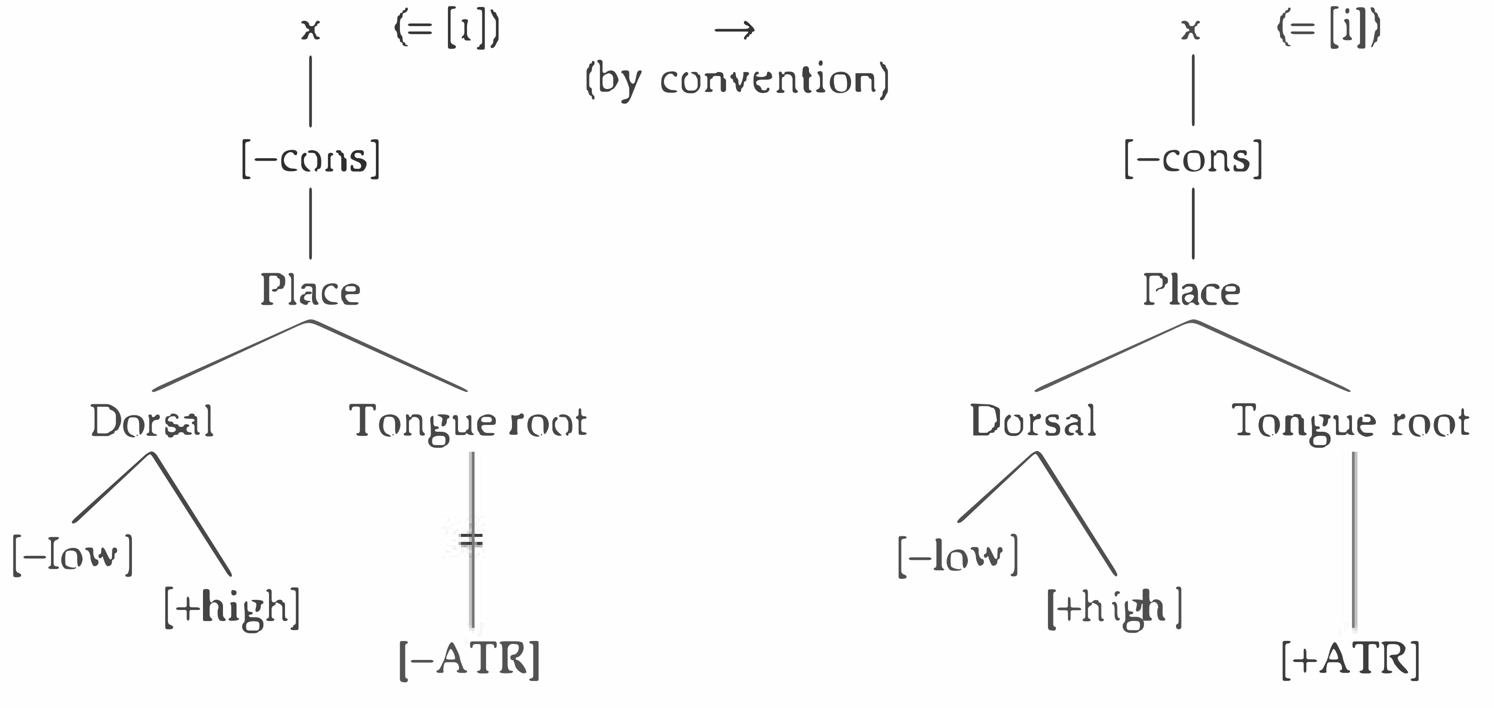
2.1 Metaphony as a unitary process

2.1.1 The analysis of metaphony in Calabrese (1985, 1998)

Calabrese (1985) was the first to analyze Italian metaphony in a generative and, more specifically, non-linear phonological framework. The dialect he studied was that of Francavilla Fontana (Ribezzo 1912) in northern Salento, where, despite some obscuring surface changes, one can easily reconstruct a metaphony system rules to the same disallowed configuration [+high, −ATR], created by metaphony in the case of mid [−ATR] vowels. In historical terms, he hypothesizes that the same rule of height assimilation spread across the southern Italian dialects. Each dialect reacted to the problem posed by the output of the application of this rule to the mid [−ATR] vowels in a different way.

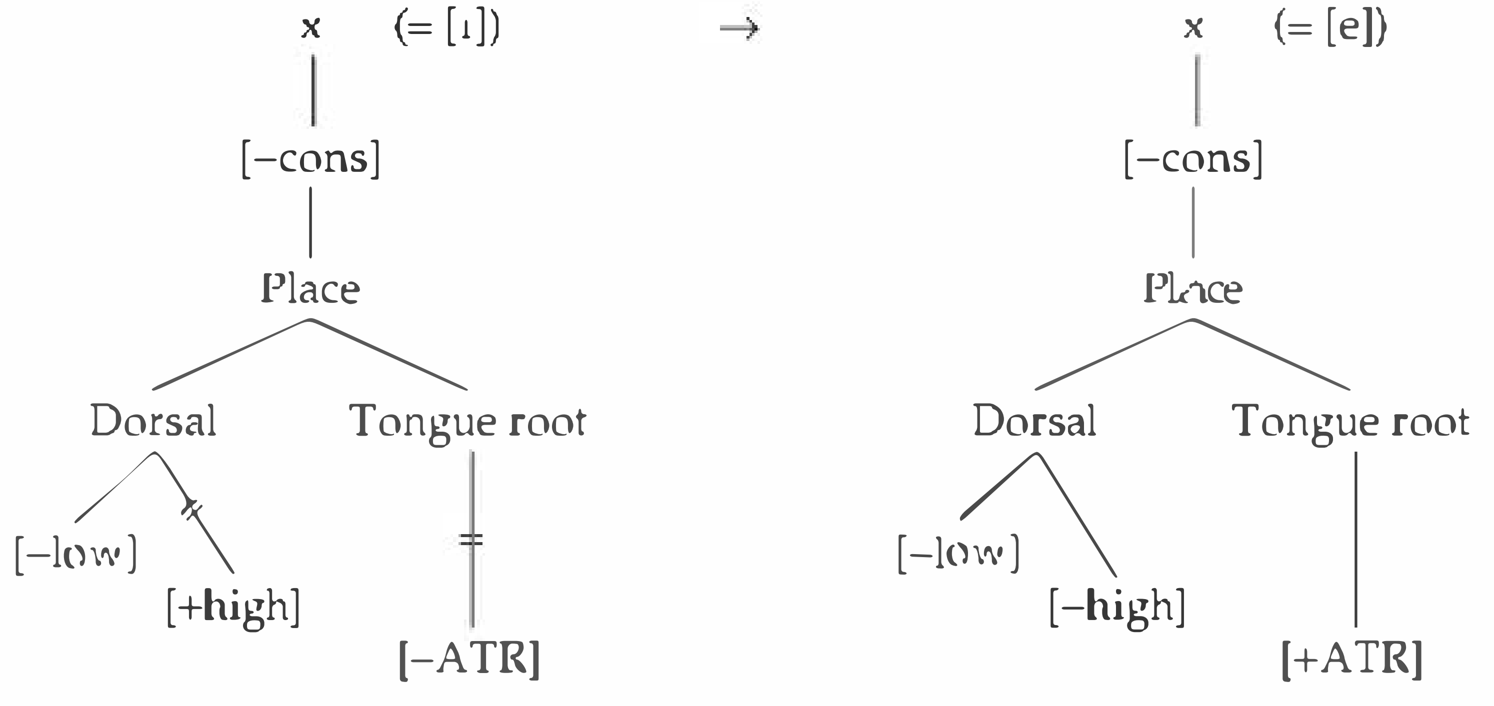
The relevant repairs he proposes are delinking and excision. Delinking involves an operation that removes one of the feature specifications disallowed by an active marking statement. The opposite feature specification is inserted by convention. This operation accounts for metaphonic raising to high vowels. In this case the feature specification [−ATR] of the disallowed configuration [+high, −ATR] created by the metaphony rule is delinked and the opposite feature specification [+ATR] is inserted, as shown in (12).

(12)



Excision involves removing both the feature specifications of the disallowed configuration and replacing them with their opposite specifications (see Calabrese 1995, and especially 2005, for further discussion of excision). If we assume that the rule of metaphony is the same across dialects and that the variation is due to the application of a different repair, we can hypothesize that excision is the relevant repair in southern Umbro. Thus, we have a repair like the one in (13).

(13)



That is, the [–ATR] high vowels produced by metaphony are changed into [+ATR] mid vowels.[10]

Calabrese (1985, 1995) accounts for metaphony systems where only mid [+ATR] vowels, but not mid [–ATR] ones, undergo metaphony (see the dialect of Grado in (6)) by postulating that there is a parametrized option that governs the interaction between rules and active constraints. If applying a rule would generate a violation of a constraint, the application of the rule could either be (i) blocked (structure preservation) or (ii) allowed to apply. Under option (ii), a repair then fixes the illicit configuration that is so produced. The dialects discussed above select this option. The dialects where metaphony does not apply to mid [–ATR] vowels select option (i): so the rule in (10) is prevented from applying to these vowels because of the active constraint *[+high, –ATR] (see Calabrese 1998 for further discussion of this case).

I will now discuss two different accounts of Italian metaphony that essentially preserve Calabrese's hypothesis that there is a single metaphonic process raising all mid vowels to high ones and that the different treatments found in the case of the mid [–ATR] vowels are due to different processes. These are the analyses of Maiden (1991) and Walker (2005). Observe that only Walker accounts for the dialectal variation we find in the case of metaphony of the mid [–ATR] vowels. Maiden (1991) accounts only for the diphthongization observed in dialects such as that of Calvello. In this section I will also consider Nibert's (1998) analysis of metaphony in Servigliano.

2.1.2 *The analysis of metaphony in Maiden (1991)*

Maiden (1991) provides a detailed study of metaphony in its diachronic development and in its synchronic phonological and morphonological aspects. In his analysis of the phonology of this sound change, Maiden treats metaphony as an assimilation in height, as proposed by Calabrese; however, he disagrees with Calabrese's account of the diphthongization process found in lower mid vowels. In his account of metaphony, Maiden assumes the framework of dependency phonology. Following dependency phonology, he proposes that vowel height in a seven-vowel system like that of the Italian dialects should be represented as in (14), where vowel space is characterized by four components: |i| (palatality or acuteness); |a| (lowness or sonority); |u| (roundness or gravity); |ə| (centrality (not used here)).

(14) {|i|}	= /i/	{|u|}	= /u/
{i;a}	= /e/	{u;a}	= /o/
{a;i}	= /ɛ/	{a;u}	= /ɔ/
{|a|}	= /a/		

(Curly brackets indicate that the segment is characterized phonologically by just that component or combination of components, and ";" symbolizes asymmetrical right-to-left "government" relationships between components.)

[10] Calabrese (1985) uses excision to account for the metaphonic raising of /a/ to a mid vowel [ɛ ɔ]. The application of the metaphony rule to /a/ creates the disallowed configuration [+high, +low] which by excision becomes [–high, –low].

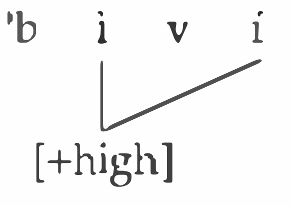
Metaphony, in other words, accomplishes improved perceptibility by extending the height features to the stressed syllable. She uses the positional licensing constraint in (18) to represent the need for high vowels to be associated with a prosodically strong position.

(18) LICENSE([+high]post-tonic, 'σ)

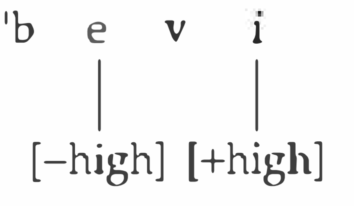
[+high] in a post-tonic syllable must be associated with a stressed syllable.

The constraint in (18) is satisfied in (19a), but not in (19b). The alternative solution in (19c), where the mid character of the stressed vowel overrides the post-tonic vowel, would result in unstressed vowel lowering. Walker eliminates this solution by assuming a preference for minimized sonority (CHAPTER 49: SONORITY) in unstressed syllables (CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE).

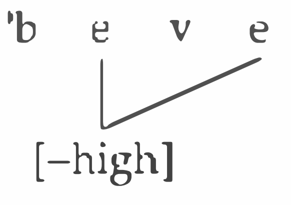
(19) a. *Licensing: satisfied*



b. *Licensing: violated*



c. *Licensing: satisfied (sonority minimization in unstressed syllables violated)*



In the OT scheme Walker proposes, all height features of post-tonic high vowels ([+high], [±ATR], [−low]) in metaphonic systems require licensing. This is implemented by LICENSE(height), which subsumes (18). In all of these systems, LICENSE(height) is ranked over IDENT[high], thus forcing spreading of the feature [high]. At the same time, *[+high, −ATR] is strictly enforced across patterns to block [+high, −ATR], as in Calabrese's analysis discussed in §2.1.1.

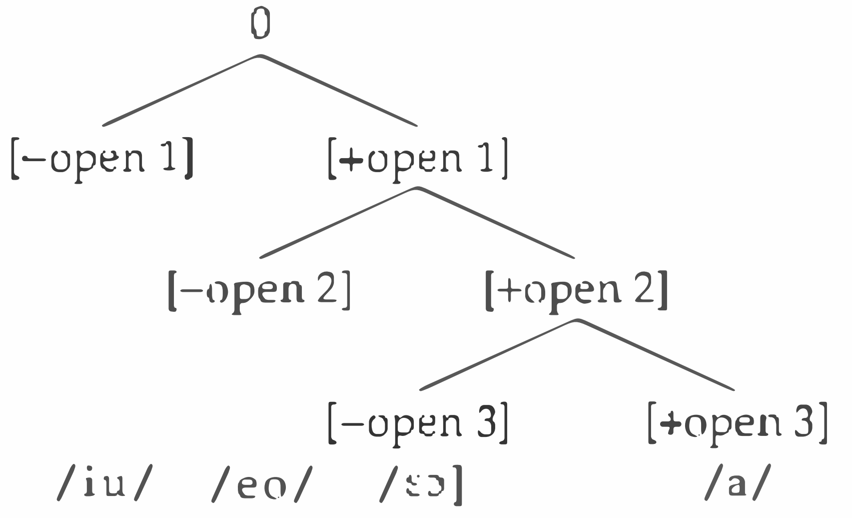
The dialectal variation we observe across the dialects is obtained as follows in Walker's OT analysis.

In Veneto, where metaphony applies only to [+ATR] mid vowels, the higher ranking IDENT[ATR] and IDENT[low] prevent [−ATR] vowels from being affected in metaphony. LICENSE(height) is ranked above IDENT[high]. Therefore the only allowed pattern is that in (6).

In the Pugliese dialect of Foggia (see (7)), where /ɛ ɔ/ raise to [i u], licensing for all height features is also capable of overriding IDENT[ATR]. Hence, in contrast to Veneto, LICENSE(height) dominates IDENT[ATR] in the dialect of Foggia.

Like Pugliese, metaphony in southern Umbro (see (5)) has the capacity to violate IDENT[ATR], but it shows gradual raising. Walker follows Kirchner (1996) in accomplishing this by local conjunction of faithfulness constraints for height features, which moderates satisfaction of height licensing (CHAPTER 62: CONSTRAINT CONJUNCTION).

(22) *Recursive expansion of* [open]



Servigliano has the seven-vowel inventory /i e ɛ a ɔ o u/. Metaphony induces a one-step raising of both the high mid vowels /e o/ and the low mid vowels /ɛ ɔ/. Nibert specifies the vowel height features for Servigliano in (23). Note that this is a center-embedding structure, as opposed to the right-branching structure in (22). The choice between the center-embedding and right-branching analysis of a four-height system depends on the behavior of the mid vowels in an individual language. (22) best represents a system in which the low mid vowels pattern with the low vowel with respect to vowel height, while (23) is an appropriate specification for languages like Servigliano, in which the low mid and high mid vowels pattern together.

(23)

	i	e	ɛ	a	ɔ	o	u
open 1	–	–	–	+	–	–	–
open 2	–	+	+	+	+	+	–
open 3	–	–	+	+	+	–	–

Nibert shows that under this analysis of vowel height, the raising of /e o/ requires assimilation of [–open 2] from the triggering vowels. But spreading [–open 2] onto /ɛ ɔ/yields the feature bundle [–open 1, –open 2, +open 3], which doesn't correspond to any of the available vowel qualities. In order to raise /ɛ ɔ/ in one step to /e o/, it is the feature [–open 3] that must be assimilated. Nibert argues that metaphony can spread either feature, [–open 2] or [–open 3], subject to structure preservation, which bans the spread of [–open 2] onto /ɛ ɔ/. In order to guarantee that [–open 2] and not [–open 3] spreads onto /e o/, since spread of [–open 3] would be vacuous, and would not accomplish raising, Nibert's rule must be modified to apply disjunctively, spreading [–open 2] unless spreading is blocked, in which case [–open 3] will spread.

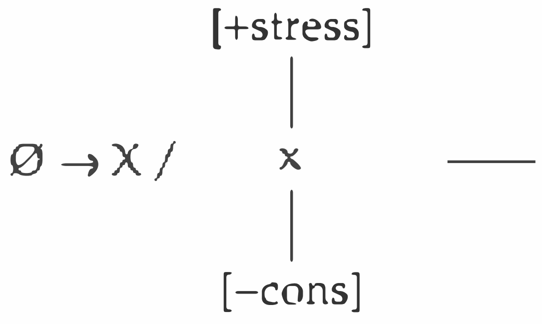
2.2 *Italian metaphony as two different processes*

The analyses that I consider in this section reject the idea that there is a unitary metaphonic process, and propose to break metaphony into two distinct processes – one targeting the mid [+ATR] vowels and raising them to high ones and the other targeting the mid [–ATR] vowels, leading to diphthongization or tensing. These are the analyses of Sluyters (1988) and Cole (1998).

2.2.1 *The analysis of metaphony in Sluyters (1988)*

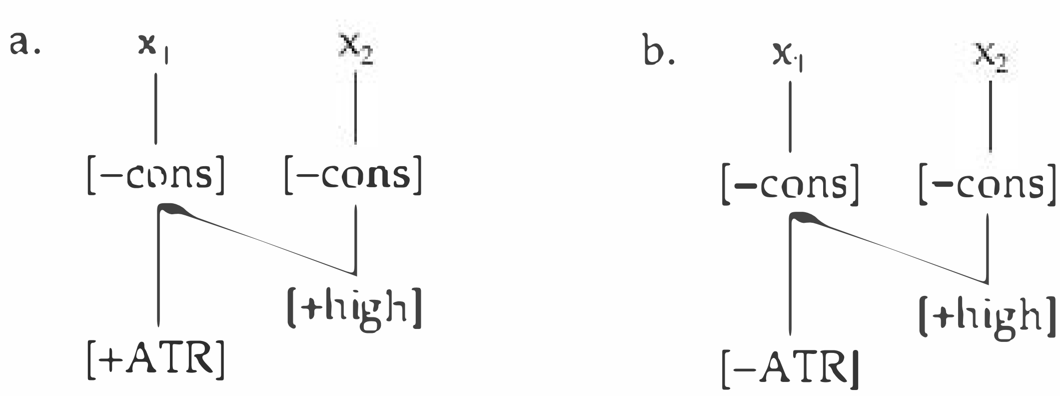
Sluyters (1988) proposes a reanalysis of Calabrese's (1985) account of metaphony in the northern Salentino dialect of Francavilla Fontana (see (9)). Assuming an underlying system as in (2) for this dialect, he adopts Calabrese's rule (10) (see §2.1.1 above) to account for height assimilation before high vowels. However, he proposes a different account of diphthongization in metaphony. Sluyters suggests that the metaphonic diphthongs /iɛ uɔ/ are characterized by a second x-slot which is contextually determined by stress and introduced by a rule like that in (24).

(24)



In Sluyters' analysis, Calabrese's rule (10) directly accounts for the raising observed in the case of mid [+ATR] vowels. According to Sluyters, however, (10) does not apply in the case of the mid [−ATR] vowels until after rule (24) has applied. However, the facts cannot be accounted for by simply applying (10) before and after (24). The first application of (10) would also affect mid [−ATR] vowels. It is therefore necessary to split rule (10) into two different rules: one applying to mid [+ATR] vowels and the other to mid [−ATR] vowels as in (25).

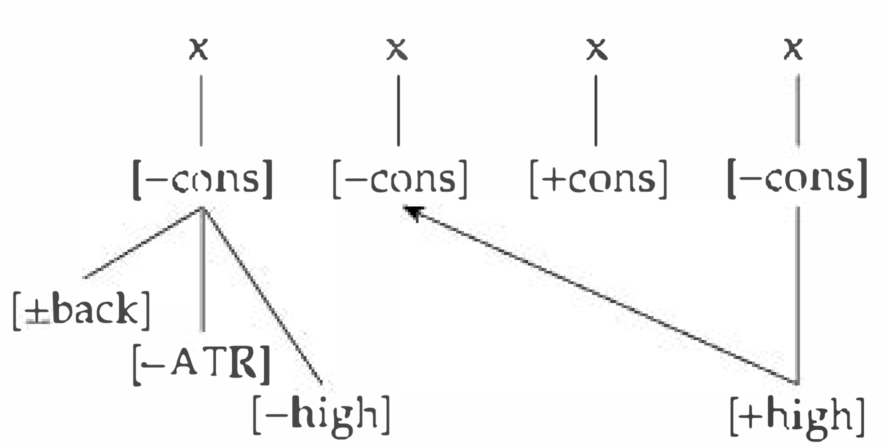
(25)



where x_1 is the head of the main stress foot

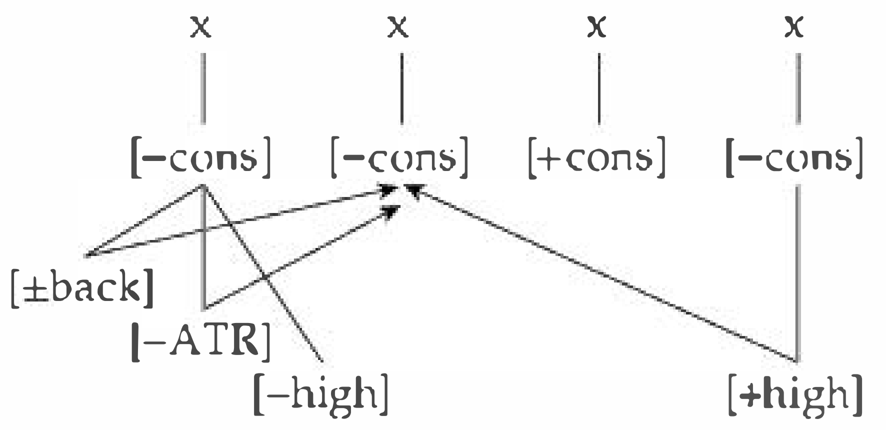
(25a) applies first. Then (24) applies. After the application of (24), rule (25b) fills in the empty x-slot with the feature [+high] as shown in (26) (tree structures are simplified).

(26)



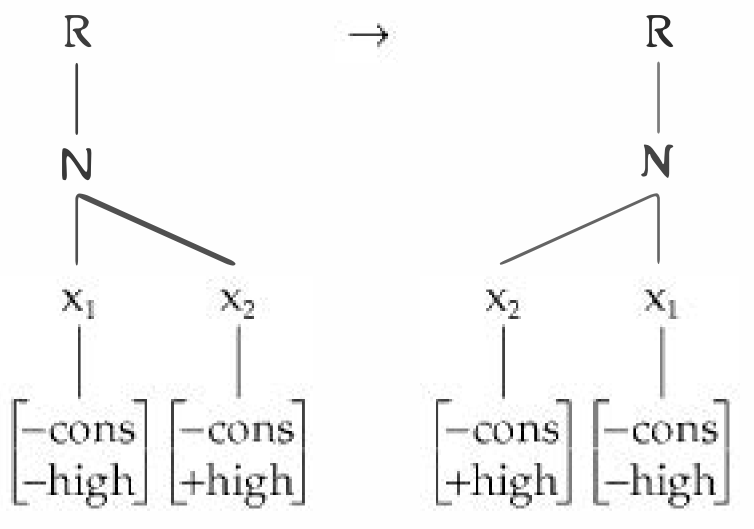
The other features of the inserted empty vowel are filled in by spreading from the preceding vowel, according to Sluyters.

(27)



Such an analysis produces the falling diphthongs /ɛj ɔw/. In order to get the correct outputs, Sluyters assumes the rule of metathesis in (28).

(28)



Application of (28) to the outputs of (27) generates the metaphonic outcomes /jɛ wɔ/.

2.2.2 *The analysis of metaphony in Cole (1998)*

Cole (1998) argues that the vowel raisings that define metaphony systems in Italian dialects do not result from a unified operation of height assimilation. Instead, metaphony is claimed to be the product of the assimilation of the mid [+ATR] vowels /e o/ to the high vowels [i u] within the stress foot (CHAPTER 40: THE FOOT), and a subsequent vowel shift by which the non-high /(a) ɛ ɔ/ are raised one step.

Like Walker (2005), Cole views the height assimilation of the mid [+ATR] /e o/ to /i u/ in functional terms. Following Cole and Kisseberth (1994), she assumes that assimilation promotes perceptual salience by reducing or eliminating marginal contrasts between two phonetically similar segments. Small phonetic differences between the trigger and target of assimilation are resolved, and in many vowel harmony systems the target emerges as fully identical to the trigger. Thus the assimilated target provides additional or extended acoustic cues for the identification of the triggering segment.

Cole then proposes that the assimilation in height of the mid [+ATR] leaves a vacancy in the vertical dimension of the vowel space, and a vowel shift takes place to raise the lower mid [−ATR] vowels, and thereby fill the gap.[11] This accounts for the tensing of the mid [−ATR] vowel observed in southern Umbro.

[11] Vowel shift is governed by two principles, according to Cole. One prohibits the neutralization of contrastive height categories, while the other one requires the preservation of the relative underlying height of raised vowels (see Cole 1998 for more discussion).

is filled by the spreading |i| component. In this way, according to the reviewer, it is possible to explain why lower mid is changed to a rising diphthong. (See Calabrese 1998 for further discussion of Maiden 1991.)

2.3.3 *Discussion of Walker's analysis*

Walker's (2005) article (see §2.1.3) not only provides an excellent description of metaphony in Veneto and other Italian dialects, but also an exhaustive overview of metaphony-like phenomena from a wide range of languages. Her analysis is the first to deal successfully and convincingly with all aspects of metaphony in an OT framework. Her constraint in (18) could provide a functional basis for the metaphony rule in (10). However, this can be true only diachronically, as an account of the phonological innovation that lead to metaphonic alternations. In contrast, it is difficult to hypothesize that the constraint in (18) is operative synchronically, at least for most Italian dialects where metaphony is morphologized or obscured by subsequent phonological processes (see §1). Functional explanations by their nature must be surface-true. This is not the case in Italian metaphony.

2.3.4 *Discussion of Nibert's analysis*

Nibert's analysis of the Servigliano dialects (see §2.1.4) holds only for this dialect and cannot account for the diphthongization or the dialectal variation we observe in the case of this phenomenon (see Zetterstrand 1998 for a criticism of Clements' 1989 approach to vowel height). Furthermore, as observed by Cole (1998), Nibert's analysis cannot deal with metaphony systems in which the low vowel also participates, undergoing a one-step raising to [e]. In such a system, a different [open] feature would have to assimilate for each of the low, low mid, and high mid vowels. Under an assignment of height features as in (24), raising [a] one step would require the assimilation of [−open 1], in addition to the assimilation of the features [−open 2] and [−open 3] required for the raising of [ɛ ɔ] and [e o], respectively. See Cole (1998) for further discussion of Nibert (1998).

2.3.5 *Discussion of Sluyters' analysis*

Sluyters' analysis (see §2.2.1) has several problematic aspects. Rule (24) is a rule of lengthening and applies independently of metaphony. It then predicts long vowels in all stressed syllables in non-metaphonic contexts. This is false. Lengthening in Italian dialects is restricted to open stressed syllables in penultimate position (Rohlfs 1966).[13] Furthermore, (24) must apply only to [−ATR] mid vowels, and not to the [+ATR] ones; otherwise it wrongly creates diphthongs in the case of the latter vowels. Rule (24) is thus simply wrong. The metathesis rule in (28) fares no better, and cannot be motivated either for Salentino or for any other dialects having metaphonic diphthongization. Furthermore, Sluyters' analysis does not account for either the tensing or the raising of mid [−ATR] vowels found in other Italian dialects.

[13] On the other hand, metaphonic diphthongization, as shown in Calabrese (1985, 1988), is not restricted to such an environment. It applies both in closed syllables and in stressed syllables of all types in antepenultimate position.

2.3.6 *Discussion of Cole's analysis*

Cole's (1998) functional analysis of metaphony (see §2.2.2) faces the same problems as Walker's (2005): it may hold diachronically for the stage of Italian in which metaphony was introduced as a phonological innovation, but it cannot hold synchronically for most, if not all, Italian dialects where metaphony is either morphologized or obscured by subsequent phonological processes. Another problematic aspect of her account involves diphthongization as a way to maintain contrasts. According to Cole:

> by diphthongizing a raised low mid vowel, the surface form unambiguously identifies the contrastive category of the vowel, countering the neutralization that would otherwise result from compressing four contrastive height categories into three. (1998: 93)

The problem is that metaphonic diphthongization of mid [−ATR] vowels also occurs in Sicilian or central Salentino dialects that have an underlying five-vowel system. Given that in these dialects there are no mid [+ATR] vowels, diphthongization cannot be explained as a way to counter the neutralization between the two sets of mid vowels.

3 Typological variation in metaphony as a height assimilation process

The term metaphony is also used to refer to processes of height assimilation in which features other than [+high] are spread. In this section I will survey the variation we find with respect to the features spread by metaphony across Romance.

3.1 *Metaphonic processes spreading* [+ATR]: *Sardinian metaphony*

There are Romance varieties in which we find metaphonic alternations that are better analyzed as involving the spreading of feature [+ATR], instead of [+high], of the high vowels. This is for example the case of Sardinian metaphony.[14] Frigeni (2003) discusses the following examples from Campidanese and Nuorese, the main varieties of Sardinian. Sardinian has the vowel system in (3).

(30) *Campidanese*

a.	*masc sg* [-u]	*fem sg* [-a]	
	'lentu	'lɛnta	'slow'
	'veru	'vɛra	'true'
	'nou	'nɔa	'new'
	'mot:u	'mɔt:a	'dead'
b.	*sg*	*pl*	
	'tempus	'tɛmpos	'time'
	'krop:us	'krɔp:os	'body'
	'betʃ:u	'bɛtʃos	'old (MASC)/parent'
	'ot:u	'ɔt:os	'vegetable garden'

[14] See also southern Salentino, discussed in §1.1, for another example of this type of metaphony (see also Grimaldi 2003).

THE REPRESENTATION OF WORD STRESS). In fact, in dialects such as Neapolitan, metaphony in words with antepenultimate stress appears to apply regularly across the non-high post-tonic vowel, although this occurs only when the final vowel is /i/: [ˈmɔnako ~ ˈmuonatʃi] 'monk (SG~PL)', [ˈtʃɛfaro ~ ˈtʃiefari] 'mullet (SG~PL)', [kaˈrɔfano ~ kaˈruofani] 'carnation (SG~PL)', [ˈfekato ~ ˈfiiekati] 'liver (SG~PL)' (Rohlfs (1966: 22). Note that these are intermediate representations; final [-o, -i] become [ə], due to a surface process of vowel reduction. However, there are also cases in which it applies before /u/: Centrache [ˈmuɔnaku] 'monk' (Lausberg 1939: 10; Rohlfs 1966: 427).[19] What matters in these cases is the target of the process – the stressed vowel – and the element that appears in post-tonic position can be disregarded. This type of situation is even more clear in the Asturian dialects that are characterized by [+high] metaphony (cf. [ˈnenos ~ ˈninu] 'child (MASC PL~SG)', [ˈtsobos ~ ˈtsubu] 'wolf (MASC PL~SG)' (Hualde 1998). In these dialects, metaphony also targets /a/, which is raised to a mid vowel. In words with antepenultimate stress, metaphony can target a stressed /a/ leaving unaffected a post-tonic /a/, as illustrated in the following examples.

(36) *Lena*

ˈpeʃaru	ˈpaʃarin	ˈpaʃara	'bird ~ little bird ~ female bird'
ˈpempanu	ˈpampanos	'old and decrepit person (MASC SG~PL)'	
ˈkendanu	ˈkandanos	'dry branch (SG~PL)'	

If metaphony is characterized as applying to the last foot, it cannot be explained why the intermediate vowel can be skipped over by this process. The best characterization of metaphony is not in terms of domain of application, but in terms of target of application: metaphony applies to stressed vowels.[20]

4.2 *Metaphony and morphology*

A debate that has existed for some time in connection with metaphony is the role of morphology (e.g. Maiden 1991; Hualde 1992). A central question is whether metaphony processes are conditioned by morphology, phonology, or some combination of the two. The fact is that in many Italian dialects, historical changes such as reduction to schwa, or deletion, of unstressed or final vowels can no longer be motivated synchronically through alternations. Since these historical processes removed the phonological conditions for metaphony, in these dialects the metaphonic changes affecting the stressed vowels have become the only surface markers of inflectional categories such as the masculine singular and

[19] In many, perhaps most, Italian dialects the phonological conditions for the application of metaphony in words with antepenultimate stress are obscured by schwa-reduction of post-tonic vowels [mɔnakə ~ muonəkə] (Campobasso; Rohlfs 1966: 22) or regressive total harmony between post-tonic vowels [ˈmɔnuku ~ ˈmwɛnitʃi] 'monk-(SG~PL)', [kɔfimu ~ ˈkwɛfini] 'barrel (SG~PL)', [ˈmjɛtucu ~ ˈmjɛtitʃi] 'physician-(SG~PL)' (Francavilla Fontana; Ribezzo 1912; Calabrese 1985). Metaphony is constrained by a further condition in many dialects: it applies only between adjacent vowels, so there is no metaphony in [ˈkɔfanu ~ ˈkɔfani] 'chest (SG~PL)', [ˈmɔnaku ~ ˈmɔnaki] 'monk (SG~PL)' (Aprigliano; Rohlfs 1966: 189, 427).

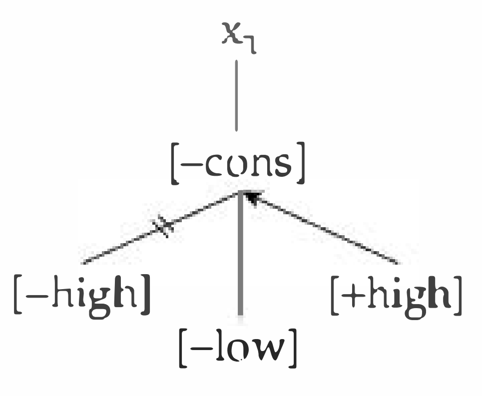
[20] Note that under such an analysis we still have to account for why the non-contrastive feature [−high] of the post-tonic vowel does not interfere with the metaphonic spreading process: either it must be underspecified or one must assume that the metaphony rule accesses only contrastive features, as in Calabrese's (1995) re-analysis of underspecification theory (see also Nevins 2010).

plural of nominal class II, the plural of nominal class III, the 2nd singular of the present indicative, etc. (CHAPTER 82: FEATURAL AFFIXES; CHAPTER 103: PHONOLOGICAL SENSITIVITY TO MORPHOLOGICAL STRUCTURE). Therefore they are most adequately analyzed in terms of a morphophonological rule. An example is provided below from Arpinate (see also (8) for examples in other morphological categories in this dialect) where the marking of the 2nd person singular is obtained solely through the application of metaphony to the stressed vowel of the verbal stem.

(37) *Metaphonic alternations in the present singular of verbs*

'vedə	'korrə	'sɛntə	'mettə	'kɔʎʎə
'vidə	'kurrə	'sjentə	'mittə	'kwoʎʎə
'vedə	'korrə	'sɛntə	'mettə	'kɔʎʎə
'see'	'run'	'feel'	'put'	'pick'

For the dialects in which metaphony is morphologized, Calabrese (1998) proposes the rule in (38), which must be specified as applying in morphological contexts such as the masculine singular and plural of nominal class II, in the plural of nominal class III, in the 2nd singular of the present indicative, etc.

(38) 

where x_1 is the head of the main stress foot

In the account developed above, metaphony was originally a phonologically motivated innovation that later became morphologized. However, an alternative view of how "morphological" metaphony developed is proposed by Devoto (1974). He assumes that, when in a language the exponence of inflectional categories is obscured or lost because of phonological processes, speakers of the language may react to the loss of morphological information by introducing alternative morphological realizations for these categories. Italian metaphony, according to him, then originated in order to realize morphologically the inflectional features whose exponence was threatened by the phonetic erosion of inflectional endings. This was done by marking the height features of the endings onto the stressed vowels. Under such an approach, metaphony is a morphologically motivated innovation. The presence of dialects that display metaphonic alternations and fully preserved inflectional endings – see for example the dialect of Servigliano in (5), which did not undergo changes neutralizing final vowels – shows that this idea is too strong (see Maiden 1991: 212 for more discussion). A recent contribution that argues for the relevance of morphology in the origin of metaphony is Finley (2009) on Lena, who labels metaphony as a type of "morphemic harmony" (see also Kurisu 2001, who analyzes umlaut as a kind of "morpheme realization").

A more nuanced position on the role of morphology in metaphony is taken by Maiden (1991), who argues that morphological conditioning does play a role in the development of at least some metaphonic processes. This is the case of the

Cruschina, Silvio. 2008. Il vocalismo della sicilia centrale: Il tratto [ATR], metafonesi e armonia vocalica. Unpublished ms., Cambridge University.

Devoto, Giacomo. 1974. *Il linguaggio d'Italia*. Milan: Rizzoli.

Dyck, Carrie. 1995. Constraining the phonology–phonetics interface, with exemplification from Spanish and Italian dialects. Ph.D. dissertation, University of Toronto.

Finley, Sara. 2009. Morphemic harmony as featural correspondence. *Lingua* 119. 478–501.

Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles.

Frigeni, Chiara. 2003. Metaphony in Campidanian Sardinian: A domain-based analysis. *Toronto Working Papers in Linguistics* 20. 63–91.

Gioscio, Joseph. 1985. *Il dialetto Lucano di Calvello*. Stuttgart: Steiner.

Grimaldi, Mirko. 2003. *Nuove ricerche sul vocalismo tonico del Salento meridionale*. Alessandria: Edizioni dell'Orso.

Grimaldi, Mirko, Andrea Calabrese, Francesco Zagona, Luigina Garrappa & Bianca Sisinni. 2010. Articulatory grounding of southern Salentino harmony processes. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech)*, 1561–1564. Makuhari, Japan: ISCA.

Halle, Morris & Kenneth N. Stevens. 1969. On the feature "advanced tongue root." *MIT Research Laboratory of Electronics Quarterly Progress Report* 94. 209–215.

Hualde, José Ignacio. 1989. Autosegmental and metrical spreading in the vowel-harmony systems of northwestern Spain. *Linguistics* 27. 773–805.

Hualde, José Ignacio. 1992. Metaphony and count/mass morphology in Asturian and Cantabrian dialects. In Christiane Laeufer & Terrell Morgan (eds.) *Theoretical analysis in Romance linguistics*, 99–114. Amsterdam: John Benjamins.

Hualde, José Ignacio. 1998. Asturian and Cantabrian metaphony. *Rivista di Linguistica* 10. 99–108.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Kaun, Abigail. 1995. The typology of rounding harmony: An optimality theoretic approach. Ph.D. dissertation, University of California, Los Angeles.

Kaze, Jeffrey W. 1991. Metaphony and two models for the description of vowel systems. *Phonology* 8. 163–170.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Kirchner, Robert. 1996. Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry* 27. 341–350.

Kurisu, Kazutaka. 2001. The phonology of morpheme realization. Ph.D. dissertation, University of California, Santa Cruz.

Ladefoged, Peter. 1964. *A phonetic study of West African languages*. Cambridge: Cambridge University Press.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford & Malden, MA: Blackwell.

Lausberg, Heinrich. 1939. *Die Mundarten Südlukaniens*. Zeitschrift für Romanischen Philologie: Beiheft 90.

Lindau, Mona. 1975. [Features] for vowels. Ph.D. dissertation, University of California, Los Angeles.

Loporcaro, Michele. 1991. The natural phonological process [V, +high] → [+Tense] and the vowel system of some southern Italian dialects. *Folia Linguistica* 15. 3–4.

Loporcaro, Michele. 1997. Puglia and Salento. In Maiden & Parry (1997), 338–348.

Maiden, Martin. 1987. New perspectives on the genesis of Italian metaphony. *Transactions of the Philological Society* 85. 38–73.

Maiden, Martin. 1991. *Interactive morphonology: Metaphony in Italy*. New York: Routledge.

Maiden, Martin & Mair Parry (eds.) 1997. *The dialects of Italy*. London: Routledge.

Mancarella, Giovann-Battista. 1974. Origine e diffusione della metafonia nel Salento. *Atti del XIV Congresso di Linguistica e Filologia Romanze* 3. 117–124.

Marotta, Giovanna & Leonardo Savoia. 1991. Diffusione vocalica in un dialetto Calabrese. Alcuni parametric fonologici. In E. Magno Caldognetto & Paola Beninca (eds.) *L'interfaccia fra fonologia e fonetica*, 19–43. Padua: Unipress.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.

Melillo, G. 1926. *I dialetti del Gargano: Saggio fonetico.* Pisa: Simoncini.

Miglio, Viola. 2005. *Markedness and faithfulness in vowel systems.* New York: Routledge.

Navarro Tomás, Tomás. 1939. *Manual de la pronunciación española.* Madrid: Consejo Superior de Investigaciones Cientificas.

Nevins, Andrew. 2010. *Locality in vowel harmony.* Cambridge, MA: MIT Press.

Nibert, Holly. 1998. Processes of vowel harmony in the Servigliano dialect of Italian: A comparison of two non-linear proposals for the representation of vowel height. *Probus* 10. 67–101.

Parodi, Ernesto. 1892. Il dialetto di Arpino. *Archivio Glottologico Italiano* 13. 299–308.

Penny, Ralph J. 1969. *El habla pasiega: Ensayo de dialectología montañesa.* London: Tamesis Books.

Penny, Ralph J. 1978. *Estudio estructural del habla de Tudanca.* Tübingen: Niemeyer.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Prokosch, Eduard. 1938. *A comparative Germanic grammar.* Baltimore: Linguistic Society of America.

Ribezzo, Francesco. 1912. *Il dialetto apulo-salentino di Francavilla Fontana.* Martina Franca.

Rohlfs, Gerhard. 1966. *Grammatica storica della lingua italiana e dei suoi dialetti: Fonetica.* Turin: Einaudi.

Sanders, Benjamin P. 1998. The Eastern Andalusian vowel system: Form and S structure. *Rivista di Linguistica* 10. 109–135.

Savoia, Leonardo & Martin Maiden. 1997. Metaphony. In Maiden & Parry (1997), 15–25.

Sluyters, Willebrord. 1988. Vowel harmony, rule formats and underspecification: The dialect of Francavilla-Fontana. In Harry van der Hulst & Norval Smith (eds.) *Features, segmental structure and harmony processes*, part II, 161–184. Dordrecht: Foris.

Steriade, Donca. 1994. Positional neutralization and the expression of contrast. Unpublished ms., University of California, Los Angeles.

Steriade, Donca. 1995. Underspecification and markedness. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 114–174. Cambridge, MA & Oxford: Blackwell.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie.* Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.

Trumper, John. 1979. La zona Lausberg ed il problema della frammentazione linguistica. In Federico Albano Leoni (ed.) *I dialetti e le lingue delle minoranze di fronte all'italiano*, vol. 1, 267–303. Rome: Bulzoni.

Valente, Vincenzo. 1975. *Puglia.* Pisa: Pacini.

Walker, Rachel. 2005. Weak triggers in vowel harmony. *Natural Language and Linguistic Theory* 23. 917–989.

Weinrich, Harald. 1958. *Phonologischen Studien zur romanischen Sprachgeschichte.* Münster: Aschendorff.

Zetterstrand, Sylvia. 1998. The phonological representation of vowel height. Ph.D. dissertation, Harvard University.

Zubizarreta, Maria Luisa. 1979. Vowel harmony in Andalusian Spanish. *MIT Working Papers in Linguistics* 1. 1–11.

111 Laryngeal Contrast in Korean

Young-mee Yu Cho

1 Introduction

Starting from the structuralist tradition (Martin 1951, 1954), and continuing into the generative framework, attempts to identify the laryngeal contrast in the obstruent system of Korean have led to intense discussion. Even today, it is one of the areas that still actively engage researchers of Korean grammar.

There is general consensus that on the surface there are three types of phonation in Korean; to be more precise, there is a three-way contrast for stops and a two-way contrast for fricatives, as shown in the consonantal inventory in (1).[1]

(1) *Korean consonantal phoneme inventory*

	labial	alveolar	prepalatal[2]	velar	glottal
plain stops	p	t	ts	k	
aspirated stops	pʰ	tʰ	tsʰ	kʰ	
tense stops	p'	t'	ts'	k'	
plain fricatives		s			h
tense fricatives		s'			
nasals	m	n		ŋ	
liquid		l			

(2) lists minimal pairs exemplifying the consonantal inventory of (1).

[1] The two fricatives /s/ and /s'/ exhibit an interesting array of paradoxical behavior. They clearly pattern as plain and tense consonants respectively in phonology, as demonstrated in such phenomena as post-obstruent tensing and the consonantal correspondences in sound symbolism, but their phonetic behavior is less clear. Building on Kagaya's characterization of laryngeal contrasts in terms of glottal width at the moment of articulatory release, Iverson (1983) concludes that the /s/ belongs to the lax category in phonology, and he claims that it even undergoes the intervoiced slackening process, if not voicing itself, as do the other lax obstruents. In contrast to its phonological behavior, /s/ could be better characterized phonetically as an aspirate. On the other hand, T. Cho *et al.* (2002) arrive at the conclusion that it is better characterized as "lenis" than as "aspirated," mainly on the strength of the observation that it is unaspirated medially and sometimes undergoes gradient voicing in a parallel fashion to stops.

[2] Regarding the place of articulation for the affricate series, there is some debate whether they are dental (H. Kim 1999) or alveopalatal, as most phonologists assume, but in this chapter I eschew the controversy and represent the affricates as /ts tsʰ ts'/. Nothing crucial hinges on this particular characterization.

(2) *Korean laryngeal contrasts exemplified*

pul	'fire'	p^hul	'grass'	p'ul	'horn'
tal	'moon'	t^hal	'mask'	t'al	'daughter'
tsa-ta	'to sleep'	ts^ha-ta	'to kick'	ts'a-ta	'to be salty'
koŋ	'ball'	k^hoŋ	'bean'	k'oŋ	'frozen'
sal	'flesh'			s'al	'rice'

Plain consonants (also called "lenis" or "lax" consonants) are produced with slight aspiration in word-initial position, while aspirated consonants are heavily aspirated. In addition, it should be noted that plain obstruents undergo systematic changes in medial positions. They become voiced between voiced segments within a phonological phrase, except for /s/, and are realized as tense consonants after any obstruent. As mentioned in footnote 1, phonetically intervoiced slackening of fricatives might be considered the same process as intersonorant voicing (T. Cho *et al.* 2002). The same slackening process in the voicing environment may be also responsible for the relatively newly emerged obligatory disappearance of intersonorant /h/, as in /tsoh-a/ → [tsoa] 'to be good-CONT'. In final position, all obstruents are unreleased and, as a result, there is no aspiration or realization of tenseness.

The tense series (sometimes termed "fortis" consonants) is characterized by some sort of "laryngeal tension." Whereas voicing and aspiration are cross-linguistically common laryngeal features for obstruent systems, having a tripartite system involving aspiration and tenseness within one language is rather unusual. This rare pattern has motivated a number of researchers to argue against the existence of a tense contrast in Korean; in more radical cases, both features of the laryngeal contrast are re-analyzed as non-laryngeal features, which will be the topic of §3. Not only tense, but also aspirated, consonants are treated as geminates, and their phonetic properties are derived from the structural configuration rather than from phonetic features.

The unsettled state of laryngeal contrasts in Korean in some ways reflects the diachronic emergence of laryngeal obstruents; according to Poppe (1965) and Starostin *et al.* (2003), the Proto-Korean consonant system lacked phonemic aspiration and tenseness. By early Middle Korean, the aspirated series had emerged – first labials and dentals, followed by velar aspirates – as evidenced by the way Chinese aspirated consonants were adopted into Middle Korean (Park 1996). By early Modern Korean, tense consonants (or geminates for those who do not assume tense consonants as a single phoneme) emerged from consonant clusters resulting from vowel deletion and the ubiquitous addition of the linking /s/ between morphemes (K. M. Lee 1985).[3]

Treating tense consonants as phonologically geminate versions of the corresponding plain consonants might have the advantage of doing away with a rather unusual phonation contrast, i.e. the coexistence of aspiration and tenseness within one and the same system, but it carries the burden of explaining why it is better to interpret surface tension as a phonetic property of length contrast derived from the geminate property.[4] The geminate account of tense consonants

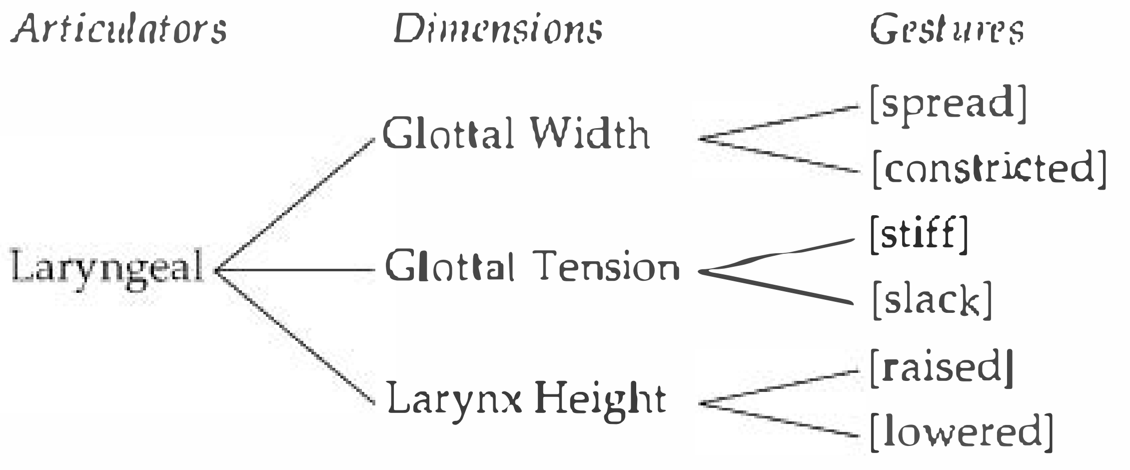
[3] Among tense consonants, the affricate /ts'/ is the last one to emerge.

[4] As an anonymous reviewer points out, there are a number of languages with a fortis/lenis contrast, such as Upper Cross languages like Lokka and Leggbo, in which there is evidence (e.g. vowel shortening before tense consonants) to treat fortis consonants as geminates.

(1971). More recently, S.-C. Ahn and Iverson (2004: 13) have drawn attention to two aspects of difference between typical ejectives and Korean tense consonants: (i) Korean in-phase constriction *vs.* persisting glottal closure of ejectives, and (ii) a raising effect on the pitch of the following vowel by Korean tense consonants, as opposed to a lowering effect by ejectives (Kingston 1985).

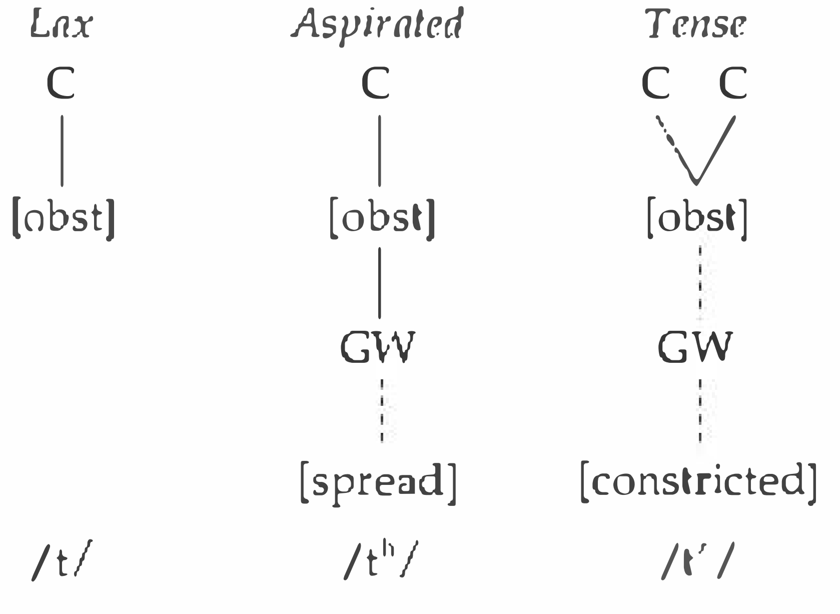
S.-C. Ahn and Iverson (2004) employ Avery and Idsardi's (2001) Dimensional Theory. In this theory, the Laryngeal articulator is defined by three dimensions, Glottal Width, Glottal Tension and Laryngeal Height, as shown in (3) (see also CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION).

(3) *Geometry of laryngeal representation in Dimensional Theory* (Avery and Idsardi 2001)



Glottal dimensions implicate antagonistic "gestures," and in particular, Glottal Width implicates two antagonistic gestures of [spread] and [constricted]. For theory-internal reasons and in an attempt to account for cross-linguistically robust laryngeal patterns, only one member of a gestural pair is stipulated to function contrastively in a given system (see also CHAPTER 2: CONTRAST). This set of assumptions predicts that no language will have tense and aspirated consonants contrasting with lax consonants at the same time. According to this theory, tense consonants in Korean are phonetically augmented with Glottal Width, resulting in surface tension, but they are characterized neither by Glottal Width nor by Glottal Tension phonologically. On the phonemic level they are laryngeally lax, with their phonetic property being derived from geminate structure, implemented by "Korean Enhancement," a language-specific redundancy. (4) illustrates the contrast with the dimension of Glottal Width only, because the other dimensions, Glottal Tension and Larynx Height, are not relevant in Korean.

(4) *Korean obstruent contrasts in Dimensional Theory* (S.-C. Ahn and Iverson 2004)



Also contributing to the complications are recently observed acoustic facts that contemporary Seoul Korean may be developing a tonal system in lieu of laryngeal contrasts (see also CHAPTER 97: TONOGENESIS). According to S.-C. Ahn and Iverson (2004), Silva (2006), and Wright (2007), over the past 40 years Seoul speakers have shown a dramatic change in their use of VOT to mark the language's distinction among tense, lax, and aspirated stops. Data indicate that while VOT for tense stops has not changed since the 1960s, VOT differences between lax and aspirated stops have decreased, in some cases to the point of complete overlap. Concurrently, the mean F0 for words beginning with tense or aspirated stops has become significantly higher than the mean F0 for comparable words beginning with plain stops.[6]

Silva's proposal of tonogenesis is quite interesting, in that it mirrors a cross-linguistically common pattern reported for Tibetan (Duanmu 1992) and Kamuu (Svantesson and House 2006), but perhaps it is too early to assume that the laryngeal contrast in Korean is being completely replaced by tones, as the new development is observed in only some speakers of Korean, and only Seoul speakers at that. Silva's position can be summarized as follows: (i) the underlying contrast between lax and aspirated stops that is maintained by younger speakers is phonetically manifested in terms of differentiated tonal melodies, rather than by laryngeal features, and (ii) laryngeally unmarked (lax) stops trigger the introduction of a default L tone, while laryngeally marked stops (aspirated and tense) introduce H, triggered by a feature specification for [stiff]. To distinguish the tense series from the aspirates, however, VOT or some sort of laryngeal features is needed, since both phrase-initial aspirates and tense stops should be marked by a High tone. In addition, the less flamboyant behavior of phrase-medial stops needs to be accounted for.[7] In order to explain this new phenomenon, Silva proposes a replacement of the feature of glottal aperture [spread glottis] and [constricted glottis] with a more abstract laryngeal "tensity feature," [stiff], in the spirit of C.-H. Kim (1965). According to Silva (2006), this single feature has the advantage of unifying the VOT distinction on the one hand and expanding to characterize the newly emerging tonogenesis.

M.-R. Kim and Duanmu's (2004) categorization of "lax" stops as voiced stops and "tense" stops as regular voiceless unaspirated stops is another attempt to do away with an additional feature of tensity as well as to explain the above-mentioned consonant–tone correlation. Their rather unorthodox classification is based on Kingston and Diehl (1994: 435), who attribute the paradoxical behavior of a voiceless unaspirated stop which may both elevate and depress F0 to the phonological specification of these phones:

> when it represents the [−voice] category, as in Hindi, Thai, Spanish, French, Portuguese, Italian, and Japanese, then F0 is elevated; but when it represents the [+voice] category, as in English, German, Swedish, Danish and Korean, then F0 is depressed.

[6] It should be noted that the affricates pattern in exactly the same way as the stops in phonology, e.g. in processes such as intersonorant voicing and accentual realization (Y. Y. Cho 1990b; S.-A. Jun 1993) as well as in morphology as manifested in sound symbolism (Y. Y. Cho 2006).

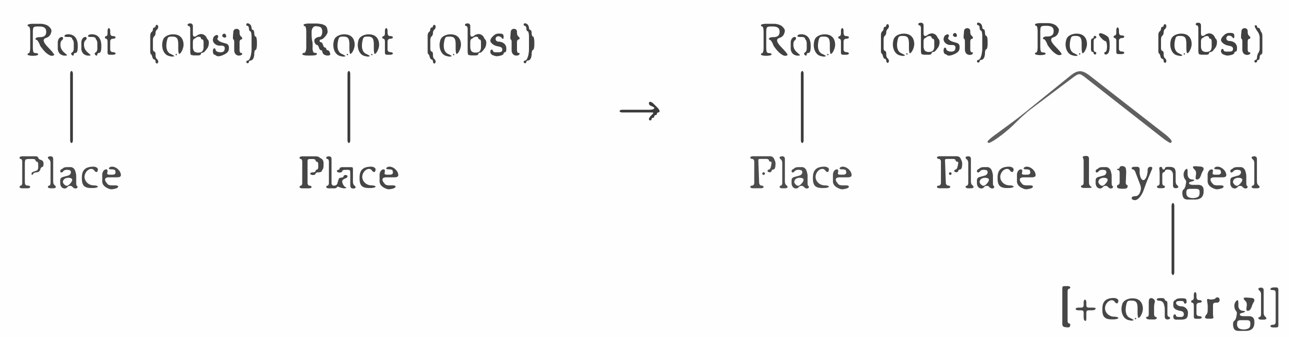
[7] Jun, in her analysis of the Chonnam Accentual Phrase (1993: 53), also observes that when the phrase-initial segment is either a tense or an aspirated obstruent, the HHL tonal pattern is chosen instead of the usual LHL. It is interesting to note here that /s/ triggers the HHL pattern in parallel with the other laryngeal obstruents.

dimension of contrast, namely gemination. Alternatively, a few more extreme unary analyses emerged where no underlying laryngeal contrast is assumed, and the so-called "tense" and "aspirated" consonants acquire their surface specifications by other means (S.-H. Kim 1989; J. Jun 1994).

The singleton analysis of tense and aspirated consonants acquires its primary support from the restrictions on Korean syllables, in particular from the fact that no clusters are allowed in onset position in Modern Korean. If tense and aspirated consonants that freely occur in syllable-initial position are singletons with a unique laryngeal specification, nothing more needs to be said with regard to tense consonants or aspirates. On the other hand, if these laryngeal consonants are geminates, a further stipulation is required about this unexpected distribution. If laryngeal consonants were geminates in Korean, these would be the only type of clusters allowed in onset position, a highly unlikely phenomenon cross-linguistically.[9] Therefore, even people who espouse the geminate nature of tense consonants treat these as singletons on the surface, with phonetic tenseness being derived from the underlying geminate structure.

J.-I. Han (1992, 1996) is worth mentioning, because her work is the most concrete and detailed in the debate. She observes that there is a significant difference in length between intervocalic tense consonants on the one hand and word-initial tense and intervocalic plain consonants on the other. Word-initial tense consonants are singletons, due to Stray Erasure, thereby avoiding the problem associated with the syllable structure condition against onset clusters. The manifestation of tenseness comes from its geminate structure; i.e. tense consonants are plain geminates underlyingly. Furthermore the relationship between length and tension is explained by stipulating that the surface tension is derived from the underlying geminate structure by a rule of Geminate Reinforcement (5), which inserts the tense feature on all geminate obstruents.

(5) *Geminate Reinforcement* (J.-I. Han 1996)



However, a problem with this account's reliance on closure duration is that, as noted by Y. Y. Cho and Inkelas (1994), not only tense consonants but also

[9] A further study is needed to support the position espoused here that all geminates are clusters. An anonymous reviewer wonders whether all languages that allow initial geminates allow non-geminate clusters initially. According to Stuart Davis (personal communication; see also CHAPTER 37: GEMINATES), there are two types of languages with initial geminates: those in which the initial geminate is the only type of initial cluster and those that allow almost any type of initial cluster, including geminates. While languages like Trukese (also called Chuukese) and Pattani Malay have initial geminates but do not have initial consonant clusters, there are languages like Leti and Moroccan Arabic that have initial geminates as well as initial consonant clusters (Muller 2001). He also points out that in languages that have geminates as well as onset clusters with sonority restrictions, the geminates typically cannot occur in initial position (e.g. Italian). This line of typological argument would not be decisive either for or against the geminate analysis of tense consonants in Korean.

aspirated consonants have been observed to be longer than plain consonants in intervocalic position (Pyo-Ih 1973; Zhi 1982, cited by S.-H. Kim 1989), although the length contrast is not as pronounced in the aspirated case (Silva 1992). According to some measurements, if duration were the primary phonetic indicator of phonological geminates, aspirated consonants would have to be considered geminates as well. These moves would lead to a drastic reduction in the number of phonemic contrasts in Korean: instead of three series of obstruents, there would be only one type of laryngeal consonant, and, moreover, all of its laryngeal members would be geminate. Furthermore, the distinction between the tense and aspirated consonants would be derived from yet another contrast, probably an unnecessary complication in the grammar. Finally, getting rid of all laryngeal contrast seems implausible in view of the cross-linguistic generalization (noted by Scobbie 1991) that in any language the inventory of geminate consonants is a subset of singleton consonants.[10]

One issue with the data in J.-I. Han (1996), as well as in subsequent publications, is what kind of speech was used in measuring the duration of intervocalic tense consonants. In particular, it has often been observed that certain intervocalic tense and aspirated consonants are phonologically geminated in emphatic speech. In particular, as illustrated in (6b), it has words whose intervocalic tense or aspirated consonant cannot be geminated.

(6) *Gemination in emphatic speech (optional)*

a. *Emphatic speech gemination*

ap'a	app'a	'Dad'
ak'a	akk'a	'a while ago'
tok'i	tokk'i	'ax'

b. *Emphatic speech gemination is not allowed in non-initial syllables*

atsəs'i	*atsəss'i	'uncle'
tsaŋats'i	*tsaŋatsts'i	'pickle'

It so happens that all of Han's examples of intervocalic tense consonants belong to the class in (6a), and can be readily geminated in slow emphatic speech. Thus it is not clear that extra duration must be attributed to the putative underlying geminate status of tense consonants, or whether it might not in fact be due to emphatic rendering of these forms by the subjects in the experiment, especially in the citation forms.

Han's other argument comes from the phenomenon of degemination, in which, as shown in (7), a sequence of homorganic plain obstruents is often simplified in allegro speech (Kim-Renaud 1987; see also CHAPTER 79: REDUCTION).

(7) *Allegro speech degemination*

ak+ki	akk'i	ak'i	'musical instrument'
tat-ta	tatt'a	tat'a	'to close'

[10] In view of the cases where there is a kind of complementarity, for instance *ww* vs. *bbw/ggw*, or Italian affricates that always surface as geminates, as pointed out by an anonymous reviewer, Scobbie's generalization may be revised to a weaker form.

contrast in emphatic speech is in itself sufficient to undermine any account in which all tense consonants are underlyingly plain geminates which tensify and degeminate by later processes.[13]

J.-I. Han (1996) uses as a piece of supporting evidence an allegedly robust typological tendency for sonorant geminates to occur with obstruent geminates in many languages, rather than having only sonorant geminates in one given system (CHAPTER 8: SONORANTS). Such would be the case in Korean, if a singleton analysis were to be adopted. However, invoking this general tendency as one of the arguments is not fully warranted at this point. While sonorant geminates are clearly distinct from a single sonorant, as exemplified in (9), the supposed minimal pairs (e.g. /tal/ *vs.* /t'al/ and /aka/ *vs.* /ak'a/) have a different status from sonorant geminates, thus undermining an attempt to establish the geminate contrast in obstruents. The short and the long forms of obstruent geminates are mere variants in the ESG environment, which the geminate account would be hard-pressed to account for.

(9) *Tautomorphemic sonorant geminates*

manna-	'to meet'	an-	'to hug'
kənnə-	'to cross'	kam-	'to wind'
nolla-	'to be surprised'	al-	'to know'
əmma	'mom'	ama	'perhaps'
pəlle	'worm'	alɛ	'below'

In sum, none of the arguments for the geminate account of tense consonants seems to be compelling enough to rule out the view of these consonants as single phonemes, at least morpheme-internally.

It has been claimed, however, that a certain explanatory economy is gained by the geminate account in morphology. What is relevant is the fact that partial reduplication (CHAPTER 100: REDUPLICATION) of bases with tense and aspirated onsets results in the loss of tenseness and aspiration on the second syllable, as illustrated in (10).

(10) *Loss of laryngeal features in partial reduplication* (J. Jun 1994)

tsulu	tsulu-lu-k	'falling rain'
asak	asa-sa-k	'crunchy'
tʰaŋ	tʰa-ta-ŋ	'bang'
p'aŋ	p'a-pa-ŋ	'exploding'
ts'ik	ts'i-tsik	'(the sound of) tearing'

In a pioneering theoretical approach, J. Jun (1994) proposes a generalized account of Korean reduplication, based on the assumptions listed in (11).

[13] In order to see if there is a clear durational difference between lax and tense consonants in intervocalic position that cannot be reduced to ESG, we need data where the laryngeally marked consonants appear in contexts where ESG cannot apply. The currently available studies examine consonants in CVCV context, and all involve a list of citation forms obtained in an experimental setting, rather than from natural speech.

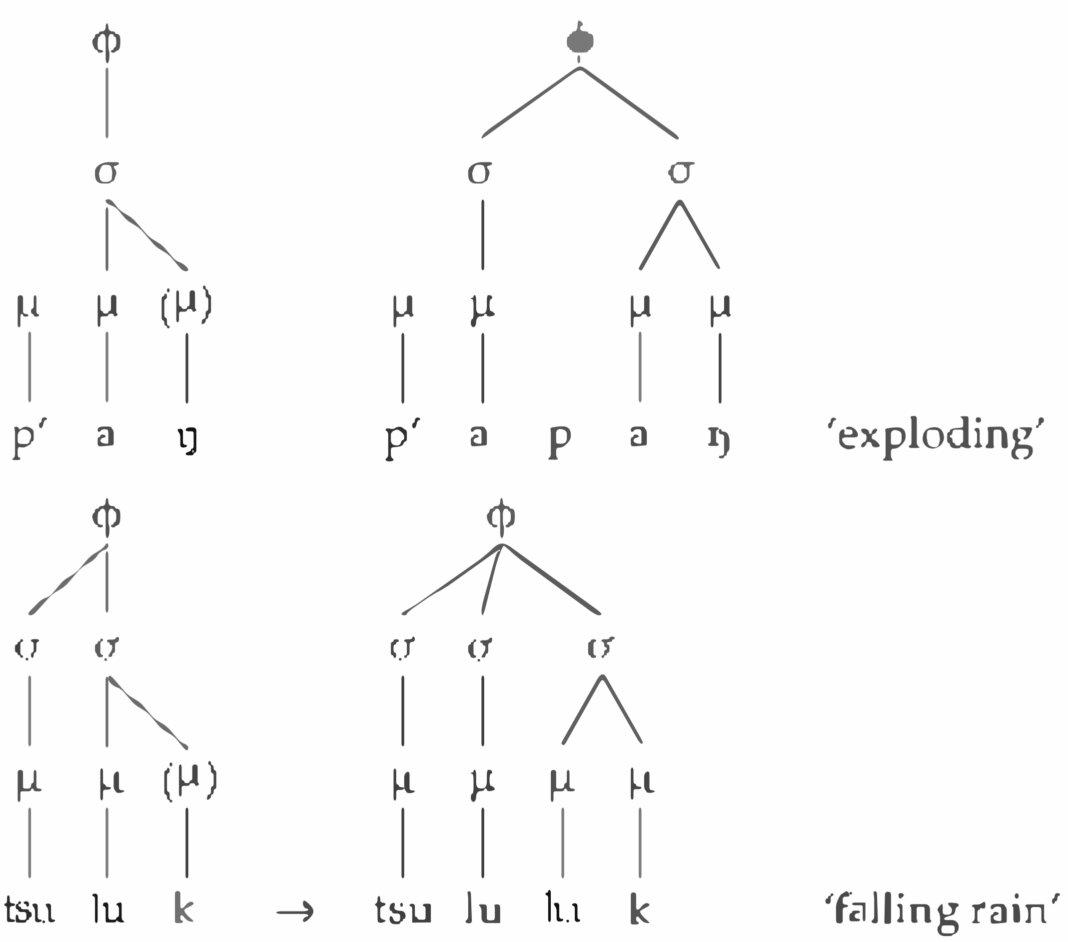
(11) *Phonological assumptions*

a. In both input and output, a closed syllable constitutes a single foot.
b. The final base syllable is suffixed, with its coda deleted.
c. Laryngeal consonants are geminates.
d. Coda consonants are moraic.

His analysis of reduplication crucially relies on the principle of Metrical Weight Consistency, which requires that the number of metrical feet in the input be conserved in the output. This principle is based on the following assumptions about foot structure, moras, and syllable weight. The base for reduplication consists of one foot with a final heavy syllable, and the output also observes the same weight constraint. He assumes the Korean foot to be quantity-sensitive, right-headed, unbounded, and constructed from the left edge, without any motivation from the rest of the Korean grammar. Moreover, treating tense and aspirated consonants as geminates is essential in calculating metrical weight. In this account, the loss of the laryngeal feature in partial suffixal reduplication (e.g. /p'aŋ/ → [p'a-paŋ], /tʰaŋ/ → [tʰa-ta-ŋ]) is accounted for by demorification (the loss of a mora), so that the output can remain as a single foot, with any number of light syllables occurring before the last mora-bearing consonant.[14] Laryngeal feature loss in the reduplicant is an automatic consequence of applying metrical weight consistency. All the inputs to partial reduplication contain a single foot (either a single heavy syllable or a disyllabic stem consisting of a light syllable and a heavy syllable). As a consequence, the output words should contain only one foot. If the output were to preserve the laryngeal features on the second syllable, as in *[p'a-p'aŋ], then the tense consonant would be moraic, making the output consist, incorrectly, of two feet.

Based on (11), some derivations are shown in (12).

(12) a. / b.



[14] One piece of evidence for this is multiple outputs such as [p'a-paŋ], [p'a-pa-paŋ], [p'a-pa-pa-paŋ], [p'a-pa-pa-pa-paŋ], etc. There are other ways to handle iteration in other approaches; one straightforward way would be to have reduplication reapply to its own output.

It is, however, extremely plausible to have an optimality account of partial reduplication whereby the loss of laryngeal features is merely a case of emerging unmarked constraints by relevant constraint ranking. In fact, the Emergence of the Unmarked (TETU; CHAPTER 58: THE EMERGENCE OF THE UNMARKED), the name given to the striking tendency of reduplicants to be "less marked" in some dimension than in the language generally, is an effect that helps elucidate phonological constraints of Universal Grammar in Optimality Theory (McCarthy and Prince 1993, 1994; Prince and Smolensky 1993). There are numerous examples of TETU in the literature that are motivated by the very architecture of Optimality Theory. For Korean partial reduplication, there is consequently no need to represent tense and aspirated consonants as geminates, nor any need to invoke J. Jun's foot type (quantity-sensitive, unbounded, and right-headed), which seems to play no role in other parts of Korean phonology. Like other TETU cases, a particular ranking produces a TETU effect of selecting a reduplicant without a coda consonant. This loss of laryngeal features is accomplished by sandwiching the markedness constraint, NoLaryngeal, between IO-faithfulness and BR-faithfulness; Ident-IO protects underlying laryngeal features, whereas the ranking NoLaryngeal >> Ident-BR[Lar] chooses the output without laryngeal specification, as in [t'a-ta-k]. The TETU effect, which militates against laryngeal features, is accounted for by the relative ranking of markedness and faithfulness constraints, in exact parallel to other TETU cases. When re-analyzed from this theoretical angle, there still seem to be no compelling arguments from Korean morphophonology for the geminate analysis of laryngeal consonants.

As laid out in the preceding discussion, researchers differ in their analyses of the two types of laryngeal consonants, some arguing that (i) tense consonants should be treated as singletons on a par with plain consonants, differing only in the specification of laryngeal features, and others that (ii) tense consonants pattern with aspirated consonants. Even though there are many parallels between tense and aspirated consonants, most studies readily accept the existence of an aspiration contrast, but not that of a tense contrast. This is due mainly to the phonetic characteristics of tense consonants and the typological rarity of a system with a simultaneous aspiration/tense contrast, but not voicing, as detailed in §2 (Kingston and Diehl 1994; S.-C. Ahn and Iverson 2004; M.-R. Kim and Duanmu 2004). Given that a substantial number of arguments for the geminate analysis apply to both of the laryngeal consonants, the burden falls on the proponents of the geminate analysis to clarify how the two types of laryngeal consonants pattern together or differently, and to choose in a principled manner which patterns are relevant for establishing the laryngeal contrast in Korean.

In the discussion of laryngeal contrasts, Korean Post-Obstruent Tensing (POT)[15] is quite relevant. Plain obstruents in Korean undergo tensification immediately after another obstruent (Kim-Renaud 1974), as illustrated in (13a).[16] The obstruents in (13b) are not subject to POT, since they are not preceded by an obstruent.

[15] POT should be distinguished from other processes of tensification such as post-nasal tensing in verbal morphology (e.g. /an-ta/ → [ant'a] 'to sit-verbal ending') and post-lateral coronal tensing in the Sino-Korean vocabulary (e.g. /il-tsa/ → [ilts'a] 'one word').

[16] As noted earlier, the behavior of lenis fricative /s/ is ambiguous. It undergoes POT, as in /pat-so/ → [pats'o] 'to receive-verbal ending'; but it is not voiced in an inter-sonorant environment, unlike the other plain obstruents, although Iverson (1983) regards the slackening of [s] as the same process of voicing.

Following Cook (1987), Y. Y. Cho and Inkelas (1994) assume that the so-called "*Bindungs*-/s/," which applies in the sub-compounding construction in (15b), involves morphological gemination. This is supported by the fact that the effects of the *Bindungs* element appear only when the second member of the compound begins with a consonant.[18] Note that even those morphologically derived geminates are, due to a preceding consonant, simplified to singletons and surface as tense. From this it can be concluded that POT is fed by gemination, as in [mukk'uk] and [pamt'otuk], regardless of whether or not the geminates which undergo POT ultimately survive to the surface.

A closely related problem has to do with the domain of Post-Obstruent Tensing. As observed in Y. Y. Cho and Sells (1995), the domain of POT is the phonological word.[19] As illustrated in (16), tensing applies within underived words, as in [kaks'i] 'bride', within lexical compounds, as in [talk k'ogi] 'chicken meat', between a clitic and its host (as in [hal s'u] 'ability to do' and [kal p'a] 'way to go'), and between a stem and a case marker or a verbal suffix. It is, however, blocked between two phonological words, as shown in (16d).

(16) *Post-Obstruent Tensing as word-level process* (Y. Y. Cho and Sells 1995)

a.	*Monomorphemic forms*			
	kaksi	→	kaks'i	'bride'
	maktɛ	→	makt'ɛ	'stick'
b.	*Compounds*			
	talk+koki	→	talkk'ogi	'chicken meat'
	pat^h+kolaŋ	→	patk'oraŋ	'furrow'
c.	*Host + clitic*			
	halC su	→	hal s'u	'ability to do'
	kalC pa	→	kal p'a	'way to go'
d.	*Phrasal combinations*[20]			
	k'amp'ak kolat'ɔlɔtsita	→	ka'amp'ak korat'ɔrɔdzida	'doze off'
	pap tewə	→	pap tewə	'warm the rice'

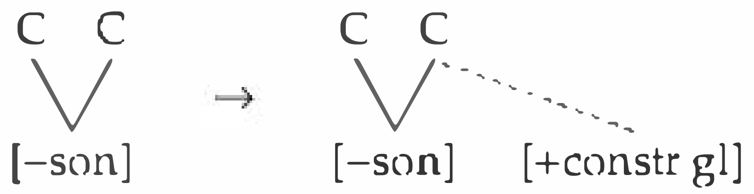
Again, if Han is to be able to relate the two tensification environments and capture Post-Obstruent Tensing and "Geminate Reinforcement" as the same process, the facts in (15) would require that Geminate Reinforcement (GR) precede POT, and therefore that it apply at the word level or earlier. This is in conflict with Han's claim that GR is a late phonetic rule. J.-I. Han (1996) collapses GR and POT as a single process in response to this particular opposition. Two separate rule formations in (17) are unified as the generalized form in (5). GR and POT each target geminates and singletons separately and are completely unrelated, although the effect in both cases is to insert the tense feature ([+constricted glottis]) on an obstruent in the post-obstruent position.

18 The only exception to this generalization involves the morpheme /u/ 'upper' as in /u+əlini/ → [utɔlin] 'elders'.

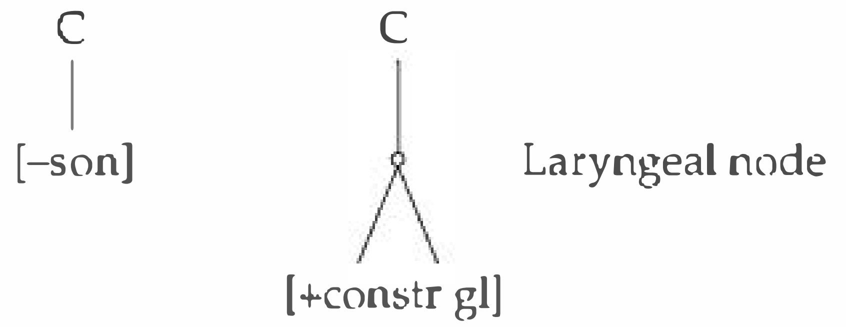
19 Choi (1991) claims that POT is a post-lexical rule whose domain is larger than a phonological word, but all of his data involve cases in which plain and tense consonants occur in free variation in casual speech and in certain dialectal forms, as in /tsuŋkuk/ and /ts'uŋkuk/ 'China', /təntsita/ and /t'əntsita/ 'to throw'.

20 S.-A. Jun (1993) proposes that POT is bound by Accentual Phrase and is attested across word boundaries in close phrasal combinations.

(17) a. *Geminate Reinforcement*

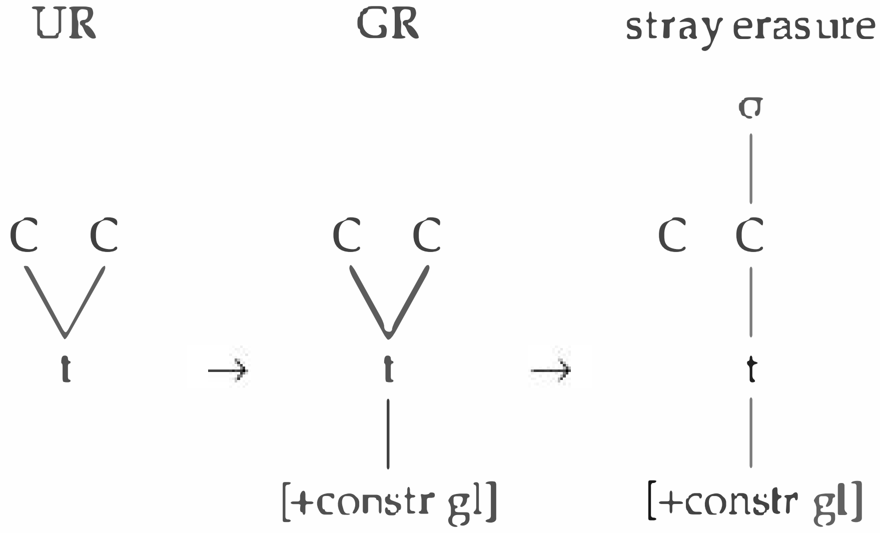


b. *Post-Obstruent Tensing*



As a process of phonetic implementation, GR interacts with stray erasure for the purpose of syllabification, and simplifies abstract geminates to their surface singleton counterparts in word-initial position, as shown in (18):

(18) *Geminate Reinforcement precedes stray erasure* (J.-I. Han 1992)



As a consequence, stray erasure must also be treated as a late phonetic rule. However, this is at odds with the fact that Korean syllabification is a cyclic process (CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME). Inasmuch as stray erasure is connected to syllabic well-formedness, it ought also to apply cyclically. There is some positive evidence for this as well. As argued in Kang (1992) and E. Han (1993), the domain of lexical syllabification is a sub-lexical constituent consisting of a stem and a suffix. Prefixes and the individual members of a compound form separate syllabification domains, as illustrated by the example /nəks+əps-i/ → [nəkəps'i] 'listlessly'. In this form, the unsyllabifiable cluster [ks] is simplified, because it precedes a compound juncture and is thus domain-final; [ps], by contrast, is not simplified, as it precedes a suffix with which it can syllabify. The fact that stray erasure, which is responsible for the simplification of obstruent clusters, is sensitive to morphological boundary distinctions confirms its lexical status and may be used to argue against any analysis that treats stray erasure as a late phonetic process.[21]

[21] An anonymous reviewer suggests an analysis whereby syllabification is bound by prosodic word, but stray erasure only applies at the end of derivation. This account goes against the received view of the morphology–phonology interaction, where earlier-level processes are closed off and are opaque to following operations (Kiparsky 1982; Inkelas 1990, forthcoming; Anttila 2002). Until an alternative theory is put forward to supersede this view of morphology–phonology interface, this option is not tenable.

4 Conclusion

The previous sections described two opposing positions regarding laryngeal contrast in Korean. One camp advocates treating tense consonants (and in some cases aspirated consonants as well) as geminates, thus eliminating the distinctive feature(s) referring to tenseness (and aspiration in some cases) in phonology and obtaining the phonetic realization of tenseness as a by-product. The other camp assumes that Korean obstruents manifest a three-way distinction, which cannot be further reduced, thus refuting the evidence from historical evolution and phonetic idiosyncrasy. When a three-way laryngeal distinction is maintained, the simple onset structure of Korean syllables can be maintained without further stipulation, while the economy gained in the account of laryngeal contrast needs to be compensated by complications in the treatment of syllabification and the different behavior of plain *vs.* laryngeal consonants. Some arguments for both positions are more theory-internal than others, but we can safely conclude this chapter with the following observations.

(23) *Generalizations about Korean laryngeal contrast*

a. Having tense and aspiration in one system is quite unusual cross-linguistically, thus motivating some researchers to do away with tense consonants in phonology, although there is no systematic study that argues for removing both tense and aspiration features from phonology.[24]

b. Phonetic details on Korean laryngeal contrast are quite complicated; such considerations as glottal width, laryngeal tension, VOT, and even High-tone association have been used to argue for a certain phonological position.

c. Tense and aspirated consonants are not always treated in the same manner in the geminate account; some treat only the former as geminates, but others treat both as geminates. In the latter case, no explanations have been offered to account for the total absence of laryngeal features in Korean.

d. Treating tense (and sometimes aspirated) consonants as geminates purely on the basis of phonetic length is ambiguous, and comparing length in initial *vs.* medial positions is not yet well established.

e. Arguments based on moraic distinction between singletons and geminates in Korean seem untenable, and thus cannot be used to support the position that laryngeal consonants are structurally geminates.

f. Laryngeal consonants in a particular medial position can be lengthened to close the previous syllable, but because of the restricted environment and the optional nature of gemination, this fact cannot be used to argue for tense consonants as geminates across the board.

[24] Keren Rice (personal communication) points out that it is not unusual to have tenseness and aspiration in one system when tenseness involves glottalization, but it is clear that Korean tense consonants are not glottalized.

g. Post-Obstruent Tensing, a robust phenomenon in Korean, is treated as the same process of Geminate Reinforcement (or Korean Enhancement) by the proponents of the gemination analysis. While it is true that all obstruents in post-obstruent position are tensed, it still remains to be demonstrated that all instances of tense consonants can be derived from POT.

REFERENCES

Ahn, Hyunkee. 1999. Post-release phonatory processes in English and Korean: Acoustic correlates and implications for Korean phonology. Ph.D. dissertation, University of Texas, Austin.

Ahn, Sang-Cheol. 1985. The interplay of phonology and morphology in Korean. Ph.D. dissertation, University of Illinois, Urbana.

Ahn, Sang-Cheol & Gregory K. Iverson. 2004. Dimensions in Korean laryngeal phonology. *Journal of East Asian Linguistics* 13. 345–379.

Anttila, Arto. 2002. Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20. 1–42.

Avery, Peter & William J. Idsardi. 2001. Laryngeal dimensions, completion and enhancement. In T. A. Hall (ed.) *Distinctive feature theory*, 41–70. Berlin & New York: Mouton de Gruyter.

Baek, Eung-Jin. 1991. Unreleasing in Korean: A phonetic explanation. *Harvard Studies in Korean Linguistics* 4. 29–35.

Cho, Taehong, Sun-Ah Jun & Peter Ladefoged. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30. 193–228.

Cho, Young-mee Yu. 1990a. The parameters of consonantal assimilation. Ph.D. dissertation, Stanford University.

Cho, Young-mee Yu. 1990b. Syntax and phrasing in Korean. In Sharon Inkelas & Draga Zec (eds.) *The phonology–syntax connection*, 47–62. Chicago: Chicago University Press.

Cho, Young-mee Yu. 1997. Liquid specification in Korean as geminate alterability. *Harvard Studies in Korean Linguistics* 7. 78–92.

Cho, Young-mee Yu. 2006. Sound symbolism in Korean. In Ho-min Sohn (ed.) *The Korean language in culture and society*, 64–73. Honolulu: University of Hawaii Press.

Cho, Young-mee Yu & Sharon Inkelas. 1994. Post-obstruent tensification in Korean and geminate inalterability. In Young-Key Kim-Renaud (ed.) *Theoretical issues in Korean linguistics*, 45–60. Stanford: CSLI.

Cho, Young-mee Yu & Peter Sells. 1995. A lexical account of inflectional suffixes in Korean. *Journal of East Asian Linguistics* 4. 119–174.

Choi, Eun Young. 1991. Postlexical phonology in Korean. Ph.D. dissertation, University of Washington.

Cook, Eung-Do. 1987. *Sai-sios* is a geminated consonant. *Harvard Studies in Korean Linguistics* 2. 360–366.

Dart, Sarah N. 1987. An aerodynamic study of Korean stop consonants: Measurements and modeling. *Journal of the Acoustical Society of America* 81. 138–147.

Duanmu, San. 1992. An autosegmental analysis of tone in four Tibetan languages. *Linguistics of the Tibeto-Burman Area* 15. 65–91.

Guerssel, Mohamed. 1978. A condition on assimilation rules. *Linguistic Analysis* 4. 225–254.

Halle, Morris & Kenneth N. Stevens. 1971. A note on laryngeal features. *MIT Research Laboratory of Electronics Quarterly Progress Report* 101. 198–213.

Han, Eunjoo. 1993. Prosodic structure in compounds. Ph.D. dissertation, Stanford University.

112 French Liaison

Marie-Hélène Côté

1 Introduction

Liaison is a type of external sandhi, one of several occurring at word boundaries in French. It corresponds to the pronunciation of a consonant (liaison consonant = LC) between two words, the second being vowel-initial, in certain liaison-triggering contexts. Liaison may apply categorically or variably, depending on the context. The data in (1) illustrate this process. The words preceding and following the LC are referred to respectively as Word1 and Word2; their canonical pronunciation in non-liaison contexts, including in isolation, is indicated in the relevant columns. In (1b), for example, the words *vous* and *allez* are pronounced in isolation [vu] and [ale]; when the two words are concatenated, an intervening [z] surfaces. LCs are underlined in all examples for ease of identification.[1]

(1)

			Word1	*LC*	*Word2*	*Context*
a.	*un homme*	'a/one-MASC man'	[œ̃	n	ɔm]	det + noun
b.	*vous allez*	'you-PL go'	[vu	z	ale]	pronoun + verb
c.	*grand ami*	'big-MASC friend'	[gʁɑ̃	t	ami]	adj + noun
d.	*très utile*	'very useful'	[tʁɛ	z	ytil]	adv + adj

Most LCs originate from fixed final consonants that progressively ceased to be pronounced between the twelfth and the sixteenth century, first before a consonant, later at the pause. These consonants remained only in prevocalic position between words displaying a high degree of cohesion. Some LCs also evolved through analogy as part of restructuring processes affecting liaison (see Morin 1986 for a discussion of final consonants in the history of French). The system of liaison has been in evolution for almost a millennium and is still subject to change and reinterpretation.

[1] In most cases, only word-by-word translations are provided. Grammatical information such as PL and MASC is added when relevant, if it is not already clear elsewhere in the example. The word-by-word translation is followed by a free translation (introduced by "=") when the literal translation is opaque or when minimal pairs with the same lexical material but different meanings are involved.

Liaison is a complex and multifaceted phenomenon, involving various components of linguistic analysis. As a segmental process that finds its historical source in unexceptional cases of coda reduction, it is considered essentially phonological. But its synchronic conditioning goes much beyond phonology as traditionally defined. Liaison in contemporary French raises a number of challenging descriptive and analytical issues, pertaining to the status of LCs and the segmental, prosodic, and morphosyntactic contexts in which they surface. It has figured prominently in a number of important theoretical debates in the last half-century, including abstractness in phonology, autosegmental representations, the syntax–phonology interface, and, more recently, the roles of frequency and usage.

The analysis of liaison is complicated by the variability and controversial nature of some of the data on which it is based (see Morin 1987). Even within what may be referred to as Standard French (see Francard *et al.* 2000–01), the use of liaison is subject to substantial sociolectal and stylistic variation (CHAPTER 92: VARIABILITY). Dialectal differences are also observed. Formal accounts of liaison tend to rely on a relatively limited set of conventionalized generalizations. Standard adult data and theory-internal arguments have been increasingly complemented, however, with results from a variety of perspectives: corpus studies of spontaneous and formal French, acquisition, psycholinguistics, phonetics, and liaison errors in adults. This empirical renewal has led to a more nuanced understanding of the functioning of liaison, its conception as a unified phonological process having partly given way to a multidisciplinary and fragmented approach.

The chapter is organized as follows. §2 provides basic data and concepts, organized around two dimensions of liaison: the LCs, and the contexts in which they occur. This presentation sets the stage for the discussion of two central analytical issues in the study of liaison: the nature and status of LCs (§3), and the characterization of liaison contexts (§4). On each topic the main theoretical proposals are introduced, and their merits and shortcomings are discussed with regard to the range of empirical results available.

2 Basic facts and concepts

2.1 *Liaison consonants*

LCs are special in at least two ways. First, they are only allowed in certain segmental and grammatical environments, while other French consonants are stable across contexts (e.g. *net* 'clear' is invariably pronounced [nɛt] with an initial [n] and a final [t]).[2] Second, all consonants are possible as stable segments, whereas LCs are restricted to a small set of consonants, essentially [z n t] (with [n] appearing only after nasalized vowels). Residual LCs include [ʁ p], which together account for less than 0.2 percent of all cases of liaison in a corpus of

[2] Final consonants are implicated in two variable processes independent from liaison: optional final consonants in some words (e.g. *but* 'goal' [by] ~ [byt]) and variable cluster reduction (e.g. *couple* 'couple' [kup(l)]). Numbers between *cinq* 'five' and *dix* 'ten' behave idiosyncratically: their final consonant is stable at the pause but subject to variable preconsonantal deletion and/or prevocalic voicing. These special items will not be considered here; see CHAPTER 36: FINAL CONSONANTS for general discussion of final consonants.

spontaneous speech (Mallet 2008: 212). Liaison in [ʁ] occurs after infinitive forms and a handful of prenominal adjectives ending in *-er* (e.g. *premier* 'first'); [p] only appears after the two adverbs *trop* 'too much' and *beaucoup* 'a lot'.[3]

LCs that follow the prescriptive norm are always represented in the orthography, usually as permanent final letters of Word1: liaison [z] corresponds to written <s>, <x>, or <z>, liaison [n] to <n>, and liaison [t] to <t> or <d>. Liaison between verbs of the *-er* conjugation and enclitic pronouns is special in being marked orthographically by consonants that only appear in liaison (e.g. *vas-y* 'go there' [vazi], *va-t-il* 'goes he' [vatil] *vs. va* 'go' [va]). Non-standard cases of liaison, many of which are mentioned in the following sections, depart from the standard written form (e.g. *cent amis* 'a hundred friends' pronounced [sɑ̃zami], *donne-m'en* 'give me of-it' surfacing as *donne-moi-en* [dɔnmwazɑ̃]).

The relative frequency (CHAPTER 90: FREQUENCY EFFECTS) of the three LCs [n t z] has been debated (see Ranson 2008). Mallet (2008: 213) provides measures based on a corpus of conversational speech containing more than 28,500 potential contexts for liaison, with a global rate of realization of 45 percent. Potential contexts involve, in that order, [z], [t], and [n] (2a); however, [n] is realized more often than [t] (2b). This reversal results from the fact that [n] is implicated essentially in contexts of categorical liaison, and [t] in cases of variable liaison (2c).

(2)		[z]	[t]	[n]
a.	% of all potential instances of liaison	49%	28%	19%
b.	% of all realized instances of liaison	46%	15%	39%
c.	Rate of realization	43%	23%	90% (Mallet 2008)

Both LCs and stable final consonants are subject to a regular process of *enchaînement* or (re)syllabification before a following vowel, favoring open syllables and CV sequences at word boundaries. The example in (1b) normally syllabifies as in (3a), and a similar example with a stable final consonant appears in (3b).

(3)	a.	LC	*vous allez*	'you-PL go'	[vu.za.le]	cf. *vous* [vu]
	b.	stable final C	*douze allées*	'twelve alleys'	[du.za.le]	cf. *douze* [duz]

Liaison may also apply without *enchaînement*, the LC syllabifying with the preceding segment, as in [vuz.(ʔ)a.le] (Encrevé 1988; Laks 2009). As noted by Durand and Lyche (2008), however, liaison without *enchaînement* is found almost exclusively in read speech or elevated discursive styles (e.g. formal presidential speeches), and essentially in cases of variable liaison. It is highly exceptional in spontaneous speech. The status and theoretical significance of liaison without *enchaînement* have been debated; it has been considered on a par with liaison with *enchaînement* (Encrevé 1988), simply irrelevant (Scullen 1993), or marginal and directly related to spelling and to a specific type of public discourse (Laks 2005, 2009; Côté 2008).

[3] Historically, LCs included [k], a consonant that is clearly no longer productive (see Morin 1987: 819), but liaison with [g] may be observed after the prenominal adjective *long* 'long-MASC'. Quebec French also has [l] as a possible LC, which surfaces between proclitics ending in [a] and vowel-initial verbs, as in *ça arrive* 'this happens' [salaʁiv] (Morin 1982).

The length of the following sequence may also influence the frequency of liaison, although this effect remains to be fully tested. Morin and Kaye (1982) offer the contrast in (9) and suggest that liaison between *mangent* and *après* is more natural in (9a) than in (9b). Post (2000) observed in her reading task more liaison when the following context was non-branching than when it was branching.

(9) a. *Ils travaillent d'abord et mangent après.*
'They work first and eat after.'
b. *Ils mangent après qu'ils aient fini leur travail.*
'They eat after that they have-SUBJ finished their work.'

It has been suggested that liaison is more frequent if Word1 ends in a vowel than a consonant, and more frequent after one consonant than after two (Delattre 1955); liaison is then predicted to apply with decreasing likelihood from (10a) to (10c). This generalization would follow from a constraint against hiatus or empty onsets (e.g. Steriade 1999; Davis 2000; Tranel 2000; CHAPTER 61: HIATUS RESOLUTION; CHAPTER 55: ONSETS), which favors liaison intervocalically and when *enchaînement* is less likely to occur. Morin (2005), however, argues against these factors being a driving force in the evolution and application of liaison, and Ågren's (1973) corpus study does not confirm the hierarchy of frequency in (10).

(10)	a.	*ponts étroits*	'bridges narrow'	[pɔ̃z̩etʁwa]
	b.	*murs étroits*	'walls narrow'	[myʁz̩etʁwa]
	c.	*portes étroites*	'doors narrow'	[pɔʁtz̩etʁwat]

2.2.3 *Lexical factors*

Keeping the number of syllables, word class, and syntactic configuration constant, there remains substantial lexical variation in the frequency of liaison. For example, the proportion of realized liaison after four monosyllabic adverbs in two corpora is given in (11) (de Jong 1994; Mallet 2008; see also Encrevé 1988 and Laks 2009). The variation is extreme and obviously involves factors independent from the phonological structure of Word1. Perhaps even more telling is the comparison between different monosyllabic forms of the verb *être* 'be', which trigger liaison at proportions varying between 0 percent and 71 percent (de Jong 1994; see also Ågren 1973).

(11)	*très* 'very'	*plus* 'more'	*bien* 'well'	*pas* 'not'
Mallet (2008)	97%	64%	43%	1%
de Jong (1994)	99%	96%	82%	7%

These statistics illustrate the extent to which liaison is lexicalized and cannot be reduced to independent structural factors. Lexical distinctions in the realization of liaison have been significantly correlated with Word1 frequency, more frequent words triggering liaison more often than less frequent ones (de Jong 1994; Fougeron *et al.* 2001a; Fougeron *et al.* 2001b; see also CHAPTER 90: FREQUENCY EFFECTS).

2.2.4 *Stylistic and dialectal variation*

The use of liaison is highly dependent on speech style and register, more elevated styles being typically associated with an increased rate of liaison realization (e.g. Lucci 1983; Moisset 2000; Fougeron *et al.* 2001a). If, for example, liaison with plural prenominal adjectives (6a) is almost categorical across styles, liaison after plural nouns (6f) is exceptional in colloquial speech and frequent only in formal production. In some contexts not listed in (5)-(6), liaison may be allowed exclusively in hyperformal registers. According to Fouché (1959: 443) and Delattre (1947), a LC may intervene between a plural subject and its inflected verb (12), but Durand and Lyche (2008) mention that liaison was never realized in this context in a reading task (a context that strongly favors the realization of liaison) performed by 100 speakers.

(12) *les gens obéirent* 'the-PL people obeyed' [leʒɑ̃z̲obeir]

Laks (2009) compares the rates of liaison realization following 28 invariable words belonging to categories (6c) and (6d) in two stylistically contrasting corpora: formal political speeches and informal conversation. The global rate of realized liaison is 78 percent in the former against 47 percent in the latter, but the distinction may be more dramatic for individual words, e.g. 78 percent against 1.5 percent after the adverb *pas* 'not'.

Dialectal variation has not inspired as much research as the sociolinguistics of liaison, and the extent of this variation remains to be more fully documented. Variation essentially concerns liaison contexts: their classification, the nature of the LC that appears in each of them, and the frequency of liaison in specific variable contexts (see notes 3 and 7 for other aspects of dialectal variation). In Quebec French, for instance, one of the better-described non-standard varieties (see Côté 2010 for references), liaison after the proclitics *on* 'we/one' and *ils/elles* 'they-MASC/FEM' (5b) is variable rather than categorical (de Jong 1993) and liaison in [t] occurs after all forms of the present tense of *être* 'be', including those that trigger liaison in [z] in the standard variety (e.g. *tu es assis* 'you are seated' [tet̲asi] rather than [tyez̲asi]; van Ameringen and Cedergren 1981).

3 The phonological and lexical status of LCs

LCs, by definition, appear between two words and their relationship to each of these words has remained a matter of debate. Five main options for the lexical affiliation of LCs have been considered. They are summarized in (13), with a schematic lexical representation of the elements contained in the sequence *deux amis* 'two friends' [døz̲ami].

(13)
a. LCs are final consonants of Word1
   i. Truncation analysis /døz/ /ami/
   ii. Suppletion analysis /dø, døz/ /ami/
   iii. Autosegmental analysis /dø(z)/ /ami/
b. LCs are epenthetic /dø/ /ami/
c. LCs are initial consonants of Word2 /dø/ /zami/
d. LCs are morphemic
   i. Prefix of Word2 /dø/ /z + ami/
   ii. Suffix of Word1 /dø + z/ /ami/
e. LCs are part of a (partially) lexicalized contruction /dø z ami/

Non-standard or marginal instances of liaison and experimental evidence have played an important role in supporting or arguing against specific proposals. Observed gaps between normative descriptions and actual usage correspond either to processes established in the grammar of non-standard (spoken) varieties or to less systematic "mistakes" often characterizable as hypercorrective. The latter type tends to occur when speakers perform under some normative pressure (e.g. non-professional speakers in media contexts), which favors liaison, and in cases of variable liaison that are not used productively in colloquial registers (see Desrochers 1994 for a typology and analysis of a corpus of liaison errors). §3.1 reviews the approaches in (13), while §3.2 specifically addresses experimental evidence drawn from aquisition, phonetics, and psycholinguistics. The relationship between liaison and morphological relatedness, relevant to the nature of adjectival LCs, is discussed in §3.3.

3.1 Competing approaches

3.1.1 LCs as final consonants of Word1

The vast majority of studies on liaison adopt the classic position in (13a), according to which LCs belong to Word1, mirroring the historical origin (and the modern spelling) of most LCs. This approach has been implemented in a number of ways, which have in common that the lexical form of *deux* contains a final /z/, but differ with respect to the representation of that consonant and the status of the non-liaison form [dø] (see Klausenburger 1984 and Encrevé 1988: ch. 3 for reviews). Three main groups of analyses may be identified: truncation, suppletion, and autosegmental, presented roughly in chronological order. All must establish a distinction between final stable consonants and those pronounced only in liaison (e.g. the liaison [z] in *deux* 'two' *vs.* the stable [z] in *douze* 'twelve').

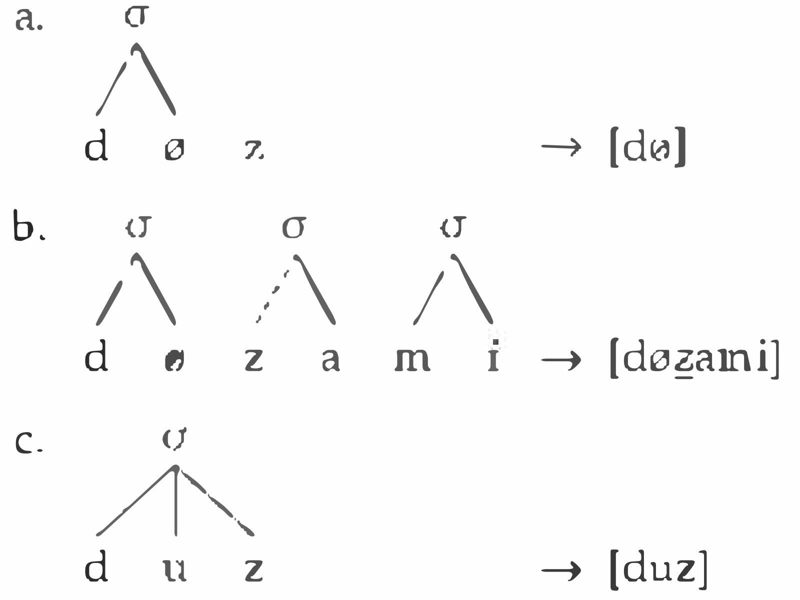
The truncation analysis coincides with the early phase of generative phonology (Milner 1967; Schane 1968; Selkirk and Vergnaud 1973; Selkirk 1974). It relies on a general rule that deletes word-final consonants in non-liaison contexts. Liaison, then, results from the non-application of deletion (CHAPTER 68: DELETION). Stable final consonants escape the truncation rule either because they are lexically specified as exceptions to it or because they are followed by a final ("protective") schwa. This schwa makes the rule inapplicable and is itself subject to a later deletion rule. One intriguing feature of this analysis is that liaison forms, which may be considered marked, result from the simplest derivation, while non-liaison forms and stable final consonants, which are in no way exceptional, involve deletion rules or special devices.

Suppletive analyses (Gaatone 1978; Klausenburger 1984; Green and Hintze 1988; de Jong 1994; Perlmutter 1998; Plénat 2008) consider that the liaison and non-liaison forms of a word are listed separately in the lexicon (e.g. /dø, døz/ for *deux*). Words ending in a stable consonant have a single allomorph (e.g. /duz/ for *douze*) (CHAPTER 99: PHONOLOGICALLY CONDITIONED ALLOMORPH SELECTION).[9]

[9] Suppletion may not characterize liaison in general but may be limited to prenominal adjectives with a different vowel in their liaison and non-liaison forms, as illustrated in (8) (Tranel 1990; Scullen 1993; Côté 2005). Adjectival LCs may also be indirectly suppletive in being derived from the listed feminine form (Steriade 1999).

The development of non-linear phonology has renewed the interest in final LCs, represented as floating segments with respect to the skeleton (CHAPTER 54: THE SKELETON), the syllable, or both (Booij 1983; Clements and Keyser 1983; Hyman 1985; Encrevé 1988; de Jong 1990, 1994; Tranel 1990, 1996, 2000; Prunet 1992; Scullen 1993; Paradis and El Fenne 1995; Davis 2000; Wauquier-Gravelines 2005, 2009; Nguyen *et al.* 2007; cf. Tranel 1995a, 1995b for an elaborate discussion and comparison of the different autosegmental accounts). Consonants left floating remain unpronounced, as is the case for *deux* produced in isolation (14a). Floating segments make it to the surface only when properly inserted into the prosodic structure, typically by filling a following empty onset position, as in (14b). As in truncation approaches, liaison and non-liaison allomorphs are derived from a unique underlying form. Fixed and liaison consonants, however, are representationally distinct, fixed ones being prosodically anchored in the lexicon, as shown in (14c) for *douze* 'twelve'. This representational approach simultaneously avoids specific deletion rules and the exception devices needed for stable consonants, but it involves more complex lexical representations (see Côté 2008 for arguments against the floating approach to LCs).[10]

(14)



Final consonant accounts are challenged by instances of LCs that appear to be separated from the preceding word by an intonational boundary or some lexical material. Liaison has regularly been shown to occur across pauses and intonational boundaries (Ågren 1973; Morin and Kaye 1982; Lucci 1983; Green and Hintze 1988; Tranel 1990; Post 2000; Morin 2003; Miller and Fagyal 2005). This phenomenon can be observed in standard parenthetical structures ((15a), from Morin 2003) and in colloquial right dislocation constructions (15b) (Tranel 1990; Côté 2005; Plénat 2008). Crucially, stable final consonants do not cross the intonational boundary in the same constructions.

(15) a. *un robuste, mais petit, enfant* [œ̃ʁɔbyst mɛpœti t̲ɑ̃fɑ̃]
'a robust, but small-MASC, child'

b. *j'en ai un, ami* [ʒɑ̃neœ̃ n̲ami]
'I of-it have one-MASC, friend'

[10] Post's (2000) account combines features of both floating and suppletive approaches. She adopts floating segments in underlying representations, but derives liaison allomorphs in the lexicon. At the point of lexical insertion, the distinction between stable and latent consonants is lost.

A verbal marker [t] has also been identified, although its morphemic status is not as firmly established as that of plural [z]; Morin and Kaye (1982) describe verbal [t] as involving a sort of "morphologized epenthesis." This [t] seems to be an extension across verbal forms of the liaison [t] that is standardly triggered by a subset of persons, verb classes, and tenses (19).

(19) a. *le corps sera exposé* [. . . sœʁatɛkspoze]
'the body will be exhibited'
b. *image qui m'est venue à l'esprit* [. . . vœnytalɛspʁi]
'image that me is come to the mind' = 'image that came to my mind'

Finally, the morphemic account has been applied to prenominal singular adjectives: the liaison [t] in *grand arbre* 'big tree' [grɑ̃taʁbʁ] has been treated as a suffix to *grand*, comparable to a case marker (Morin 1992), and as a prefix to Word2 (Klausenburger 2001; Morin 2003).

3.1.4 LCs as initial consonants of Word2

The affiliation of LCs to Word2 is taken a step further by Ternes (1977), who conceives liaison as a case of initial consonant mutation: LCs belong to the following word and vowel-initial words subject to liaison have several variants, for example [ami], [tami], [nami], and [zami] for *ami* 'friend'. The correct variant is selected by the preceding context. The initial approach to LCs has been applied more restrictively to LCs before enclitics (5c): *y, en* in imperative constructions and *il(s), elle(s), on* in subject inversion are lexicalized as [zi zɑ̃ til tɛl tɔ̃] (Morin 1979, 1986; Côté 2005, 2008). This analysis is supported, among other evidence, by non-standard imperative forms as in (20), with an obligatory [z] after another enclitic.

(20) *donne-moi-*[z]*-en* 'give me of-it' [dɔnmwazɑ̃]

3.1.5 LCs as part of larger constructions

Finally, LCs have been viewed as part of constructions larger than the word that are (partially) lexicalized (Bybee 2001a, 2001b, 2005; Dugua 2006; Chevrot *et al.* 2007, 2009). Constructions may be more or less specific: general constructions like [DET NOUN] coexist with more specific ones like [*un* 'a/one' NOUN]$_{MASC}$ or [*un* [-n-] (vowel) NOUN]$_{MASC}$. This last construction establishes a liaison [n] between *un* and a masculine vowel-initial noun. The assumption is that more specific constructions apply first but are subject to a generalizing pressure from the less specific (and more frequent) constructions. Liaison is involved in highly specific constructions, whose strength and maintenance are argued to depend on their frequency of usage.

LCs occur at the boundary of Word1 and Word2, which prevents an a priori determination of their lexical affiliation with either word. We should not be surprised, then, by the available range of conceptions of LCs, which covers the full distance between Word1 and Word2. A schematic representation of the distance between LCs and their adjacent words appears in (21). LCs analyzed as final and initial consonants are most closely associated with Word1 and Word2, respectively. Suffixes and prefixes are partially independent from their host. LCs in epenthetic or constructionist approaches are equally (in)dependent from both words and lie at an equal distance from them.

and perceptual weakness of LCs also challenges morphemic accounts, in that consonants with a morphemic status have been shown to be longer than corresponding non-morphemic consonants, for example the inflectional [s] in English *wrecks vs.* the non-morphemic [s] in *Rex* (Walsh and Parker 1983; Losiewicz 1992). If LCs were morphemic (prefixes or suffixes), their shorter duration would be unexpected. One issue in interpreting these acoustic/perceptual results, however, is that they combine different liaison contexts, without distinguishing LCs that are associated with specific morphological categories (in particular the plural [z]) from those that appear to be strictly lexical (following invariable words or singular adjectives). It is in fact the latter type, in particular prenominal singular adjectives, that has been used predominantly, sometimes exclusively, in most studies. A morphemic analysis of other types of LCs therefore remains to be investigated.

A number of experiments have also tested the productivity of liaison with words that do not normally appear in liaison contexts (Morin 1992; El Fenne 1994; Sampson 2001). Adjectives that are either novel or usually appear postnominally (as most adjectives do in French) are tested in adjective + noun sequences. Liaison applies variably in this situation, a result that has been used to argue both for and against a final-consonant analysis of LCs (Paradis and El Fenne 1995; Côté 2005). But since the variability of liaison in noun + adjective sequences is now acknowledged, it is difficult to draw firm conclusions.

The acquisition of liaison by children is a growing area of investigation. Categorical liaison is mastered by the age of five; variable liaison develops until much later, adults varying considerably in their control and use of it. The acquisition process is characterized by numerous forms that depart from the target ones. Five main categories of mistakes involving liaison are identified (see Dugua 2006: 114–137): the substitution of the expected LC by an erroneous one (23a), the omission of a LC in categorical liaison contexts (23b), the insertion of a LC in non-liaison contexts (23c), the substitution of a stable initial consonant by a LC (when the initial consonant corresponds to a possible LC) (23d), and the omission of a stable initial consonant (23e).

(23)				*child*	*adult*
	a.	*un âne*	'a/one-MASC donkey'	[œ̃t̲an]	[œ̃n̲an]
	b.	*des étoiles*	'DET INDEF stars'	[deetwal]	[dez̲etwal]
	c.	*papa aigle*	'Daddy eagle'	[papan̲ɛgl]	[papaɛgl]
	d.	*un zèbre*	'a/one-MASC zebra'	[œ̃n̲ɛbʁ]	[œ̃zɛbʁ]
	e.	*un zèbre*	'a/one-MASC zebra'	[œ̃ɛbʁ]	[œ̃zɛbʁ]

These mistakes are sensitive to different contextual factors, and their relative frequency evolves over time. On the basis of both experimental results and the analysis of spontaneous productions, different developmental scenarios have been elaborated. It is generally agreed that in the early stages of the acquisition of liaison, around the age of two to three, LCs appear to be encoded as initial consonants of Word2, in conformity with a preference for CV syllabifications (e.g. *nami, tami, zami* for *ami* 'friend') (see Chevrot *et al.* 2005, 2009; Chevrot *et al.* 2007 for arguments and references). The acquisition of liaison then consists, for the child, in identifying the contexts in which each of these variants should be used.

Subsequent stages are more controversial. Morel (1994), Wauquier-Gravelines (2005, 2009) and Wauquier-Gravelines and Braud (2005) consider that LCs belong syntactic relationship holding between *vous* and *écouter* in the second example. (26c) displays a noun–adjective structure that excludes liaison with singular nouns, while (26d) corresponds to an adjective-noun sequence typically involving liaison. The pair (26e, 26f), from Prunet (1992: 42), differ in whether the PP introduced by *à* is a complement of *dire* or *partait*; only in the latter case is liaison between *partait* and *à* possible.

(26) a. *allez-vous écouter* [alevuekute] (no liaison)
'go you-PL listen' = 'are you going to listen'
b. *allez vous écouter* [alevuz̲ekute] (categorical liaison)
'go you-PL listen' = 'go listen to yourselves'
c. *savant anglais* [savɑ̃ɑ̃glɛ] (no liaison)
'savant English-MASC'
d. *savant Anglais* [savɑ̃(t̲)ɑ̃glɛ] (variable liaison)
'knowledgeable-MASC Englishman'
e. *il fallait dire qu'on partait à Marie* [. . . diʁkɔ̃paʁtɛamaʁi]
'it needed say that one/we left to Marie' (no liaison)
= 'somebody had to tell Marie that we were leaving'
f. *il fallait dire qu'on partait à Paris* [. . . diʁkɔ̃paʁtɛ(t̲)apaʁi]
'it needed say that one/we left to Paris' (marginal liaison)
= 'somebody had to say that we were going to Paris'

Traditional descriptions of liaison essentially present liaison contexts as sequences of standard parts of speech such as noun and preposition. Such lists are viewed as carrying little explanatory power, and more theoretically oriented analyses have tried to subsume these contexts under unified configurations, prosodic or syntactic. At a more conceptual level, syntactic approaches also raise the question of whether a phonological process (insofar as liaison is such a process) may refer directly to syntactic structure. Alternatively, frequency of co-occurrence has been proposed as an explanatory factor of the degree of cohesion between Word1 and Word2. Structural and frequency-based approaches are briefly reviewed below.

4.1 Structural approaches

Syntactic accounts of liaison contexts follow in part the evolution of syntactic theory and approaches to the syntax–phonology interface. They establish a syntactic configuration or domain in which liaison applies. Proposals rely, directly or indirectly, on a small set of morphosyntactic distinctions: (i) a distinction between lexical and functional categories; (ii) a distinction between specifier–head and head–complement relations; (iii) a distinction between inflected heads, marked for person or number, and non-inflected heads. A representative schema is given in (27) and exemplified with the NP *mes amis américains* 'my friends American'. Liaison is categorical between a specifier (the determiner *mes*) and a head (the noun *amis*), and variable between an inflected head (the plural *amis*) and a following complement (the adjective *américains*). Note that a singular noun would not trigger liaison, hence the inflectional restriction on liaison in the head–complement configuration.

According to de Jong (1990), the categorical–variable distinction in liaison contexts corresponds to different prosodic levels. Alternatively, one may exploit the variability in prosodic parsing: liaison is systematic within PhPs but its optionality in certain contexts results from the possible restructuring of adjacent PhPs. If Word1 and Word2 belong to different PhPs, these may merge if Word1 is inflected, Word2 is in the complement of Word1, and the second PhP is non-branching; see Post (2000) for a discussion and assessment of this type of approach.

More recently, Davis (2000) has proposed to associate the domain of liaison with the concept of Extended Projection (EP), a syntactic unit defined as a combination of a functional and a lexical category. This approach differs from the preceding ones in that all word categories project a phrasal level, rendering irrelevant the distinction between heads and specifiers. Liaison applies obligatorily within EPs, and its variability in several contexts is attributed to independent factors: variable underlying representations, with and without a final latent consonant, and interference from prosodic structure and the number of syllables. Unlike all previous approaches, the EP domain predicts liaison to be impossible after lexical categories (except prenominal adjectives). Occurrences of liaison after nouns and verbs, which remain marginal in informal speech, follow from the possibility of using the left edge of a PhP (a prosodic category) rather than that of an EP as a liaison boundary.[13]

Although they invoke similar concepts, these analyses make different predictions. The sentence in (28), which contains nine potential contexts of liaison, is used to compare four specific proposals: X″ (Selkirk 1974), PhP (de Jong 1990), c-command (Prunet 1992), and EP (Davis 2000). For each context, Table 112.1 indicates the prediction of each analysis and, for comparison, its relative frequency in real speech, based on the generalizations in (5), (6), and (7) above, complemented with Morin and Kaye's (1982) observations. The contexts are listed in order of frequency. We may consider that all contexts involve a relationship between a head and a syntactically dependent element, except 2 and 8.

(28) le petit enfant a vait appris aux hommes attentifs à ses prouesses à
1 2 3 4 5 6 7 8
voler sans ailes
9
'the small-MASC child had taught to-the men attentive to his feats to fly without wings'

In only the first and last contexts do all the analyses make exactly the right prediction. Otherwise all underpredict and overpredict liaison in at least some contexts. The adjective + noun context (*petit enfant*) is wrongly predicted to be obligatory, but this is consistent with the earlier classification of this context as categorical. A past participle followed by its prepositional complement (*appris aux hommes*) is predicted not to trigger liaison, since Word1 (*appris*) is not inflected. Liaison is indeed rare, but not unattested, as shown by Morin and Kaye (1982). On the other hand, a plural adjective followed by its complement (*attentifs à ses*

[13] This makes the questionable prediction that the use of the PhP increases with formality. Since formal liaison is closer to the written form, it is not expected to be more strongly driven by prosodic structure than informal liaison.

Table 112.1 Status of liaison in the nine contexts in (28), as observed in real speech and predicted by four analyses

	context	*observed*	Selkirk 1974	de Jong 1990	Prunet 1992	Davis 2000[14]
5	Det+N (4a)	categorical	obligatory	obligatory	obligatory	obligatory
1	Adj+N (5a)	frequent	obligatory	obligatory	obligatory	obligatory
9	Prep+NP (5b)	variable	obligatory	frequent	formal style	variable
3	Aux+Past part (5d)	variable	obligatory	frequent	obligatory	obligatory
6	N_{PL}+Adj (5e)	infrequent	variable	rare	formal style	no
4	Past part+PP (5d)	marginal	no	no	no	no
7	Adj_{PL}+PP	no	variable	rare	formal style	no
2	Subj+V (6a)	no	no	no	no	no
8	N+non-comp	no	no	no	no	no

prouesses, context 7) should trigger variable liaison, as nouns and verbs do in similar configurations, but the absence of liaison after postnominal adjectives has often been noted (e.g. Morin and Kaye 1982; de Jong 1994; Davis 2000).

Syntactic approaches, direct or indirect, contrast with strictly prosodic ones, where prosodic domains are defined by intonational and durational patterns rather than syntactic structure. It is traditionally claimed that liaison is only realized inside a rhythmic or accentual group, therefore never between a stressed vowel and a following word (e.g. Béchade 1992; Léon and Léon 1997: 40; cf. Morrison 1968). A more recent and theoretically driven attempt to establish the Accentual Phrase (AP) as the domain of liaison is presented in Scarborough and Jun (2003).

Structural accounts of the relationship between Word1 and Word2 raise a number of general objections, and have failed to receive solid support from corpus-based or experimental testing. Fougeron *et al.* (2001a) found a strong positive correlation between liaison and syntactic distance (measured as the number of nodes separating Word1 and the common ancestor of Word1 and Word2); this confirms the basic intuition that Word1 and Word2 tend to be syntactically close in order for liaison to apply. But studies of read speech corpora indicate that c-command, the PhP, and the AP are not good predictors of liaison, either because liaison is often realized outside of the proposed domains (i.e. without c-command and across PhP boundaries, but only rarely across APs) or because liaison is not realized in a large proportion of potential liaison contexts within these domains (liaison is not produced in 50 percent of liaison contexts with c-command, against 49 percent within APs and 24 percent within PhPs) (Fougeron *et al.* 2001a; Fougeron and Delais 2004; see also Post 2000).

Generally, syntactic approaches predict a homogeneous behavior of liaison across categories, admitting a distinction between functional and lexical ones. This is not the case. On the one hand, non-prenominal adjectives almost never trigger liaison, unlike other lexical categories, nouns and verbs. Adverbs, which tend to

[14] These predictions are based only on the EP part of Davis's analysis and do not take account of the possibly interfering factors mentioned above.

(Culetto 2008). Fougeron *et al.* (2001a) and Fougeron *et al.* (2001b) observe that the frequency of the Word1-Word2 sequence is a better predictor than transitional probability, but a general statistical model is not yet in sight.

5 Conclusions

French liaison offers a microcosm of the challenges of linguistic analysis. Although typically classified as a phonological process, it lies at multiple interfaces, involving internal and external dimensions: morphology, syntax, and the lexicon, dialectal and sociolectal variation, registers, and spelling. This complexity has motivated the inclusion of liaison in a number of important conceptual debates, such as those surrounding abstractness, representations, and usage.

The last 50 years of study of liaison have seen a shifting balance between formal elegance and empirical adequacy, data idealization, and analytical scattering. One central goal of this chapter has been to convey a sense of the empirical and conceptual richness involved. With respect to the lexical status of LCs, the traditional relationship between the LC and the preceding word has been loosened, by establishing a morphological or semantic motivation to liaison that is independent of Word1, introducing prosodic or lexical material between Word1 and the LC, or dissociating liaison from Word1's morphological paradigm. Studies of liaison contexts have revealed the deficiencies of global structural accounts, syntactic or prosodic. It is hoped that the recent surge of experimental and corpus-based research will lead to new analytical developments on both issues.

ACKNOWLEDGMENTS

I would like to thank Jacques Durand for his support and Chantal Lyche, Sophie Wauquier, two anonymous reviewers, and the *Companion* editors for their contribution to this chapter.

REFERENCES

Ågren, John. 1973. *Remarques sur quelques liaisons facultatives dans le français de conversation radiophonique*. Uppsala: Acta Universitatis Upsaliensis.

Ameringen, Arie van & Henriette J. Cedergren. 1981. Observations sur la liaison en français de Montréal. In Sankoff & Cedergren (1981), 141–149.

Ashby, William J., Marianne Mithun, Giorgio Perissinotto & Eduardo Raposo (eds.) 1993. *Linguistic perspectives on the Romance languages*. Amsterdam: John Benjamins.

Béchade, Hervé-D. 1992. *Phonétique et morphologie du français moderne et contemporain*. Paris: Presses Universitaires de France.

Bonami, Olivier, Gilles Boyé & Jesse Tseng. 2005. Sur la grammaire des consonnes latentes. *Langages* 158. 89–100.

Booij, Geert. 1983. French C/Ø alternations, extrasyllabicity and Lexical Phonology. *The Linguistic Review* 3. 181–207.

Bybee, Joan. 2001a. Frequency effects on French liaison. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*, 337–359. Amsterdam & Philadelphia: John Benjamins.

Bybee, Joan. 2001b. *Phonology and language use*. Cambridge: Cambridge University Press.
Bybee, Joan. 2005. La liaison: Effets de fréquence et constructions. *Langages* 158. 24–37.
Chevrot, Jean-Pierre, Damien Chabanal & Céline Dugua. 2007. Pour un modèle de l'acquisition des liaisons basé sur l'usage: Trois études de cas. *Journal of French Language Studies* 17. 103–128.
Chevrot, Jean-Pierre, Céline Dugua & Michel Fayol. 2005. Liaison et formation des mots en français: Un scénario développemental. *Langages* 158. 38–52.
Chevrot, Jean-Pierre, Céline Dugua & Michel Fayol. 2009. Liaison acquisition, word segmentation and construction in French: A usage-based account. *Journal of Child Language* 36. 557–596.
Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.
Coquillon, Annelise & Corine Astésano. 2008. A multiparametric phonetic investigation of liaison in standard and meridional French. Paper presented at the Conference Structures des français en contact, Tulane University. Available (July 2010) at www.projet-pfc.net/index.php?option=com_docman&task=cat_view&gid=943&Itemid=179&limitstart=15.
Côté, Marie-Hélène. 2005. Le statut lexical des consonnes de liaison. *Langages* 158. 66–78.
Côté, Marie-Hélène. 2008. Empty elements in schwa, liaison and h-*aspiré*: The French holy trinity revisited. In Jutta M. Hartmann, Veronika Hegedüs & Henk van Riemsdijk (eds.) *Sounds of silence: Empty elements in syntax and phonology*, 61–103. Amsterdam: Elsevier.
Côté, Marie-Hélène. 2010. Le statut des consonnes de liaison: L'apport de données du français laurentien. In Franck Neveu, Valelia Muni Toke, Thomas Klingler, Jacques Durand, Lorenz Mondada & Sophie Prévost (eds.) *2ème congrès mondial de linguistique française*, 1279–1288. Paris: Institut de Linguistique Française.
Culetto, Thomas. 2008. Prediction of liaison in French by measures of information theory. In Miltiadis Kokkonidis (ed.) *Proceedings of LingO 2007*, 59–66. University of Oxford: Faculty of Linguistics, Philology, and Phonetics. Available (July 2010) at www.ling-phil.ox.ac.uk/events/lingo/proceedings.htm.
Davis, Julia L. 2000. French liaison: A case study of the syntax–phonology interface. Ph.D. dissertation, Indiana University.
Dejean-de-la-Bâtie, Bernadette. 1993. Word boundary ambiguity in spoken French. Ph.D. dissertation, Monash University, Melbourne.
Delattre, Pierre. 1940. Le mot est-il une entité phonétique en français? *Le français moderne* 8. 47–56.
Delattre, Pierre. 1947. La liaison en français: Tendances et classifications. *French Review* 21. 148–157.
Delattre, Pierre. 1955. Les facteurs de la liaison facultative en français. *French Review* 29. 42–49.
Desrochers, Richard. 1994. Les liaisons dangereuses: Le statut équivoque des erreurs de liaison. *Linguisticae Investigationes* 18. 243–284.
Dugua, Céline. 2006. Liaison, segmentation lexicale et schémas syntaxiques entre 2 et 6 ans: Un modèle développemental basé sur l'usage. Ph.D. dissertation: Université Stendhal, Grenoble.
Durand, Jacques & Chantal Lyche. 2008. French liaison in the light of corpus data. *Journal of French Language Studies* 18. 33–66.
Durand, Jacques, Benoit Habert & Bernard Laks (eds.) 2008. *Congrès mondial de linguistique française*. Paris: Institut de Linguistique Française.
El Fenne, Fatimazohra. 1994. La flexion verbale en français: Contraintes et stratégies de réparation dans le traitement des consonnes latentes. Ph.D. dissertation, Université Laval.
Encrevé, Pierre. 1988. *La liaison avec et sans enchaînement: Phonologie tridimensionnelle et usages du français*. Paris: Éditions du Seuil.
Fouché, Pierre. 1959. *Traité de prononciation française*. Paris: Klincksieck.

Léon, Pierre & Monique Léon. 1997. *La prononciation du français*. Paris: Nathan.
Losiewicz, Beth L. 1992. The effect of frequency on linguistic morphology. Ph.D. dissertation, University of Texas, Austin.
Lucci, Vincent. 1983. *Étude phonétique du français contemporain à travers la variation situationnelle*. Grenoble: Publications de l'Université des Langues et Lettres.
Mallet, Géraldine. 2008. La liaison en français: Descriptions et analyses dans le corpus PFC. Ph.D. dissertation, Université Paris Ouest, Nanterre La Défense.
Mazzola, Michael L. 1993. French rhythm and French segments. In Ashby *et al.* (1993), 113–126.
Miller, Jessica S. & Zsuzsanna Fagyal. 2005. Phonetic cues to common and special cases of liaison: Looking for a prosodic domain. In Randall Gess & E. J. Rubin (eds.) *Theoretical and experimental approaches to Romance linguistics*, 179–196. Amsterdam & Philadelphia: John Benjamins.
Milner, Jean-Claude. 1967. French truncation rule. *MIT Quarterly Progress Report* 86. 273–283.
Moisset, Christine. 2000. Variable liaison in Parisian French. Ph.D. dissertation, University of Pennsylvania.
Morel, Éliane. 1994. Le traitement de la liaison chez l'enfant: Études expérimentales. *Travaux neuchatelois de linguistique (TRANEL)* 21. 85–95.
Morin, Yves-Charles. 1979. La morphophonologie des pronoms clitiques en français populaire. *Cahier de linguistique* 9. 1–36.
Morin, Yves-Charles. 1981. Où sont passés les s finals de l'ancien français? In Sankoff & Cedergren (1981), 35–47.
Morin, Yves-Charles. 1982. De quelques [l] non étymologiques dans le français du Québec: Notes sur les clitiques et la liaison. *Revue québécoise de linguistique* 11. 9–47.
Morin, Yves-Charles. 1986. On the morphologization of word-final consonant deletion in French. In Henning Andersen (ed.) *Sandhi phenomena in the languages of Europe*, 167–210. Berlin: Mouton de Gruyter.
Morin, Yves-Charles. 1987. French data and phonological theory. *Linguistics* 25. 815–843.
Morin, Yves-Charles. 1990. La prononciation de [t] après *quand*. *Linguisticae Investigationes* 14. 175–189.
Morin, Yves-Charles. 1992. Un cas méconnu de la déclinaison de l'adjectif français: Les formes de liaison de l'adjectif antéposé. In André Clas (ed.) *Le mot, les mots, les bons mots*, 233–250. Montreal: Presses de l'Université de Montréal.
Morin, Yves-Charles. 2003. Remarks on prenominal liaison consonants in French. In Stefan Ploch (ed.) *Living on the edge: 28 papers in honour of Jonathan Kaye*, 385–400. Berlin & New York: Mouton de Gruyter.
Morin, Yves-Charles. 2005. La liaison relève-t-elle d'une tendance à éviter les hiatus? Réflexions sur son évolution historique. *Langages* 158. 8–23.
Morin, Yves-Charles & Jonathan Kaye. 1982. The syntactic bases for French liaison. *Journal of Linguistics* 18. 291–330.
Morrison, Allen S. 1968. A critical bibliography of studies of liaison in French speech since 1800. Ph.D. dissertation, Columbia University.
Nespor, Marina & Irene Vogel. 1982. Prosodic domains of external sandhi rules. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part 1, 225–255. Dordrecht: Foris.
Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
Nguyen, Noël, Sophie Wauquier-Gravelines, Leonardo Lancia & Betty Tuller. 2007. Detection of liaison consonants in speech processing in French: Experimental data and theoretical implications. In Pilar Prieto, Joan Mascaró & Maria-Josep Solé (eds.) *Segmental and prosodic issues in Romance phonology*, 3–23. Amsterdam & Philadelphia: John Benjamins.
Pagliano, Claudine & Bernard Laks. 2008. Problématiques de la liaison dans l'analyse d'un corpus de français oral actuel. In Chantal Parpette & Marie-Anne Mochet (eds.) *L'oral en représentation(s): Décrire, enseigner, évaluer*, 69–91. Louvain: Éditions EME.

8 Sonorants

Bert Botma

1 Introduction

The class of sonorants includes vowels, semi-vowels, liquids, and nasals, with the sonorant consonants sometimes being referred to as "resonants" (e.g. Laver 1994). There is abundant cross-linguistic evidence for natural class behavior of these segments. For example, in languages where consonants can be syllabic, the class of syllabic segments is almost always restricted to sonorants (Imdlawn Tashlhiyt Berber is an exception; see Dell and Elmedlaoui 2002). Further, there are quantity-sensitive languages in which the set of weight-bearing segments is restricted to sonorants, or to a subset thereof (see Zec 2007 and CHAPTER 57: QUANTITY-SENSITIVITY for examples). Similarly, in languages where consonants can support tone contrasts, the consonants involved are almost always sonorant (e.g. Yip 2002). In addition, in languages which permit complex onsets, the second element is typically restricted to a sonorant consonant (disregarding sC clusters). Taken together, these observations suggest a distinctive feature, e.g. [sonorant], which distinguishes sonorants from obstruents (i.e. plosives and fricatives). These examples further suggest that the most straightforward evidence for the class behavior of sonorants comes from their patterning with respect to suprasegmental aspects such as syllable structure, moraic structure, and tone – an observation to which we return in §3.

While the class of sonorants is well established, there are a number of issues relating to sonorants which are a matter of debate. This chapter focuses on three such issues:

(i) How should sonorancy be represented in a theory of segmental structure? (§2)
(ii) How should less straightforward cases of sonorant class behavior be accounted for, in particular those cases in which voiced stops and voiced fricatives display sonorant-like behavior? (§3)
(iii) How should sonorancy be defined phonetically? (§4)

On closer inspection, each of these issues pertains to the relation between sonorancy and voicing. Any approach to sonorancy must do justice to the observation that voicing in sonorants is never contrastive (see §2.3), but may nevertheless be phonologically active. §2 considers a range of proposals that have been advanced to account for this observation, and also provides a more general discussion of laryngeal modifications in sonorants. The relation between sonorancy and voicing is also relevant in the class behavior of sonorants and voiced stops and fricatives, which is considered in §3. §4 reviews a number of possible phonetic correlates of sonorancy. It will be suggested that sonorants are best characterized acoustically, in terms of a clearly marked formant structure – a position that is perhaps most closely identified with the work of Ladefoged (e.g. 1971, 1982, 1997). It will also be suggested, somewhat more tentatively, that the class behavior of sonorants and voiced obstruents is phonetically natural to the extent that both are voiced, with the latter involving "active voice facilitation," although there is some reason to believe that aerodynamic correlates may also be relevant.

In what follows, my concern will be with the class of sonorants as a whole. Thus I will have little to say about the specific properties of sub-types of sonorants, some of which are discussed in other contributions (e.g. CHAPTER 15: GLIDES; CHAPTER 30: THE REPRESENTATION OF RHOTICS; CHAPTER 31: LATERAL CONSONANTS), or about the relation between sonorancy and syllable structure (for this, see e.g. CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE; CHAPTER 49: SONORITY; CHAPTER 53: SYLLABLE CONTACT).

2 How should sonorancy be represented in a theory of segmental structure?

2.1 *Representations of sonorancy*

Different proposals have been made regarding the formalization of sonorancy in phonological theory. The mainstream view, as espoused in Chomsky and Halle (1968; *SPE*) and subsequent work in Feature Geometry (e.g. Clements 1985; Sagey 1986; McCarthy 1988; Clements and Hume 1995; Halle *et al.* 2000), recognizes a major class feature [sonorant]. An alternative proposal within Feature Geometry is the "SV hypothesis," in which [sonorant] is replaced by a "Spontaneous Voicing" or "Sonorant Voice" node (e.g. Rice and Avery 1989, 1991; Piggott 1992, 1993; Rice 1993; Avery 1996). A quite different approach is taken in Dependency Phonology and related frameworks such as Radical CV Phonology, where sonorancy is viewed as a particular manifestation of a vocalic "component" or "element," e.g. |V| (Anderson and Ewen 1987; van der Hulst 1995) or |L| (Botma 2004; Botma and Smith 2006, 2007). However, while these frameworks assume different "atoms" of representation, they share important insights with both *SPE* and Feature Geometry, as we will see below. Finally, some approaches, e.g. most versions of Element Theory (see Harris and Lindsey 1995; Harris 1996), do not recognize a feature or "element" representing sonorancy. This section considers each of these approaches in some detail.

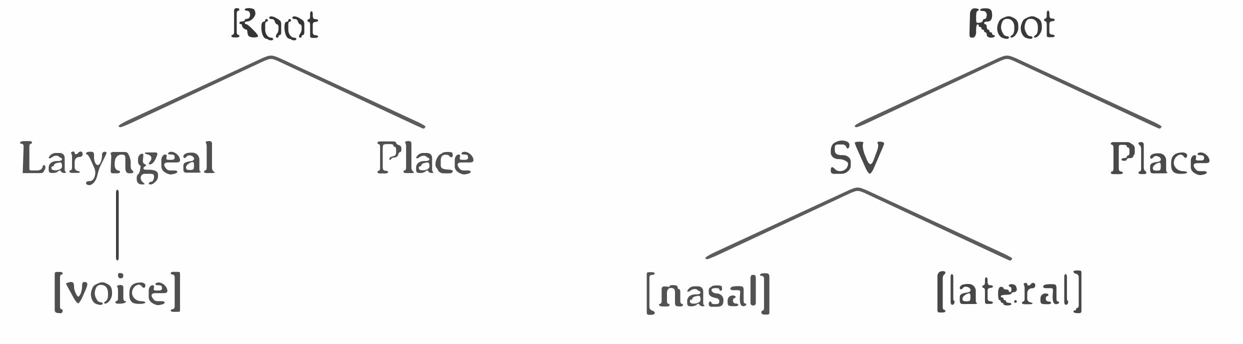
In *SPE*, [sonorant] is assumed to be a "major class feature," together with [consonantal] and [continuant]. What sets major class features apart from other features is that the former are not bound to any particular articulator. Thus, as Kenstowicz (1994: 36) puts it, [sonorant] specifies "phonologically relevant degrees of constriction imposed by essentially any articulator." However, since segments in *SPE* consist of unordered feature bundles, major class features do not have any independent theoretical status.

of the output already inheres in the lenition target. This option is not considered in Harris (1996), though it is perhaps compatible with Harris's more recent approach to sonority (see Harris 2006).

Harris goes on to make a compelling case for rejecting [consonantal], noting in passing that there are few processes in which major class features function as the target or trigger of phonological processes. However, it is questionable whether this last point can be maintained for the feature [sonorant], not just in light of the examples of sonorant class behavior mentioned in §1, but also in view of the kind of arguments advanced by proponents of the SV hypothesis.

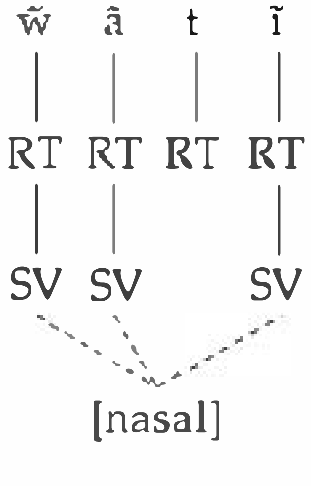
The SV hypothesis can be regarded as an attempt to accord [sonorant] the same status as other, non-major class features. The key insight of this approach is that there are two types of phonological voicing: one realized through a laryngeal feature ([voice], dominated by the Laryngeal node) and a property of obstruents (2a), and one through SV ("Spontaneous Voice" or "Sonorant Voice") and a property of sonorants (2b).

(2) a. *Obstruent voicing* b. *Sonorant voicing*



The SV node organizes sonorant features such as [nasal] and [lateral], and acts as both a trigger and a target for autosegmental processes such as spreading and delinking. For example, Piggott (1992) accounts for the nasal harmony pattern of Southern Barasano (discussed in more detail in §3) on the assumption that all targets are specified for SV, which provides a landing site for a harmonic feature [nasal]. Thus a form like [w̃ãtĩ] 'devil' involves association of [nasal] to all SV-specified segments in the word:

(3)



Spreading of the SV node itself is assumed to take place in sonorant-induced voicing phenomena, such as post-nasal voicing assimilation (see §2.2).

The function of the SV node is similar to class nodes such as Laryngeal and Place, in that it designates functional feature groupings (see especially Piggott 1993 for an elaboration of this idea). In addition, some phonologists have argued that the unmarked dependent of the SV node is the feature [nasal], which captures the observation that nasals are the unmarked type of sonorant consonants. According to this view, nasals are underlyingly specified for SV only, with their nasal aspect supplied by a default rule which fills in [nasal] (e.g. Rice and Avery 1991; Rice 1993). However, this assumption seems difficult to combine with an account of post-nasal voicing assimilation in terms of SV spreading (see §2.2).

A difference between class nodes and the SV node is that the latter is construed as having phonetic content. This assumption is not unproblematic. While a distinction between sonorant voicing and obstruent voicing seems well motivated on phonological grounds, the question whether there is phonetic support for this distinction is rather more contentious (see §4.1 for further discussion). To this extent at least, a more promising approach is taken in Dependency Phonology and related frameworks, where sonorancy and voicing are viewed as manifestations of a single component, whose interpretation depends on its position in the phonological structure. For example, van der Hulst (1995) represents sonorancy in terms of a |V| in the "head" component of a segment (4a) and voicing in terms of a |V| in the "dependent" component (4b). (The possibility of expressing the relative prominence of voicing can also be used to represent the difference between sonorants and sonorant-like obstruents, as we will see in §3.)

(4) a. *Sonorancy* b. *Voicing*

Van der Hulst further assumes that in (4b) the head may be either an obstruent, i.e. |C|, or a sonorant, i.e. |V|. The latter assumption attempts to do justice to the observation that there are languages in which sonorants trigger voicing. Such processes cannot involve the head |V|, given van der Hulst's (1995: 96) assumption that "dependent properties can spread independently, while heads can only spread together with their dependents." This makes head components similar to major class features, while the possibility of "active" voicing in sonorants is reminiscent of the SV hypothesis.

2.2 *The relationship between sonorancy and voicing*

Phonetically, the natural state of sonorants is to be voiced. Phonologically, this is reflected by the observation that voicing in sonorants is unmarked. Inspection of the UPSID database (Maddieson 1984) reveals that most languages have voiced sonorants only (96.6 percent) and that the presence of non-voiced sonorants in a language implies the presence of voiced ones. Furthermore, voicing in sonorants is often redundant. This has been interpreted to mean that [voice] is underspecified for segments specified as [sonorant] and supplied by default in the post-lexical phonology (e.g. Kiparsky 1985; Archangeli 1988; Lombardi 1994; Steriade 1995).

One phenomenon which suggests that sonorants are underlyingly unmarked for voice is Japanese Rendaku (e.g. Itô and Mester 1986; Itô *et al.* 1995; Nasukawa 2005). This concerns a voicing process which targets an initial voiceless consonant of the second member of a compound, converting it into its voiced counterpart, as in /onna + kokoro/ → [onna gokoro] 'woman's heart'. Rendaku interacts with a constraint known as Lyman's Law, which bans multiple occurrences of voiced

3 Fuzzy class behavior: Sonorant obstruents

While the class of sonorant consonants traditionally includes liquids and nasals, it has been observed that voiced fricatives and voiced stops sometimes also pattern as sonorants, at least with respect to certain phenomena. Following Rice (1993), such sounds will be referred to as "sonorant obstruents." This section considers some examples of sonorant obstruents that have been reported in the literature and discusses the kinds of arguments that have been advanced for treating these sounds as sonorants. The recognition of a class of sonorant obstruents also has repercussions for the phonetic underpinnings of sonorancy (see §4).

Consider first the following data from Turkish. In this language, syllable-final voiced plosives undergo devoicing (7a) while voiced fricatives retain their voicing in this position (7b), like sonorants (7c) (cf. Rice 1993: 332–333):

(7)	a.	*sara*[p]	'wine-NOM SG'	*vs.*	*sara*[b]*i*	'wine-ACC SG'
	b.	*a*[z]	'few'			
		e[v]	'house'			
	c.	*gü*[n]	'day'			
		güze[l]	'pretty'			

Two possible analyses of the Turkish pattern suggest themselves. The first limits the structural description of devoicing to stops, e.g. by specifying the targets of devoicing as [−sonorant, −continuant]. As Rice notes, the problem with this analysis is that it does little more than stipulate that devoicing is limited to stops. The alternative is to specify the voiced fricatives for the same feature as sonorants, e.g. [sonorant] or, in Rice's analysis, SV. This account is more parsimonious, since devoicing can now be restricted to whatever feature it is that obstruents have in common. In Rice's account this is the Laryngeal node, which dominates the feature [voice]. In this analysis, [voice] is therefore a property of voiced obstruents, while the voicing of sonorants, including "sonorant fricatives," is supplied by SV. Rice offers a similar account of the voiced fricatives of Athapaskan languages such as Navajo and Chipewyan.

There is reason to suspect that "sonorant fricatives" may in fact be rather widespread. For example, voiced uvular /ʁ/ and pharyngeal /ʕ/ often pattern as sonorants (e.g. CHAPTER 25: PHARYNGEALS). Further, /v/ has been shown to display sonorant-like behavior in languages such as Norwegian (Kristoffersen 2000), Russian (Padgett 2002), Hungarian (Bárkányi and Kiss 2006), and Icelandic (Botma 2008). Hamann (2006) proposes the same for German /v/, providing acoustic measurements (duration, intensity, and harmonics-to-noise ratio) which suggest that the sound is phonetically a narrow approximant. More generally, Maddieson (1984: 48) observes that in UPSID, "bilabial, dental and palatal non-sibilant fricatives are found to occur without a voiceless counterpart more often than with one." To the extent that this is due to a difference in markedness between voiced and voiceless fricatives (and not to diachronic lenition, say), specifying these voiced fricatives for [voice] seems ill advised. Their relative frequency could instead support an analysis in which voiceless fricatives are specified for [spread glottis] (e.g. CHAPTER 28: THE REPRESENTATION OF FRICATIVES). However, it might also be the case that at least some of these voiced fricatives are in fact sonorants. While such a hypothesis is phonetically feasible (vocal fold vibration leads to lower airstream velocity, making it relatively difficult to produce turbulence), it must of course be demonstrated for each of the sounds in question that they pattern as sonorants phonologically.

One type of evidence that is often adduced for the sonorant status of voiced stops is the presence of oral–nasal alternations such as [b ~ m] (e.g. Piggott 1992; Rice 1993; Clements and Osu 2002; Botma 2004). Implicit in this approach is the claim that only sonorants can be nasalized. Phonetically, this claim is not unreasonable. The presence of nasal airflow is antagonistic to the buildup of oral air pressure required for obstruents. This rules out nasalized plosives (provided a different interpretation is given to nasal contours) and makes nasalized fricatives distinctly rare. Voiced nasalized fricatives have been reported in languages such as Inor, Itsekiri, and Umbundu (see Walker 2000 and references there), though it remains to be verified instrumentally whether these sounds are truly fricatives phonetically. Ohala and Ohala (1993) observe that the high rate of airflow required for friction is difficult to combine with a lowered velum, suggesting that many of the sounds described as voiced nasalized fricatives may in fact be approximants. Some phonetic studies have also reported nasal airflow in voiceless fricatives, e.g. in Coatzospan Mixtec (Gerfen 1999). Such nasalization seems to be possible only if there is an adjacent nasalized vowel, and so is presumably the result of co-articulation.

Oral–nasal alternations involving voiced stops occur in many languages with nasal harmony (see also CHAPTER 78: NASAL HARMONY). In a sub-type of nasal harmony displayed by a number of Amazonian languages, all voiced segments in the harmonic domain are nasalized, including what appear to be voiced obstruent stops phonetically. This gives rise to complementary distribution between voiced stops and nasals, as illustrated in (8) for Southern Barasano, a Tucanoan language of Colombia (cf. Piggott 1992: 46; see also Smith and Smith 1971). In this language, nasalization spreads rightwards from the leftmost nasalized vowel, skipping any intervening obstruents, as well as leftwards to an immediately preceding nasalizable consonant:

(8) a. *"Oral words"* b. *"Harmonic words"*

[^{m}ba^{ŋ}go ~ ^{m}bago]	'eater'	[kãmõkã]	'rattle'
[ta^{m}boti ~ taboti]	'grass'	[mãnõ]	'none'
[^{n}diro]	'fly'	[ẽõnõ]	'mirror'
[wesika]	'above'	[mãsã]	'people'
[wati]	'going'	[w̃ãtĩ]	'going'

Smith and Smith (1971: 82) note that voiced stops are optionally prenasalized. Their transcriptions suggest that this prenasalization is obligatory in word-initial position.

The complementary distribution of voiced stops and nasals in Southern Barasano suggests that they share a single underlying representation. This is corroborated by patterns of allomorphy of the kind in (9) (cf. Piggott 1992: 47; see also Piggott and van der Hulst 1997):

(9) a. *"Oral words"* b. *"Harmonic words"*

[ji-re]	'to say'	[ɲãɲõ-nẽ]	'to speak'
[^{ŋ}gahe-ja]	'another stream'	[mĩnõ-ɲã]	'leaf stream'
[wa-^{m}bi]	'I went'	[ĩã-mĩ]	'I saw'

Different accounts have been offered as to the underlying representation of alternating voiced stops. According to one analysis, nasal harmony in languages like Southern Barasano targets all segments specified for [voice], nasalizing sonorants and turning voiced stops into nasals (e.g. Pulleyblank 1989; Noske 1995). The problem with this analysis is that it is stipulative, since it is unclear why voicing and nasalization should have this affinity. Alternatively, it has been suggested that this type of harmony is limited to sonorants (e.g. Piggott 1992; Rice 1993; Botma 2004; Botma and Smith 2007). This analysis implies that the voiced stops in Southern Barasano are "sonorant stops," and as such permits a uniform account of the nasalization process (see e.g. Piggott and van der Hulst 1997, who offer an analysis of Southern Barasano in terms of the spreading of nasality at the level of the syllable).

Botma (2009) maintains that some Amazonian languages with a similar harmony pattern, e.g. Yuhup, are best analyzed as having underlying nasals, which denasalize in certain contexts. It is reasonable to assume that this scenario mirrors the historical situation, given the cross-linguistic frequency of nasals. (Perhaps Southern Barasano represents an innovation of this pattern.) Denasalization is also the diachronic source of alternating sonorant stops in other language families. For example, Krauss and Leer (1981) observe that in many Athapaskan languages **m* (< **w*) and **n* have developed into /ᵐb/, /ⁿd/ (e.g. in Han, Tanacross, and Southern Slave) or into /b/, /d/ (e.g. in Tahltan, Sekani, and Bearlake Slave). In all of these languages these stops still alternate with nasals before nasalized vowels. In some of them, e.g. in Bearlake Slave, the nasal realization is also still found in certain morphological contexts (Rice 1993).

Further illustration of the sonorant-like behavior of voiced stops comes from West African languages such as Cama (Ebrié), Gbe, and Ikwere. For example, Botma and Smith (2006) observe that Cama, a Kwa language of the Ivory Coast, contrasts two types of voiced stops, which Stewart (1973) describes as "fortis" and "lenis", respectively. The voiced lenis stops pattern with sonorants in that (unlike the voiced fortis stops) they alternate with nasals before nasalized vowels and do not trigger tone lowering. The voiced labial stop of Gbe, a Kwa dialect cluster of Togo and Benin, also displays oral–nasal alternations, in contrast to voiced alveolar and velar stops (Capo 1981). This /b/ derives from an earlier implosive *ɓ, which in many African languages patterns as sonorant (e.g. Kaye 1981; Clements 2000): like sonorants, implosives have unmarked voicing, are disfavored in NC clusters, often display alternations with both liquids and nasals, and fail to trigger tone lowering.

These properties are also characteristic of the voiced "non-explosive" labial stops of Ikwere, an Igboid language of Nigeria. However, Clements and Osu (2002: 337) observe that these sounds do not pattern with sonorants with respect to their "sonority-related distributional properties," nor do they behave as sonorants with regard to suprasegmental aspects such as tone and weight. This ambivalence appears to be representative of sonorant obstruents in general, and suggests that the internal structure of these sounds contains both obstruent and sonorant properties. To this end, Clements and Osu propose two distinct binary features, [obstruent] and [sonorant]. The former is an articulatory feature defined in terms of "air pressure build-up in the oral cavity," the latter an acoustic feature defined in terms of "a periodic, well-defined formant structure" (cf. Clements and Osu 2002: 338). Sonorants in this account are [–obstruent, +sonorant], while non-explosives have been criticized on the grounds that they are unable to account for phenomena that are acoustically motivated, such as the frequent interaction between labials and velars (e.g. Hyman 1973). The appropriateness of an articulatory characterization of [sonorant] has also been questioned, perhaps most notably in the work of Ladefoged (e.g. 1971, 1982, 1997), who maintains that sonorancy is best defined in acoustic terms, in line with the previously dominant approach to features of Jakobson *et al.* (1952).

Central to the *SPE* approach to sonorancy is the notion of "spontaneous voicing," which Chomsky and Halle claim is different from obstruent voicing, both phonologically and phonetically. The former claim is also the main tenet of the SV hypothesis, as we have seen in §2.1. However, it is not clear whether proponents of this hypothesis also take sonorant voicing to be phonetically distinct from obstruent voicing (cf. e.g. Rice 1993: 341), nor is it clear whether in the SV approach distinct features are required to have distinct phonetic correlates.

There are two general problems with spontaneous voicing as used in *SPE*. The first is that while in *SPE* spontaneous voicing is a key ingredient of sonorancy, Chomsky and Halle's definition does not in fact require sonorants to be voiced. This has a number of unfortunate consequences. One of these concerns the interpretation of "voiceless" sonorants, which are usually voiceless (at least for part of their duration; see §2.3) despite the fact that their vocal tract shape is conducive to voicing. Another is that in *SPE* the set of [+sonorant] segments includes the laryngeals, [h ʔ]. While these are like sonorants with respect to their vocal tract shape, their laryngeal setting is antagonistic to voicing, and for [ʔ] is in fact incompatible with it. This leads Ladefoged (1971: 109) to qualify this proposal as "counter-intuitive to say the least." An analysis of [h ʔ] as sonorants is also dubious on phonological grounds. Consider the interpretation of the lenition processes in (12), both of which are common sound changes in languages of the world (see e.g. Campbell 2004 for examples).

(12) a. f s x > h

b. p t k > ʔ

McCarthy (1988) argues that these processes are best treated as involving the loss of oral place features, which leaves behind segments specified for laryngeal structure only. According to this interpretation, [h ʔ] are therefore "defective" obstruents, like the "complete" obstruents from which they are derived. Notice further that [h ʔ] also do not display class behavior with sonorants with respect to the properties mentioned in §1. Laryngeals are never syllabic or tonal, nor do there seem to be any languages in which they are weight-bearing to the exclusion of non-sonorants.

The second problem with spontaneous voicing is that the phonetic evidence which Chomsky and Halle adduce for it is questionable, as pointed out by Ladefoged (1971), among others. Twenty-five years later, in joint work with Maddieson, Ladefoged sees no reason to change this view:

> The physiological position for modal voice can be regarded as one in which the arytenoid cartilages are in a neutral position for speech, neither pulled apart nor pushed together (Stevens 1988). The vocal folds would be very slightly apart, if there were no air flow. We assume that the same position as occurs in ordinary voiced

> vowels and in voiced continuant consonants such as nasals is normally maintained in stops that are phonologically voiced. (Ladefoged and Maddieson 1996: 50)

If sonorants and voiced obstruents do not differ in the laryngeal physiology of their voicing, then the difference between them must lie in the effect that the supralaryngeal configuration has on vocal fold vibration. Sonorants involve an unconstricted vocal tract, so that in these sounds sustained voicing is possible. Obstruents, on the other hand, involve a constricted vocal tract, which for aerodynamic reasons inhibits vocal fold vibration (e.g. Ohala 1983). The effect of this is that voiced obstruents lack a well-defined formant structure. This is presumably the reason why in many languages sonorants behave as a class to the exclusion of both voiceless and voiced obstruents. A case in point is a language like Dutch, where final devoicing targets voiced obstruents (13a) but leaves sonorants unaffected (13b).

(13)	a.	*bed*	/bɛd/	[bɛt]	'bed'	cf.	[bɛd-ə]	'beds'
		reis	/rɛis/	[rɛis]	'journey'		[rɛiz-ə]	'journeys'
	b.	*bel*	/bɛl/	[bɛl]	'bell'		[bɛl-ə]	'bells'
		trein	/trɛin/	[trɛin]	'train'		[trɛin-ə]	'trains'

For a detailed discussion of final devoicing in Dutch and other languages, see e.g. Warner *et al.* (2004), CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION, and references there.

A possible challenge to the claim that sonorants and obstruents do not differ in the physiology of their voicing comes from tone. Research on tone languages has shown that obstruent voicing often conditions a lower tone on a following vowel, and is in many cases the diachronic origin of low tone (e.g. Matisoff 1973; Bradshaw 1999; Yip 2002; see also CHAPTER 97: TONOGENESIS). Sonorants generally do not have such an effect, although there are exceptions, such as in Ewe (Smith 1968, cited in Yip 2002). It could therefore be suggested that the characteristics of obstruent voicing differ from those of sonorant voicing after all. This is the position taken by e.g. Yip, who, following Halle and Stevens (1971), maintains that voiced obstruents involve slacker vocal folds than sonorants. The reasoning behind this is aerodynamic: the oral constriction characteristic of obstruents raises the supraglottal air pressure, which slows down the rate of vocal fold vibration; to counteract this, the vocal folds are slackened, which lowers the pitch of a following vowel. The degree of oral constriction in sonorants, on the other hand, does not raise the air pressure to such an extent that this inhibits vocal fold vibration.

While this is not the place to focus on consonant–tone interaction in any detail, a couple of brief comments are in order. The experimental findings reported in Hombert *et al.* (1979) show that both voiced obstruents and sonorants trigger pitch lowering in a following vowel, both in tonal languages like Yoruba and in non-tonal ones like English. This is unexpected if obstruent-induced lowering is caused by aerodynamic factors. Hombert *et al.* reject this kind of explanation, arguing instead that in voiced obstruent stops the larynx is actively lowered during the latter part of the stop's closure phase, and that this is responsible for pitch lowering. However, as Hombert *et al.* themselves note, this account, too, fails to explain why sonorants, whose realization does not normally involve downward movement of the larynx, also lower the pitch of a following vowel.

best that can be said is that sonorant stops are consistently realized with *active* voice facilitation, and that the degree of voice facilitation is such that these sounds are voiced throughout their duration.

Thus, while the presence of voice facilitation cannot be used as a diagnostic for the sonorant status of a sound, two reasonable hypotheses can be made. The first is that active voice facilitation is a prerequisite for sonorant but not for obstruent stops. The second is that in languages where sonorant stops contrast with obstruent stops, the former involve a greater degree of voice facilitation than the latter. Support for the latter hypothesis comes from Bearlake Slave, a Northern Athapaskan language, which contrasts plain, aspirated, and ejective stops, and an additional series of voiced alternating stops. The description in Rice (1994, 2006) suggests that the realization of the plain stops is similar to the lenis stops of English. The voiced alternating series, on the other hand, is fully voiced and optionally prenasalized, suggesting a greater degree of voice facilitation. This phonetic difference correlates with a phonological difference. The two types of stops behave differently under perfective formation, where alternating stops are nasalized (14a) while non-alternating ones remain unaffected (14b) (cf. Rice 1993: 322–333; the vowel changes are due to an independent ablaut process):

(14) a.	*Alternating stops*		b. *Non-alternating stops*	
	-de	'win-IMPERF'	-da	'move-IMPERF'
	-nǭ	'win-PERF'	-dǭ	'move-PERF'

Rice analyzes the alternating stops as sonorants, specified for an SV node, and the plain stops as obstruents, unspecified for any laryngeal structure.

Some cases of class behavior of sonorant obstruents and sonorants appear to support the relevance of aerodynamic aspects of sonorancy. The most compelling evidence for this comes from Ikwere (Clements and Osu 2002). This language has a contrast between voiceless and voiced "explosive" stops, and an additional contrast between voiced and glottalized labial "non-explosive" stops. We saw in §3 that the latter display sonorant-like behavior (they alternate with nasals, do not occur in NC-clusters, and fail to trigger tone lowering). Clements and Osu transcribe the non-explosive stops as /ḅ/ and /'ḅ/, noting that they derive from historical labial-velars. Phonetically, /ḅ 'ḅ/ are non-implosive, and they are characterized by the absence of heightened air pressure. /'ḅ/ is pre-glottalized, and as a result voiceless throughout the initial part of the closure phase; the latter part of /'ḅ/ has modal voicing. /ḅ/ has modal voicing throughout. Importantly, the same is true of the voiced explosive /b/, which also has roughly the same duration as /ḅ/. The contrast between /b/ and /ḅ/ thus resides first and foremost in their air pressure characteristics. This difference certainly also leads to an acoustic difference (otherwise children acquiring Ikwere would not be able to discriminate between the two sounds), but Clements and Osu's data show that this difference does not seem to lie in the sounds' voicing characteristics. If the class behavior of non-explosives with sonorants is taken to be phonetically natural, then this can be achieved either by extending the correlates of sonorancy to include air pressure characteristics or, as Clements and Osu argue, by differentiating between an acoustic feature [sonorant] and an articulatory/aerodynamic feature [obstruent].

Alternatively, it might be argued that the phonetic motivation for class behavior of non-explosives with sonorants lies in the history of Ikwere. Such an account would have to show that the historical labial-velars can be reasonably treated as acoustically closer to sonorants. It is worth noting in this respect that labial-velars are sometimes produced with an ingressive airstream mechanism (e.g. Ladefoged 1968; Cahill 2008). This makes them similar to implosives, which frequently pattern as sonorants (see §3). It is also worth noting that Ikwere /ḅ 'ḅ/ are velarized, which, as Clements and Osu suggest, may be a reflex of their earlier velar articulation (cf. also Ladefoged and Maddieson 1996: 89, 343). Clements and Osu conjecture that velarization expands the oral cavity to such an extent that the non-explosive release of /ḅ 'ḅ/ is characterized by ingressive airflow, though, as noted, this does not appear to lead to more prominent voicing.

5 Discussion and conclusion

This chapter has considered three issues in the phonology of sonorants: (i) the representation of sonorancy in segmental structure (and its relation to laryngeal contrasts); (ii) the interpretation of fuzzy class behavior of sonorants and sonorant obstruents; and (iii) the phonetic correlates of sonorancy. As the preceding discussion makes clear, the thread that runs through these issues is the relation between sonorancy and voicing.

The main challenge facing a theoretical interpretation of sonorancy is to provide an adequate account of the observation that voicing is never contrastive in sonorants, but may nevertheless be required in their phonological specification. A further challenge concerns the observation that if voicing is not contrastive in sonorants, then sonorants evidently entertain a different relationship with laryngeal contrasts than do other segment types, such as stops and fricatives. It would seem that none of the segmental theories currently on offer can account for this difference in any straightforward way. A further challenge is posed by the class of "voiceless" aspirated sonorants, whose sonorant status is not entirely clear.

The relation between sonorancy and voicing is also relevant in the interpretation of fuzzy class behavior of sonorants and voiced stops and fricatives. In some languages where this is observed, such as Southern Barasano, the relevant class consists of all voiced sounds of the language, viz. sonorants and voiced stops, suggesting that the latter function phonologically as sonorants. In other languages, such as Cama and Bearlake Slave, some voiced sounds pattern with sonorants whereas others do not. In such languages, it appears to be the case that the sonorant obstruents involve active voice facilitation, manifested phonetically by such gestures as prenasalization and implosion, making their voicing more prominent than that of their voiced obstruent congeners.

Finally, the relation between sonorancy and voicing is also important with regard to the phonetic underpinnings of sonorant class behavior. There is good reason to believe that class behavior of true sonorants is phonetically natural in that these sounds are characterized by a clearly defined formant structure, an acoustic property which sonorants share to the exclusion of other sounds. Such a characterization is not only simpler than an articulatory one but also avoids the dubious notion of spontaneous voicing, for which little phonetic support has been found. The class behavior of sonorants and sonorant obstruents seems natural to the extent that both involve sustained vocal fold vibration. However, in some languages, such as Ikwere, sonorant stops do not appear to differ from their obstruent counterparts in terms of their voicing characteristics, suggesting that in this language sonorancy correlates with a combination of periodicity and lack of supraglottal air pressure buildup. While this is perhaps a "far-fetched notion," as Ladefoged asserts, there is recent evidence that listeners do indeed make use of the multisensory integration of perceptual events. Gick and Derrick (2009) show that *pa/ba* syllables are more likely to be heard as aspirated by English listeners (i.e. causing them to mishear *b* as *p*) when these are accompanied by cutaneous air puffs at the right hand or the neck, suggesting that information from the auditory and the tactile domain may combine to form a salient psychological percept. We should not be surprised, then, to find that similar multisensory integration is found within the auditory domain itself.

ACKNOWLEDGMENTS

The author acknowledges the assistance of the Netherlands Organization of Scientific Research (NWO) for funding project no. 2008/01214/GW. Thanks are also due to two anonymous reviewers and to the editors of this volume for useful and constructive comments, suggestions, and criticism. One reviewer deserves special thanks for drawing my attention to Gick and Derrick (2009). Needless to say, none of them is responsible for any remaining errors of fact and interpretation.

REFERENCES

Anderson, John M. & Colin J. Ewen. 1987. *Principles of dependency phonology*. Cambridge: Cambridge University Press.

Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5. 183–207.

Avery, Peter. 1996. The representation of voicing contrasts. Ph.D. dissertation, University of Toronto.

Bárkányi, Zsuzsanna & Zoltán Kiss. 2006. A phonetically-based approach to the phonology of [v] in Hungarian. *Acta Linguistica Hungarica* 58. 175–226.

Blaho, Sylvia & Patrik Bye. 2005. Cryptosonorants and the misapplication of voicing assimilation in biaspectual phonology. Paper presented at the 13th Manchester Phonology Meeting.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Botma, Bert. 2004. Phonological aspects of nasality: An element-based dependency approach. Ph.D. dissertation, University of Amsterdam.

Botma, Bert. 2005. On the interpretation of laryngeally modified nasals. In van Oostendorp & van de Weijer (2005), 255–286.

Botma, Bert. 2008. Aspiration, fricatives and the multiple feature hypothesis in Icelandic. Paper presented at the Workshop on Phonological Voicing Variation, Meertens Institute, Amsterdam.

Botma, Bert. 2009. Transparency in nasal harmony and the limits of reductionism. In Kuniya Nasukawa & Philip Backley (eds.) *Strength relations in phonology*, 79–111. Berlin & New York: Mouton de Gruyter.

Botma, Bert & Norval Smith. 2006. A dependency account of the fortis–lenis contrast in Cama. In Jeroen van de Weijer & Bettelou Los (eds.) *Linguistics in the Netherlands 2006*, 15–27. Amsterdam & Philadelphia: John Benjamins.

Botma, Bert & Norval Smith. 2007. A dependency-based typology of nasalization and voicing phenomena. In Bettelou Los & Marjo van Koppen (eds.) *Linguistics in the Netherlands 2007*, 36–48. Amsterdam & Philadelphia: John Benjamins.

Bradley, David. 1979. *Proto-Loloish*. London & Malmö: Curzon Press.

Bradshaw, Mary M. 1999. A crosslinguistic study of consonant–tone interaction. Ph.D. dissertation, Ohio State University.

Cahill, Michael C. 2008. Why labial-velar stops merge to /gb/. *Phonology* 25. 379–398.

Campbell, Lyle. 2004. *Historical linguistics: An introduction*. 2nd edn. Cambridge, MA: MIT Press.

Capo, Hounkpatin C. 1981. Nasality in Gbe: A synchronic interpretation. *Studies in African Linguistics* 12. 1–43.

Catford, J. C. 1977. *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. 2000. Phonology. In Bernd Heine & Derek Nurse (eds.) *African languages: An introduction*, 123–160. Cambridge: Cambridge University Press.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

Clements, G. N. & Sylvester Osu. 2002. Explosives, implosives and nonexplosives: The linguistic function of air pressure differences in stops. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology 7*, 299–350. Berlin & New York: Mouton de Gruyter.

Corrigan, Roberta, Fred Eckman & Michael Noonan (eds.) 1989. *Linguistic categorization*. Amsterdam & Philadelphia: John Benjamins.

Dell, François & Mohamed Elmedlaoui. 2002. *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht: Kluwer.

Durand, Jacques & Francis Katamba (eds.) 1995. *Frontiers of phonology: Atoms, structures, derivations*, 34–79. London & New York: Longman.

Ewen, Colin J. (1980). The characterization of glottal stricture in dependency phonology. *York Papers in Linguistics* 8. 35–47.

Ewen, Colin J. & Harry van der Hulst. 2001. *The phonological structure of words*. Cambridge: Cambridge University Press.

Flemming, Edward, Peter Ladefoged & Sarah Thomason. 1994. Phonetic structures of Montana Salish. *UCLA Working Papers in Phonetics* 87. 1–33.

Gerfen, Chip. 1999. *Phonology and phonetics in Coatzospan Mixtec*. Dordrecht: Kluwer.

Gick, Bryan & Donald Derrick. 2009. Aero-tactile integration in speech perception. *Nature* 462. 502–504.

Giegerich, Heinz J. 1992. *English phonology: An introduction*. Cambridge: Cambridge University Press.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Malden, MA & Oxford: Blackwell.

Hale, Mark & Charles Reiss. 2000. Phonology as cognition. In Noel Burton-Roberts, Philip Carr & Gerard J. Docherty (eds.) *Phonological knowledge: Conceptual and empirical issues*, 161–184. Oxford: Oxford University Press.

Halle, Morris. 1992. Features. In William Bright (ed.) *Oxford international encyclopedia of linguistics*, vol. 3, 207–212. New York: Oxford University Press.

Halle, Morris & G. N. Clements. 1983. *Problem book in phonology*. Cambridge, MA: MIT Press.

Halle, Morris & Kenneth N. Stevens. 1971. A note on laryngeal features. *MIT Research Laboratory of Electronics Quarterly Progress Report* 101. 198–213.

Halle, Morris, Bert Vaux & Andrew Wolfe. 2000. On feature spreading and the representation of place of articulation. *Linguistic Inquiry* 31. 387–444.

Hamann, Silke. 2006. The hybrid nature of voiced labiodentals (especially in German). Paper presented at the 14th Manchester Phonology Meeting.

Hankamer, Jorge & Judith Aissen. 1974. The sonority hierarchy. *Papers from the Annual Regional Meeting Chicago Linguistic Society: Parasession on natural phonology*, 131–145.

Harris, John. 1996. Phonological output is redundancy-free and fully interpretable. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, vol. 1, 305–332. Salford: ESRI.

Harris, John. 2006. The phonology of being understood: Further arguments against sonority. *Lingua* 116. 1483–1494.

Harris, John & Geoff Lindsey. 1995. The elements of phonological representation. In Durand & Katamba (1995), 34–79.

Hjelmslev, Louis. 1953. *Prolegomena to a theory of language*. Baltimore: Indiana University Publications in Anthropology and Linguistics.

Hombert, Jean-Marie, John J. Ohala & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55. 37–58.

Hulst, Harry van der. 1995. Radical CV phonology: The categorial gesture. In Durand & Katamba (1995), 80–116.

Hyman, Larry M. 1973. The feature [grave] in phonological theory. *Journal of Phonetics* 1. 329–337.

Hyman, Larry M. 1975. *Phonology: Theory and analysis*. New York: Holt, Rinehart & Winston.

Itô, Junko & Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17. 49–73.

Itô, Junko, Armin Mester & Jaye Padgett. 1995. Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26. 571–613.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Kaye, Jonathan. 1981. Implosives as liquids. *Studies in African Linguistics*. Supplement 8. 78–81.

Kehrein, Wolfgang. 2002. *Phonological representation and phonetic phasing: Affricates and laryngeals*. Tübingen: Niemeyer.

Kehrein, Wolfgang & Chris Golston. 2004. A prosodic theory of laryngeal contrasts. *Phonology* 21. 325–357.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kingston, John & David Solnit. 1988. The tones of consonants. Unpublished ms., Cornell University & University of Michigan.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Krauss, Michael & Jeff Leer. 1981. *Athabaskan, Eyak and Tlingit sonorants*. Fairbanks: Alaska Native Language Centre.

Kristoffersen, Gjert. 2000. *The phonology of Norwegian*. Oxford: Oxford University Press.

Kuipers, Aert. 1974. *The Shuswap language: Grammar, texts, dictionary*. The Hague & Paris: Mouton.

Ladefoged, Peter. 1968. *A phonetic study of West African languages: An auditory-instrumental study*. 2nd edn. Cambridge: Cambridge University Press.

Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics*. Chicago: Chicago University Press.

Ladefoged, Peter. 1982. *A course in phonetics*. 2nd edn. New York: Harcourt Brace Jovanovich.

Ladefoged, Peter. 1997. Linguistic phonetic descriptions. In W. J. Hardcastle & John Laver (eds.) *The handbook of phonetic sciences*, 589–618. Cambridge, MA & Oxford: Blackwell.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford & Malden, MA: Blackwell.

Lass, Roger. 1984. *Phonology: An introduction to basic concepts*. Cambridge: Cambridge University Press.

Laver, John. 1994. *Principles of phonetics*. Cambridge: Cambridge University Press.

Li, Fang Kuei. 1977. *A handbook of comparative Tai*. Honolulu: University Press of Hawaii.

Lombardi, Linda. 1994. *Laryngeal features and laryngeal neutralization*. New York: Garland.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Malsch, Derry L. & Roseanne Fulcher. 1989. Categorizing phonological segments: The inadequacy of the sonority hierarchy. In Corrigan *et al.* (1989), 69–80.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In Larry M. Hyman (ed.) *Consonant types and tone*, 71–95. Los Angeles: University of Southern California.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

Mielke, Jeff. 2005. Ambivalence and ambiguity in laterals and nasals. *Phonology* 22. 169–203.

Nasukawa, Kuniya. 2005. *A unified approach to nasality and voicing*. Berlin & New York: Mouton de Gruyter.

Nathan, Geoffrey S. 1989. Preliminaries to a theory of phonological substance: The substance of sonority. In Corrigan *et al.* (1989), 55–67.

Noske, Manuela. 1995. The ternary use of distinctive features: In defence of the number three. Ph.D. dissertation, University of Chicago.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.) *The production of speech*, 189–216. New York: Springer.

Ohala, John J. 1990. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26(2), 319–338.

Ohala, John J. & Manjari Ohala. 1993. The phonetics of nasal phonology: Theorems and data. In Mary K. Huffman & Rena A. Krakow (eds.) *Nasals, nasalization, and the velum*, 225–249. San Diego: Academic Press.

Ohala, John J. & Carol J. Riordan. 1979. Passive vocal tract enlargement during voiced stops. In Jared J. Wolf & Dennis H. Klatt (eds.) *Speech communication papers*, 89–92. New York: Acoustical Society of America.

Oostendorp, Marc van & Jeroen van de Weijer (eds.) 2005. *The internal organization of phonological segments*. Berlin & New York: Mouton de Gruyter.

Padgett, Jaye. 2002. Russian voicing assimilation, final devoicing, and the problem of [v]. Unpublished ms., University of California, Santa Cruz (ROA-528).

Perkell, Joseph S. 1969. Physiology of speech production: Results and implications of a quantitative cineradiographic study. Cambridge, MA: MIT Press.

Piggott, Glyne L. 1992. Variability in feature dependency: The case of nasality. *Natural Language and Linguistic Theory* 10. 33–77.

Piggott, Glyne L. 1993. The geometry of sonorant features. Unpublished ms., McGill University.

Piggott, Glyne L. & Harry van der Hulst. 1997. Locality and the nature of nasal harmony. *Lingua* 103. 85–112.

Pulleyblank, Douglas. 1989. Patterns of feature cooccurrence: The case of nasality. In S. Lee Fulmer, Masahide Ishihara & Wendy Wiswall (eds.) *Proceedings of the Arizona Phonology Conference* 2, 98–115. Tucson: University of Arizona.

Rice, Keren. 1993. A re-examination of the feature [sonorant]: The status of "sonorant obstruents." *Language* 69. 308–344.

Rice, Keren. 1994. Laryngeal features in Athapaskan languages. *Phonology* 11. 107–147.

Rice, Keren. 2006. "Free" variation in Slave (Northern Athapaskan): Relating fieldwork and phonology. Paper presented at the 14th Manchester Phonology Meeting.

Rice, Keren & Peter Avery. 1989. On the interaction between sonorancy and voicing. *Toronto Working Papers in Linguistics* 10. 65–82.

Rice, Keren & Peter Avery. 1991. On the relationship between laterality and coronality. In Carole Paradis & Jean-François Prunet (eds.) *The special status of coronals: Internal and external evidence*, 101–124. San Diego: Academic Press.

113 Flapping in American English

Kenneth J. de Jong

1 Original insights

In the mid-1930s, several scholars of American English from around the USA noted a peculiarity that, apparently, was becoming a pervasive and characterizing feature of the production of coronal stops in certain intervocalic positions. A high-profile notation of the phenomenon was by John Kenyon in the sixth edition of *American Pronunciation*, where he notes the presence of /t/s that are voiced in American, as opposed to British, productions. What gathered more original scholarly attention to this phenomenon is the question of how to characterize these productions. Haugen (1938) noted in particular that the productions of /t/ in words like *pity* had become so different from that in words like *typical* that classifying them as the same segment was becoming untenable, and so he posited the presence of an allophone of /t/ in American English.

Haugen noted a variety of aspects of the allophonic variant. Along with Kenyon, he noted the dialectally specific occurrence of voicing and lack of aspiration, but noted in addition a shortening and weakening of the closure. The closure weakening, then, contributes to a classificatory issue as to whether the allophone of /t/ is, in fact, not just a voiced version of the same consonant, but actually crosses into the same phonetic region inhabited by liquid sonorants. In order to evaluate this possibility, he compared the allophone with other liquid sonorants, with British rhotics, e.g. between British *porridge* and American *pottage*, with cross-language mapping comparisons, and with non-native substitution of rhotics for the allophone, e.g. Italian American *siri* for *city*. In addition, there had been many earlier observations of British dialectal variants involving substitution of /r/ for /t/, noted by Jespersen (1928), which probably indicate the seeds of the characterizing American feature in British dialects. These observations led various scholars, such as Bloomfield (1933), to describe the allophone as a ballistic flap (or "flip"). In these early discussions of segmental classification, a prominent issue is how to deal with the apparent level of variability in the voiced allophone, with Haugen (1938) suggesting that the variability in production is general to the nature of these sorts of ballistic rhotics. Oswald (1943), on the other hand, suggests three distinguishable outcomes – a voiced stop, a tap, or a "flip" (presumably, a coronal approximant) – but also states his belief that the three representations are probably just artificial divisions of a single continuum.

Complicating the analysis of the phenomenon was the realization by Kenyon (1936) that the process is sometimes a neutralizing process, collapsing productions of /t/ with those of /d/. Again, earlier works had noted dialectal (Wright 1905) or lexically specific (Krapp 1919) substitution of /d/ for /t/ in the flapping environment. While the general discussions in the 1930s usually did not treat the process as neutralizing (Bloomfield 1933; Twaddell 1935; Trager 1942), both Haugen and Kenyon noted that actual neutralization of the /t/–/d/ contrast can occur in the speech of some speakers. Even more striking, Oswald (1943) a little later documented the pervasive nature of the neutralization, in an experiment in which high school students were asked to produce and identify 14 pairs of potential homophones in sentential contexts; there was no difference in identification of items produced with orthographic *t* and *d*. However, Haugen (1938) in particular noted that the duration of the preceding vowel might not be neutralized, and suggests this as an important area of research. Also of interest in these discussions is the question of whether the outcome of the process is directional, creating [d] from /t/, or whether, as claimed by Haugen (1938), the process collapses both /t/ and /d/ into a third category, as is typical of later analyses.

A second major subject of the early discussions concerned how to characterize the context in which the allophonic shift happens. Haugen (1938), for instance, corrected Uldall's (1934) claim that voicing shift happens in final position. He then went on to summarize Kenyon's claims with a fairly simple formula for the allophone's occurrence:

> [I]t cannot occur unless it is *preceded* by a vowel or a sonorant (*n, l, r*), and is *followed* by an unstressed syllable-forming element (vowel, *l, r*, but not the homorganic *n*). (1938: 631)

His basic analysis has withstood the test of time; it is remarkably similar to those found in most of the subsequent discussions of the phenomenon (e.g. Kahn 1976; Hayes 1995).

One complicating aspect of the early observations is a failure to distinguish cases which later would be treated as a slightly different phenomenon, the glottalization of /t/ in word-final position. Thus, Kenyon's (1936) treatment considered the lack of aspiration in a sequence like *get here* as another example of the same flapping phenomenon, complicating his analysis of the overall pattern of occurrence.

Haugen (1938) noted various segmental restrictions on the rule as well. Specifically, he noted that a preceding consonant does not block the weakening, nor do following syllabic consonants, except /n/. He also noted that the flapping of /t/ is gradiently sensitive to consonant context, with greater weakening after /n/ than /r/; Oswald similarly suggests that the outcome of the process after /r/ is actually a voiced stop. Further, in the post-nasal context, Haugen noted that both the stop and the homorganic nasal are subject to weakening. He also noted greater weakening before /r/ than before /l/.

While the early works on flapping thus covered much of what is currently known about the phenomenon, although only in seminal form, one aspect of the phenomenon that is wholly absent is the question of how and whether this segmental effect might generalize beyond the coronal consonants. In all of the treatments in process might be profitably understood in terms of a gradient set of production constraints whose application is determined by the exigencies of speech rate and segmental context. This gradient variation, then, becomes quantized in the acoustic domain, yielding quasi-categorical output variation. Fujimura (1986) similarly suggested that flapping might be the output of tongue–jaw interactions, such that lower jaw positions tend to weaken and retract coronal closures, yielding the typical acoustic effects of flapping.

If this is the case, it raises the question of whether the flapping phenomenon requires any categorical shift at all, i.e. it could be, then, that flapping results from the dynamics of the speech motor system, rather than from a specification of a categorically different speech goal. To make this sort of explanation tenable, it is necessary to link the conditions on the occurrence of flaps to processes that would yield the appearance of flapping as a by-product without the need to stipulate an overt rule. The advantage to this approach is potentially on two fronts. First, if a general characterization of the prosodic and segmental conditions of the flapping rule will get the outcome of flapping automatically, the statement of an additional phonological process becomes superfluous. Developing models of prosodic and segmental dynamics is a pressing research goal in its own right, and if it obviates a phonological rule, so much the better. Second, this approach would likely be amenable to explaining the troublesome variability in the outcome of the rule that has been noted since the earliest detailed descriptions of the phenomenon.

De Jong (1998) reported a series of analyses that attempt to explain flapping in this way, as it occurs in a small corpus of singleton consonants in coda position before a vowel-initial word. This study corroborated Oswald's (1943) impressions that the stops vary in transcription along a continuum from a flap to a fully articulated [t], with many instances of orthographic *t* being transcribed as a fully articulated [d]. While this continuum of weakening would appear to be amenable to explanation as an articulatory-to-acoustic by-product, the acoustic and articulatory analyses in de Jong (1998) failed to find any straightforward way of characterizing the articulatory processes that would yield the right impressionistic outcomes. The candidates examined there included jaw dynamics and timing models, both of which failed to make the right predictions concerning causal factors. For example, while jaw mechanics would suggest that flaps result from lower jaw positions, the data indicate no systematic difference in actual jaw position during the occlusion for flaps and stops. Similar problems were found for treating flapping as a sort of durational undershoot induced by the proximity of neighboring vowels. Here, while there are different tongue body positions for flaps and stops, such that the tongue body is closer to the positioning for the vowel with flaps, this difference in overall lingual posture is more apparent during the closure itself than it is during the vowel that is supposed to be driving the occurrence of flapping.

Fukaya and Byrd (2005) examine the dynamics of overlap between the consonant and vowel in the /t/-flapping cases in greater detail, adding an additional condition in which a phrase boundary follows the target stop. They find that the transcribed flaps in different conditions with different talkers exhibit different kinematic patterns. One general observation from their data is that, despite the traditional description of flaps as being "ballistic" in nature, there is no apparent increase in the acceleration peaks for flaps over non-flaps. In general, what is consistent across events transcribed as flaps is the occurrence of a short acoustic with glottal trans-illumination (de Jong *et al.* 2002) find three very distinct glottal movement patterns for /t/s in different conditions. Onset and typical slow-rate coda /t/s exhibit large glottal openings corresponding to aspiration; in coda /t/s the glottal opening may fluctuate with a glottal closing gesture corresponding to glottalization; at fast rates, productions of coda /t/ typically exhibit no large glottal movements, yielding what would be perceived as instances of flapping. Thus, even in extreme production circumstances, glottalization is a distinct outcome from flapping.

However, this does not solve the larger problem of understanding how the environments for flapping and glottalization get resolved. Selkirk (1982) noted the inherent competition for targets between these two processes, and differentiated them by reference to a release feature on the target; consonants in the flapping environment have a release characteristic in the relevant prosodic domain, while glottalized targets are final and unreleased. Thus the apparent difference between the two would be one of prosodic domain, whether a boundary corresponds to the consonant release. However, there are obvious cases in which coronal stops in the glottalization position are flapped, even though the constraints on the flapping rule are not met. An example of this is in function word clusters such as *at all*. Here, the medial stop is typically flapped, even though the following vowel is stressed (cf. *atoll*, which has an aspirated stop). These forms suggest a possible role of word sequence frequency in triggering the rule, although it is not clear how such a trigger would work in descriptions of the process.

Other research in prosodic conditioning of flapping locates the level of affiliation at a lower level, the level of the syllable. All logically possible affiliations of the affected stops have been espoused. Selkirk (1982), due to the general observation that the onset allophone is really the odd man out, suggested that the consonant weakening is due to the stop being resyllabified (or in the case of cross-word-boundary application, syllabified) as a coda. Davis and Summers (1989) concur with this analysis. Kahn (1976) and Gussenhoven (1986) – to create a structure that would distinguish the flapping environment from the coda environment which would trigger glottalization – posit ambisyllabicity as the conditioning factor, doubly associating the medial consonant. De Jong (1998) – to complete the typology of possible analyses – suggested that the weakened consonants are onsets to the following syllable, based on the analogy of flapping with vowel reduction; both effects would be weakening processes that target unstressed syllables. Elicited syllabifications of flaps in Eddington and Elzinga (2008) show that, while aspiration is strongly associated with onset syllabification, flapped stops tend to be syllabified in various ways, depending on a number of contextual factors.

This disagreement in syllabic representation indicates a more fundamental difference of opinion as to the underlying mechanisms that should account for the occurrence of flapping, and that is whether flapping is refllex of syllable-level or higher-level (foot- or phonological phrase-level) position, or whether it is more properly viewed as a segmental effect.

The prosodic position is taken up by Davis (2005), who along with Steriade (2000) points out that the medial *t* in words like *Mediterranean* should be flapped, in the classic specification of the flapping environment. However, stops in this position, in parallel with non-coronals (VanDam 2003), are systematically non-neutralized and even lightly aspirated. This, according to Davis (2005) (Steriade espouses a different position, and notes these forms as exceptions), is due to the stop occurring at the beginning of a metrical foot (see CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE and CHAPTER 40: THE FOOT for further instances of foot-internal segmental processes). The flapping rule, then, is an indicator of a foot-medial and post-stress position. Pre-stress positions are regularly marked with aspiration with voiceless stops, and so fit with the "normally articulated" stops that Kenyon and Haugen posited for onset positions. A complicating factor in these analyses, and one that is the focus of other recent analyses, is a typical difference noted by Withgott (1982), and confirmed by Steriade (2000), that words like *capitalistic* with a base form containing a /t/ in flapping position tend to be flapped, while words like *militaristic* with a base form with /t/ in a non-flapping pre-stress position tend not to be. One problem with the *militaristic/capitalistic* pair is that the unflapped *t* is before a rhotic, which can syllabify with the previous *t*, creating a cluster, while the flapped *t* is before a lateral, which cannot form a cluster. However, this sort of paradigm uniformity effect (see CHAPTER 83: PARADIGMS) can also be seen in the data in Turk (1992), where the /t/ in *diameter* is clearly flapped, while the /d/ in *pyramidal* appears not to be (since closure durations are exceptionally long). Here, the /d/ in the base form is in final position. Also, Withgott (1982) notes a similar contrast between unflapped *militarize* and flapped *parameterize*, both of which have coronal stops followed by a rhotic.

One final point about prosodic analyses is that the flapping phenomenon is similar to consonant lenition processes in other languages, such as "spirantization" (actually "approximantization," usually) of voiced stops in most varieties of Spanish (see CHAPTER 66: LENITION). Lewis (2001) shows conclusively that this weakening process is an indicator of a medial position in a prosodic phrase. Flapping would be similar, although the domain indicated would be smaller, sometimes smaller even than a lexical item.

2.5 *The scope of flapping*

The previous discussion of prosodic conditions on flapping points out that the flapping rule, although often not noted as such, should be situated within an array of parallel lenition processes that affect other consonants. The parallelism is more obvious if we examine the voiceless stops in onset position. In onset (and foot-initial) positions, all three voiceless stops in English exhibit aspiration. This aspiration is systematically modulated by various prosodic factors, such as the occurrence of a preceding sibilant, which removes aspiration from all three stops. In addition, the presence of heavy stress increases aspiration duration for all three stops, and the aspiration that is foot-initial is weaker than the aspiration that is immediately preceding a stressed vowel (VanDam 2003). All of these facts about aspiration are systematically found for all three stops.

When we switch to non-initial positions, we also find some parallelism, in that all three stops can be subject to glottalization, although glottalization is much more common in the coronals than in the others. A very similar situation holds with the flapping environment; there is a greater proportion of voicing in closure and a much shorter closure not only for the coronal targets of the flapping rule, but also for dorsal and labial consonants as well. This observation is one of the driving factors in such works as de Jong (1998), which seek to explain flapping as part of a larger prosodic organizational process. While positing a flapping rule of some sort will explain the coronal process, such a rule would not account for the parallel, although less obvious, weakening processes in the other points of articulation.

On the reverse side, a persistent challenge to explaining flapping as a general prosodic process is the fact that the coronals are markedly different from the labials and dorsals in degree of weakening. While coronals are commonly noted as being neutralized, labial and dorsal consonants in the same environment, while perhaps exhibiting some voicing confusion, cannot be said to be neutralized in speakers of American English. For example, Davis and Summers (1989) compare coronals with labial and dorsal stops, finding much shorter closure durations for coronals, as well as very little difference between /t/ and /d/ closure which differs from stops at other points of articulation that have shorter closures with voiced stops. Turk (1992) finds even more striking differences by point of articulation.

The approach taken in de Jong (1998) was to point to the linkage of the coronal articulation to the jaw. Previous work on interarticulator variability, e.g. de Jong (1995), notes that coronal consonants tend to impinge on jaw movement patterns more than do dorsal and labial consonants. This linkage to the jaw, then, if we reverse it, would suggest that coronal articulations are particularly vulnerable to jaw perturbation by other speech goals, such as those for the neighboring vowels. Thus, coronals may be particularly susceptible to neighboring segment effects. It should be noted that such a model was not particularly successful at explaining the articulatory patterns in de Jong (1998).

A problem, then, with this generalized approach is that lenition processes, e.g. "spirantization" in Spanish, do not exhibit this radical difference between coronal and non-coronal segments. If anything, studies of spirantization such as Lewis (2001) indicate that the dorsal closures are most weakened by vocalic environment as would be expected on the basis of overlap of tongue body articulation between the stop and the neighboring vowels. Considering the weakening of voicelessness, Lewis (2001) finds labial stops exhibit the greatest amount of decrease in voicelessness and noise generation. While there may be some additional consideration, such as the fact that Spanish coronals tend to exhibit more anterior articulations than those in American English, the difference between the lenition processes in the two languages is striking.

There are further differences that emerge as one tries to generalize the flapping rule as an example of general medial lenition. For example, a persistent issue for functional explanations of lenition processes is the different outcomes for different languages. Hock's (1991: 83) treatment of historical lenition processes shows clearly that lenition processes of this kind can run in two different directions; either they can tend toward closure weakening and hence spirantization, or they can tend toward aerodynamic weakening and hence voicing and sonorantization. The traditional treatment of Spanish lenition suggests a lenition process of the former variety, while flapping appears to be of the latter variety.

It is possible that part of the explanation of the difference in outcome between, for example, Spanish and English medial lenition lies in the historical paths of the two phenomena. An intriguing oddity in Haugen's (1938) description of the weakened coronal stops is that they should be analyzed as spirants, in parallel to the weakened sounds at other points of articulation. This description is parallel to traditional descriptions of Spanish. While it might indicate an analytic bias on Haugen's part (perhaps even from his contemporaries' analyses of languages such as Spanish), it could also indicate that the "flaps" of the time period were more like fricatives.

A different line of reasoning with respect to the difference between Spanish and English would lie in the phonemic inventory of the languages and the potential for a neutralization of the stop categories with short rhotics in Spanish (Gurevich 2004). Gurevich proposes a limiting constraint on lenition processes of a functional pressure to maintain contrasts. Since English has an approximant rhotic, the weakened coronal stops can safely become flaps without incurring neutralization with the rhotics; the same outcome in Spanish would create (near-)neutralization with the Spanish rhotic. (This observation was inherent in the discussion of stop weakening as a flapping rule in the 1930s.) On the other hand, spirantization of the stops in English would incur (near-)neutralization of the stops with the dental fricatives. (See also CHAPTER 30: THE REPRESENTATION OF RHOTICS.)

One aspect of this analysis is that is incomplete, however, is the fact that it does not explain why it is that the American English coronal is exceptional, while the Spanish coronal is not. While Gurevich examines cases of lenition irregularity across point of articulation, there is no apparent source for the coronal exceptionality, based on contrastive pattern. While the dorsal non-sibilant fricative does not exist in English, the parallelism between non-sibilant labials and coronals is fairly striking, and yet the coronal is subject to extensive sonorantization lenition, while the labial is not. (For further discussion of cross-linguistic coronal exceptionality, see CHAPTER 12: CORONALS.)

2.6 *The larger context of flapping*

If we consider general models, it becomes readily apparent that, whatever model of explaining flapping might be attempted, the model must account for a critical fact about the rule, that it specifically indicates American English. The fact that Spanish does not flap might be attributed to the fact that Spanish has different phonemic contrasts. However, it is not only Spanish that does not flap; other varieties of English do not either. Whatever information can be gleaned from generalizing models, this historically rooted fact needs to be accounted for.

Interestingly, Haugen's (1938) analysis posited the outcome of the rule as being a spirant. This is curious, given the fairly accurate level of observational detail in the analysis in other respects. It is possible that he was simply in error in this observation. However, it is also possible that the outcome of the lenition process in the 1930s was a spirant of some sort. If this is the case, the process has actually changed within the 70 years of observation, since spirantization is not found in any later instrumental analysis. Current varieties of Southern British English, however, are replete with spirantized productions of coronal stops in the flapping environment.

What these observations suggest is that flapping historically might be better seen as at least a two-stage process, including an initial weakening and neutralization stage with a more obvious connection between the segmental and prosodic triggers of the process and the weakened outcome of the process. Currently, however, it is clear that the process goes much further than what would be expected from a general weakening process. If this is true, the peculiar feature of American English lenition as creating flaps could be a later development from a more typical development of spirants as seen in other languages. Hence, what we see currently is a highly grammaticized version of a lenition process.

This is what would be expected from a process of grammaticalization that takes variability rooted in the exigencies of natural production and categorizes this gradient variability into a separate segment. The new category generally fits what we would expect if it were merely an outcome of general segmental co-articulation, or if it were just another example of prosodic modulation. However, upon closer inspection, we find that the co-articulatory effects and prosodic effects in this case are not the same as in parallel cases, since there appears to be a different category of segment. This segment has the prosodic contextual effects "packed in it"; the formerly transparent contextual triggers' effects get encoded in the consonant itself.

Two further considerations should be included at this juncture, however. The first concerns the conditions that would grammaticize the lenition process. While it is quite difficult to ascertain at this juncture, the documentation examined above would suggest that, while sporadic cases of coronal lenition appear in most varieties of English, and lexically specific cases are documented in the early twentieth century, it was in the 1930s that the literature on the phenomenon expanded rapidly and instrumental studies begin finding robust support for it. While this is quite speculative, it does appear as if the specific lenition of coronal stops to flaps rapidly became a marker of American English. This general historical sketch would suggest that an important consideration in the development of categorical flapping is specifically social in nature, connected with the continued development of a specifically American English in the mid-twentieth century, following the First World War.

The second consideration that must be brought in at this juncture concerns the pervasive issue of variability in production (see CHAPTER 92: VARIABILITY). One strike in favor of the segmental or prosodic by-product approach to flapping is that of understanding the gradient nature of the outcome of the process. This remains an important consideration for anyone wishing to understand the process; while the various linguistic factors in the process – segmental and prosodic – do much for characterizing the process, the actual occurrence of the phenomenon in running speech is very messy. One consideration in explaining this degree of variability is the fact that the process is largely (despite descriptions by structuralist phonologists up to the present day) a neutralizing process. While much is made of the non-homophonous nature of flapped *t* and flapped *d*, the fact is that /t/–/d/ minimal pairs are distinguished very poorly relative to other segmental contrasts.

This neutralizing property is something that speakers of American English, most of whom are literate, must come to grips with during the process of learning to read (even if they never notice any morphological relationship between words with flapped and non-flapped stops). This focus on orthography ensures that nearly all speakers under the right conditions will produce a flapped stop as its unflapped variant. This pool of unflapped versions is continuously being contributed to the pool of experiences that speakers encounter. In addition, the presence of unflapped variants is even more likely in speech directed at children in the learning process. This, in turn, ensures that there will be some proportion of speakers that, during the process of acquisition will do such things as produce words such as *Florida* with a voiceless [t], creating cases in which the distribution

Stetson, Raymond H. 1928. *Motor phonetics*. Haarlem: Société Hollandaise de Science.

Stetson, Raymond H. 1950. *Motor phonetics: A study of speech movements in action*. Amsterdam: North-Holland.

Stone, Maureen & S. Hamlet. 1982. Variation in jaw and tongue gestures observed during the production of unstressed /d/s and flaps. *Journal of Phonetics* 10. 401-415.

Trager, George L. 1942. The phoneme "T": A study in theory and method. *American Speech* 17. 144–148.

Treiman, Rebecca, Marie Cassar & Andrea Zukowski. 1994. What type of linguistic information do children use in spelling? The case of flaps. *Child Development* 65. 1318–1337.

Turk, Alice. 1992. The American English flapping rule and the effect of stress on stop consonant durations. *Working Papers of the Cornell Phonetics Laboratory* 7. 103–133.

Twaddell, W. Freeman. 1935. *On defining the phoneme*. Baltimore: Waverly Press.

Uldall, H. J. 1934. ə nout ɔn vauəl-leŋθ in əmerikən iŋgliʃ. *Le Maître Phonétique* 48. 97–98.

Umeda, Noriko. 1977. Consonant duration in American English. *Journal of the Acoustical Society of America* 61. 846–858.

VanDam, Mark. 2003. Voice onset time of American English stops with prosodic correlates. *Journal of the Acoustical Society of America* 113. 2328.

Withgott, Mary Margaret. 1982. Segmental evidence for phonological constituents. Ph.D. dissertation, University of Texas, Austin.

Wolfram, Walter & Robert Johnson. 1982. *Phonological analysis: Focus on American English*. Washington, DC: Center for Applied Linguistics.

Wright, Joseph. 1905. *The English dialect grammar*. Oxford: Frowde.

Zue, Victor W. & Martha Laferriere. 1979. Acoustical study of medial /t, d/ in American English. *Journal of the Acoustical Society of America* 66. 1039–1050.

114 Bantu Tone

Laura J. Downing

1 Introduction

1.1 Some background

There are roughly 500 Bantu languages, spoken in Africa south of a line going from Cameroon, in the west, to southern Somalia, in the east. It is a large language family both in terms of number of languages and in terms of number of speakers: roughly 240 million Africans speak Bantu languages, and several of the more well-known Bantu languages – like Swahili, Kikuyu, Zulu, or Kongo – have a few million speakers each (Nurse and Philippson 2003a).

Proto-Bantu is reconstructed with two contrasting tone levels.[1] While many Bantu languages remain poorly described, especially with regard to their prosodic systems, the available descriptions show that most modern Bantu languages have maintained this two-tone system. Only a few either are non-tonal, like Swahili, or have more than two phonological tone levels, like Kamba. Tonal minimal pairs are not common, but one finds them, as shown by the Jita data in (1a). The data in (1b) show that tone can be morphemic in Jita, as in other Bantu languages. Note that only tone pattern distinguishes between the perfective and distant past II tenses; segmentally, they are identical.[2]

[1] See Kisseberth and Odden's (2003) overview chapter on Bantu tone for a more detailed discussion of the issues summarized very briefly in this section. Most of the individual language and language group chapters in Nurse and Philippson's (2003b) compendium discuss tone and so provide the interested reader with an idea of the commonalities and the diversity of tone systems in this large language family.

[2] The data presented in this chapter copy the transcription conventions of the source cited, unless explicitly indicated otherwise. Readers should therefore not mistake these transcriptions for IPA representations, as they are influenced by the orthographic conventions of the respective languages. The original sources should, instead, be consulted for more information about phonetic details of pronunciation that are outside the scope of the issues discussed in this chapter. Long vowels are written as geminates and tone is marked independently on each vowel: e.g. [áá] is a long High-toned 'a'; [áa] is a long Falling-toned 'a'; [aá] is a long Rising-toned 'a'; [aa] is a long Low-toned 'a'.

(1) *Jita tonal minimal pairs* (Downing 1996: 48, 63, 183)

a.	oku-ííga	'to look for'	oku-iiga	'to pass (a test)'
b.	aa-lamuuye	's/he has already decided (PERFECTIVE)'	aa-lamúúyê	's/he decided (DISTANT PAST II)'

As Kisseberth and Odden (2003) note, it is an analytical problem for each Bantu language to determine whether the mora or the syllable is the "tone-bearing unit" (TBU). In the Bantu languages which have lost the Proto-Bantu contrast in vowel length, the choice is moot, in most cases. If one assumes, as most authors do, that only open syllables are licensed in Bantu languages, then the syllable and mora define identical TBUs in a language with only short vowels.[3] However, predictably lengthened penult vowels and moraic (or syllabic) nasals in NC sequences potentially provide evidence for a choice. When relevant, the TBU is mentioned in presenting the data in the examples below.

It has been argued since work like Stevick (1969) that High tone is the active tone in many Bantu languages, while Low tone is often best analyzed as a default (or underspecified) tone, predictable from context (see also CHAPTER 45: THE REPRESENTATION OF TONE). For example, in (1b), above, the Low component of the final falling tone is an intonational final Low tone, which, as Downing (1996) shows, predictably occurs at the end of statements in Jita. The asymmetry in the phonological activity of High and Low tones is reflected in the tone marking convention adopted by Bantuists (and used in this chapter) of only marking High tone with an acute accent, while leaving Low tone generally unmarked. As we will see in this chapter, the analysis of Bantu tone patterns is mainly preoccupied with accounting for the surface distribution of High tones.

Another asymmetry found in many Bantu tone systems, noted since at least McCawley (1970, 1978), is that in verbs only the initial syllable of the verb root contrasts for tone, while in nouns, every syllable can bear contrastive tone. Again, data from Jita illustrate:

(2) *Jita lexical tone contrasts in verbs* vs. *nouns* (Downing 1996)

	High-toned stems		*Low-toned stems*	
a.	*Verbs* (infinitive form is shown; *oku-* is the infinitive prefix)			
	oku-lyâ	'to eat'	oku-sya	'to grind'
	oku-βóna	'to see'	oku-βuma	'to hit'
	oku-ííga	'to look for'	oku-iiga	'to pass a test'
b.	*Nouns* (stem follows hyphen; initial morpheme is the class agreement prefix)			
	omu-gási	'woman'	omu-saani	'friend'
	omu-tuungâ	'rich person'	omu-lamusi	'judge'
	li-nanáji	'pineapple'		
	li-darinâ	'tangerine'		

These two asymmetries will be discussed in more detail in §2.

[3] See Downing (2005) for critical discussion of this assumption.

As Clements and Goldsmith (1984a) and Philippson (1998) argue, Meeussen's Rule is one of the important synchronic and diachronic processes lending an accentual character to Bantu tone systems, because it results in culminative prominence of High tone within some large morphological domain.

Another common process resulting in culminativity is tone shift. The Jita data in (4) provide examples of this process. As we can see, an input High tone is systematically realized one syllable to the right of its input sponsor unless the sponsor is in the penult or final syllable. The example in (4e) shows that tone shift crosses word boundaries. (We return to such phrasal tone processes in §4.)

(4) *Jita tone shift*

a.	oku-[βó̲na	'to see'
b.	oku-[βo̲nána	'to see each other'
c.	oku-[βuma	'to hit'
d.	oku-mu̲-[βúma	'to hit him/her (class 1)'
e.	oku-[fwa̲ kúmugera	'to die by the river'
	(oku-[fwá̲ 'to die'; kumugera 'by the river')	

As Kenstowicz (1993), Cassimjee and Kisseberth (1998), and Kisseberth and Odden (2003) show, tone shift is clearly related to the assimilatory process of tone spread. What is puzzling is why the High tone should delink from its sponsor syllable, since other types of feature assimilation do not typically involve such delinking. Kenstowicz (1993) argues that delinking is best understood as an accentual process, leading to High tone culminativity. By eliminating the sequence of High tones on adjacent syllables derived by tone spread, it makes one syllable in the word more prominent than the others.

Most Bantu languages appear to be like Jita. The tone system is not entirely stress-like, since it preserves a contrast between stems which have a High tone and those which are toneless. High tone assignment is not obligatory like stress would be. However, some Bantu languages have taken a step further toward having a stress-like accentual system in that all words have a single High tone, consistently assigned to a limited number of positions in the word. For example, Odden (1988) shows that in Bena, every noun must have a High tone, realized on either the penult or the pre-stem vowel, and most verb forms require a High tone on the penult:

(5) *Bena* (Odden 1988: 236)

a. *Nouns*		b. *Verbs*	
mú-goosi	'man'	kwaamíle	'put to pasture (SUBJ)'
hí-fuva	'chest'	ndi-líma	'I will cultivate (NEAR FUT)'
mu-gúunda	'field'	ndaa-limága	'I used to cultivate'
lu-fwiíli	'hair'	ndaa-limiíge	'I was cultivating'
li-fulúha	'cloud'	ndihaa-limíle	'I cultivated (INTERM PAST)'
		ndaa-limíle	'I cultivated (FAR PAST)'
		hu-limíla	'to cultivate for'

Other languages like Bena are discussed in Odden's (1988, 1999) critical overviews of what he calls "predictable" tone systems in Bantu languages.

edges. These properties have all been argued to motivate the interaction of tone with prosodic structure in work beginning with Goldsmith (1976). Let us briefly review, in roughly chronological order, some of the leading proposals.

In Goldsmith's (1976, 1984a, 1984b) accent theory, Bantu words do not have an underlying High tone *vs.* toneless contrast. Instead, the contrast is formalized as the presence *vs.* absence of accent (*). For example, the Jita High-toned words in (2) would have an accent on the underlyingly High-toned syllable, while Low-toned words would be unaccented, as shown in (8a). Processes like Meeussen's Rule (MR) would be motivated by stress clash, formalized to eliminate sequences of accents (rather than High tones), as shown in (8b; cf. (3b), above). Tone melodies are associated with accents at a later stage of the derivation:

(8) *Accentual analysis* (à la Goldsmith 1976, 1984a, 1984b) *of the Jita data in* (2)–(4)

a. Underlying contrast:

   /oku-βo*na/ 'to see' *vs.* /oku-βuma/ 'to hit'
   accented — unaccented

b. MR as clash resolution:

   /oku-mu*-βo*na/ 'to see him/her' → [oku-mu*-βona]

(Accents (*) are printed above the accented syllables: βo in /oku-βona/; mu and βo in /oku-mu-βona/; mu in [oku-mu-βona].)

The use of accent to formalize High tone culminativity was mostly abandoned by the mid-1980s, after influential work like Hyman and Byarushengo (1984) and Pulleyblank (1986) argued that the lexical High tone *vs.* toneless (or inactive default tone) contrast motivated by the output tone patterns can be straightforwardly represented by associating High tones with particular input syllables and underspecifying Low tone. The OCP-motivated process, Meeussen's Rule, eliminates sequences of identical adjacent High tones without appealing to stress clash. It is an unnecessary complication to have linked accents at one stage of the derivation and linked High tones at another, when an analysis that appeals only to High tones can account for this "accentual" property. (See Hyman 1989, Odden 1999, and Yip 2002 for detailed discussion.)

Accent continued to be appealed to in work like Goldsmith (1987), Kisseberth (1992), and Downing (1996), though, to account for the demarcative output position of High tone in many Bantu languages, illustrated by the Digo data in (6) and (7) above. This alternative accentual approach accounts for the fact that High tones tend to surface near word or morpheme edges – that is, syllables that are often assigned stress in stress languages – by assigning metrical prominence – accent – to these edges. In the unmarked case, High tones are "attracted" to these accented syllables (Goldsmith 1987; de Lacy 2002). In other words, in this theory, accent is redefined as metrical prominence, and accentual languages are defined as those in which (High) tone and metrical prominence interact. These points can be illustrated by an accentual analysis of the Digo data in (7a). Recall that in Digo the rightmost (or only) High tone surfaces on the penultimate mora (i.e. in the stressed syllable), while a second High tone, if any, surfaces on the verb-stem initial mora. This attraction of High tones to particular positions can be accounted for, as shown in (9), by assigning them an accent. (To formalize the generalization that if there is only one High tone, it surfaces on the penultimate mora, one could assign it a higher degree of metrical prominence than the stem-initial mora):

We can find other arguments supporting Cassimjee and Kisseberth's (1998) proposal that High tone realization in Bantu languages is mediated by prosodic tone domains, if we briefly return to Odden's (1999) non-accentual analysis of positional High tone realization in languages like Digo, sketched in (10). The non-accentual analysis was argued to be superior to an accentual one, as it requires no special prosodic representations, appealing to processes like extratonality, tone spread, and delinking of multiple associations which are commonly found in Bantu languages. Under closer examination, though, one notices that only tone spread – i.e. a form of assimilation – is considered an unmarked featural process, common outside of Bantu tone. It would be very surprising, though, to find a vowel harmony process, for example, which excluded the final syllable of a word through "extra-harmonality." It would also be unusual to find a vowel harmony system where only the rightmost vowel of a harmony span surfaced with the harmonizing feature, while it delinked from the other vowels in the span. In contrast, non-finality (extrametricality) is a typical property of prosodic constituents and accentual processes like stress assignment. And, as noted above, Kenstowicz (1993) argues that delinking of High tones from all syllables except the rightmost in a span falls out if the rightmost syllable is a prosodic head. Delinking results in a prosodic prominence asymmetry that is a defining property of prosodic heads.

In short, one must conclude that the jury is still out on whether the prosodic properties, including accentual properties, of Bantu tone systems are best accounted for by appealing to prosodic representations. One can account for many accentual properties of Bantu tone systems without appealing to accent or prosody. However, in doing that one often loses explanatory force, by ignoring the fact that many common restrictions on Bantu tone realization are also common restrictions on headed prosodic constituents but rare for other types of feature assimilation processes.

3 Depressor consonants

Bantu languages are well known for their so-called "depressor consonants": that is, sets of consonants, often voiced, which block High tone spread and interfere with productive processes of High tone realization by lowering the pitch of a following vowel in some way. (An example of depressor interference with High tone shift was noted in presenting the data from Digo in (6) above.) Bantu depressor effects, at first blush, appear to illustrate the typologically common pattern of phonologizing the phonetic lowering effect voiced obstruents have on the pitch of a following vowel (Hyman and Schuh 1974; Hombert 1978; Kingston and Diehl 1994; see also CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). One basic approach to formalizing this lowering effect is to propose that depressor consonants introduce a Low tone into the representation as an automatic result of bearing a [voice] feature: Halle and Stevens (1971); Kisseberth (1984); Harris (1994); Halle (1995); Hyman and Mathangwane (1998); and Bradshaw (1999). The other basic approach is to propose that phonetically grounded implicational constraints (*à la* Archangeli and Pulleyblank 1994) penalize realizing a High tone (and/or optimize realizing a Low tone) on a syllable with a depressor onset: Peng (1992); Cassimjee (1998); and Hansson (2004). In this section, I first present on the depressor syllables violates the no-line-crossing constraint, as it interrupts the multiply associated High tone sequence derived by autosegmental High tone spread.

To resolve this problem, Hyman and Mathangwane (1998) propose that a process of tone "fission" (somewhat simplified here) applies to derive a well-formed autosegmental output:

(21)

However, the tone fission analysis raises the question of why the Low tone associates with the depressor syllable in the first place, when this violates the no-line-crossing constraint invoked to account for why depressors block High tone spread in other contexts. Further, one wonders why fission is commonly triggered by depressor Low tones but has not been motivated for Low tones from other sources or when spreading non-tonal features.

Problems like these have led to the development of an alternative approach to depressor transparency (and depressor blocking), inspired by the use of phonetic grounding constraints to account for transparency in vowel harmony systems. For example, in many West African ATR harmony systems, [i] does not participate in the harmony; it is either transparent or opaque. Archangeli and Pulleyblank (1994) argue that this can be accounted for through constraints which formalize that [+high]/[RTR] is a phonetically antagonistic feature combination. The high, front position of the tongue required for [i] is in conflict with the low, retracted position that defines [RTR]. Transparency of [i], like its harmonic opacity, follows from a grammar which penalizes vowels for realizing this antagonistic combination.

The transparency of depressor syllables for some High tone displacement processes in Bantu languages can be given a similar account. As work like Peng (1992), Cassimjee (1998), and Hansson (2004) argues, tone–segment interactions can be formalized in terms of grounding constraints which define both enhancing and inhibiting interactions between tone and particular laryngeal features. For example, Cassimjee (1998: 53) proposes that the grounding constraint in (22) defines the basic interaction between depressor syllables and High tone in Nguni languages like Xhosa or Ndebele:

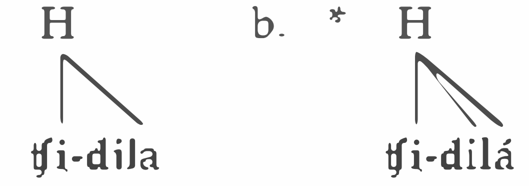
(22) *(Depress, H)

If depressor syllable, then not High tone.

The output of the High tone spread processes relevant to the form in (21) in this analysis is given below, adopting an autosegmental representation for ease of comparison rather than Cassimjee's (1998) tone domain account. The representation in (23b) is ill formed, as it violates the grounding constraint. The output in (23a) satisfies the constraint and is well formed if one adopts Archangeli and Pulleyblank's (1994) proposal that No Gap (or its OT equivalent) – a constraint penalizing representations where potential targets of High tone association are skipped over – is violated in systems with harmonic transparency:

(23) *Grounded depressor transparency*

a. b. *



Notice that the grounding constraint in (22) captures the generalization that High tones are not optimally realized on a syllable with a depressor onset without directly associating a Low tone with the depressed syllable. This means that the association of a Low tone at some stage of the derivation plays no role in this approach in accounting for why depressors are opaque for some tonal processes. For this reason, no representational problem arises if depressors are transparent for other processes in the same language. Instead, in an OT analysis of Bantu High tone realization like Cassimjee (1998), it is the relative ranking of the grounding constraint with other constraints which accounts for this variability.

The grounding constraint approach also has the advantage of linking depressor consonant transparency to the wider phonological problem of how best to account for transparency in harmony assimilation systems. While the representation in (23a), like the one in (21), violates an autosegmental well-formedness constraint, No Gap is routinely violated by grounded accounts of transparency in other assimilatory systems (Archangeli and Pulleyblank 1994). In contrast, the featural fission represented in (21), required by analyses which account for depressor blocking with a linked Low tone, has not been motivated for other phonological processes involving other features.

The analysis of the opacity and transparency of depressor consonants to High tone displacement in Bantu languages, then, raises questions that are of general interest to phonological theories which aim to account for harmonic opacity and transparency. Another important theoretical question raised by the analysis of depressor consonants is which phonetic or phonological consonantal feature grounds the tone lowering effect that accompanies these consonants. Single-source theories of consonant–tone interaction make the strong claim that: "Depressor effects [. . .] always indicate a special relationship between voicing and L[ow] tone" (Bradshaw 1999: 43). As work like Halle and Stevens (1971), Halle (1995), and Harris (1994) argues, the same laryngeal configuration that favors vocal fold vibration (slack vocal cords) also lowers fundamental frequency. And, conversely, work like Bradshaw (1999: 163) claims that Low tone implies vocal fold vibration. These theories implement the strong correlation between depressor effects and phonetic voicing of a preceding consonant by proposing that [voice] and Low tone are reflexes of a single laryngeal feature, with different realizations depending on whether the feature is linked to a consonant or a vowel.

The proposal that [voice] and Low tone are expressions of a single feature faces two problems in accounting for Bantu depressor effects. First, as we have seen in the preceding discussion, uniformly associating a Low tone with a depressor accounts well for depressor opacity (i.e. blocking effects); however, accounting for depressor transparency requires some special theoretical assumptions or devices, like tone fission or No Gap violations. A more serious problem for a single-source approach is that depressor consonants are not all voiced in all Bantu languages. For example, voiceless depressors are found in Nambya and Kalanga, two languages of the Shona group, and also in the Nguni languages. As Downing (2009) shows, these sounds are not only synchronically voiceless, there is no evidence that they were voiced historically. Instead, they have other phonetic properties that plausibly correlate with pitch lowering: frication duration (for the Shona group) and lax vocal fold tension (for the Nguni group). Surveys like Yip (2002) show that, in fact, a number of laryngeal properties besides [voice] condition tone realization and correlate with pitch lowering. Single-source theories of depressor consonants are flawed, then, in linking tonal depression only to [voice].[8]

The grounded constraint approach can get around this problem by defining an implicational relationship between tone realization and a "depressor syllable," with a depressor syllable further defined as one with an onset consonant that has a phonetic property which demonstrably correlates with lowered fundamental frequency: negative VOT, vibrating vocal cords, breathy voice release, long frication duration, lax vocal fold tension (Downing 2009). This list of possible correlates of depressor consonants serves to highlight, though, that the phonetics of consonant–tone interactions remains poorly understood.

4 Phonology–syntax interface

So far, we have discussed (High) tone realization within words, noting only in passing, in presenting the Jita tone shift data in (4), that High tones commonly cross word boundaries in Bantu languages. More interesting is that tonal processes are often sensitive to syntactic information: i.e. High tones cross some word boundaries but not all. For this reason, Bantu tonal phenomena have played an important role in the development of theories of the phonology–syntax interface: see e.g. Hyman and Byarushengo (1984); Selkirk (1986, 2000, forthcoming); Odden (1987, 1990, 1995b); Kanerva (1990); Truckenbrodt (1995, 1999, 2007); and several papers in Inkelas and Zec (1990). A central question for these theories is, as Chen (1990) so neatly puts it, "What must phonology know about syntax?" One finds two leading approaches, which provide two very different answers to this question, in work on Bantu phrasal tonology. In the direct reference approach (Odden 1987, 1990, 1995b), phonology has relatively unrestricted access to information in the morphosyntactic representation. In contrast, in the End-based (indirect reference) approach (Selkirk 1986, 2000; Truckenbrodt 1995, 1999, 2007), phonological phrasing algorithms can only refer to the edges of major syntactic constituents. Another central question is, naturally, whether syntax is all that phonological phrasing algorithms need to know.

In this section, I first present data illustrating a range of syntactic contexts for phrasal tonal processes. Then I briefly review the indirect and direct reference approaches to accounting for these syntactic contexts. As we shall see, a challenge for both of these approaches is that reference to non-syntactic factors like prosodic branching or focus is necessary to account for phrasing in some languages.

From the earliest work on phrasal phonology, including tonology, in Bantu languages – Byarushengo *et al.* (1976) on Haya, and Kisseberth and Abasheikh

[8] As Hyman (forthcoming) notes, a further challenge for single-source theories of consonant–tone interactions is that while the laryngeal properties of consonants are uncontroversially characterizable in terms of distinctive features, solid arguments for a universal set of tonal features are hard to come by.

phrasal constituent as its founding tenet. That is, it proposes that phonology needs to know very little about syntax: only where the edge – right or left – of major constituents like XP, *v*P, and CP are located in the string. The Edge-based approach is considered an indirect reference theory, as phonological processes refer directly to prosodic constituents like Phonological Phrase, and so only indirectly reference the syntactic constituents that define the prosodic ones. The repertoire of prosodic phrasal domain types is also limited in this theory to just those found in the Prosodic Hierarchy (Selkirk 1986, 1995): Phonological Phrase, Intonation Phrase, and Phonological Utterance. In OT, an Edge-based parse of a string into prosodic phrase domains is implemented straightforwardly by means of alignment constraints (Selkirk 1995, 2000; Truckenbrodt 1995, 1999, 2007). For example, ALIGNR(XP, Phonological Phrase) would be relevant for a language like Tsonga, where Phonological Phrase breaks (indicated by parentheses) are found following XPs: the NP subject and the first of two post-verbal objects: $_{NP}$[Subj)] $_{VP}$[V $_{NP}$[Obj)] $_{NP}$[Obj)]]. Recent Edge-based OT analyses of prosodic phrasing in Zulu (Cheng and Downing 2009), in Mwiini, Matuumbi, and Chewa (Truckenbrodt 1995, 1999, 2007), and in Northern Sotho (Zerbian 2006, 2007) demonstrate how effectively this minimalistic approach accounts for many phrasal phenomena in Bantu languages.

The limits of the Edge-based theory are thoughtfully surveyed in Odden (1995b), who argues instead for the direct reference approach to defining syntactic contexts for phrasal processes. Odden points out two general problems with the Edge-based approach. First, the prosodic phrase types are limited to those provided by the Prosodic Hierarchy: Phonological Phrase, Intonation Phrase, Phonological Utterance. For languages with complex phrasal phonology, though, this might not be enough levels. A second problem is that phrasal phonology is often sensitive to more specific morphosyntactic information than just the edges of major syntactic constituents. Tsonga (Kisseberth 1994) illustrates both of the problems, as Odden (1995b) shows. We saw in (26) that the verb's High tone spreads to the penult of the first following object noun only, not to the second object noun of affirmative verbs. However, we find a different pattern with negative verbs: the grammatical High tone of the negative verb is realized through the *final* vowel of the entire verb phrase if both following objects are toneless (27a, 27b). The negative High tone spreads through the *penult* of a postposed subject, though – the same pattern found with affirmative verbs with a final High tone (cf. (c) and (d)). As (27e) shows, the High tone from a verb prefix does *not* spread to a postposed subject:

(27) *Tsonga negative VPs* (Kisseberth 1994: 150, 162, 163; Odden 1995a: 66)[9]

a. a-vá-xav!êlí xí-kóxá nyáámá cf. (26a)
NEG-they-buy old woman meat
'They are not buying meat for the old woman'

b. a-ndzí-nyíkí mú-fáná tí-n-gúúvú cf. (26b)
NEG-I-give boy clothes
'I am not giving the boy clothes'

[9] In this dataset, an exclamation mark indicates that the following High tones are downstepped (i.e. lowered in pitch register) with respect to preceding High tones.

c. vá-ꜜfáámbá váá-nhu 'They are going, the people.'
d. a-má-yángá má-láándza 'They did not go, the servants.'
e. v-á-tíírha) vaa-nhu 'They are working, the people.'

Finally, the alert reader will have noticed that these data illustrate another phrasal rule of Tsonga, namely, phrase penult vowel lengthening. As Odden (1995b) points out, an inspection of the data shows that the domain for this process is also often distinct from that of High tone spread to the penult (HTS). For example, in (26), the domain for HTS is the verb plus first object noun, while the domain for lengthening is the entire sentence; conversely, in (27d), the domain for HTS is the entire sentence (verb plus postposed subject), while both the verb and the postposed subject are in separate domains for penult lengthening. In short, in Tsonga, we have several phrasal processes – HTS to penult from a verbal prefix; HTS to penult from a verb stem; HTS to final syllable from a negative verb; penult lengthening – and each process takes a slightly different domain. Further, some of these processes must refer to specific morphosyntactic information, like negative aspect or verb prefix, as the source of the High tone to determine the phrasal domain. It is hard not to agree with Odden (1995b) that the Edge-based approach does not straightforwardly provide either enough domains or access to enough morphosyntactic information to define the contexts for all of these processes. Other examples discussed in Odden (1995b) reinforce this point.[10]

Let us close this section by showing that syntactic properties of a string are not all that phonology needs to know to define the contexts for phrasal tone domains. As Bickmore's (1990) study of Nyambo shows, another factor that can condition phrasing is branchingness: that is, whether or not a syntactic constituent (in this case, maximal XP) has more than one member. The role of branchingness is motivated by data like those in (28), which illustrate the phrasal process of High tone deletion (HTD): the rightmost High tone of a word is deleted if the following word has a High tone. Examples (a) and (c) show that the entire sentence can be the domain of HTD if none of the XPs contained in it branch. Examples (b) and (d) show that a branching XP (underlined) is not phrased with what follows:

(28) *Nyambo prosodic phrasing* (Bickmore 1990: 14–15)

a. (Ba-kuru bá-ka-júna).
Mature ones helped
'The mature ones helped.'

b. (Aba-kozi bakúru) (bá-ka-júna).
workers mature helped
'The mature workers helped.'

c. (Nejákworech' ábakoz' émbwa.)
he will show workers dog
'He will show the workers the dog.

d. (Nejákworech' ómukama w'ábakózi) (émbwa).
he will show chief of workers dog
'He will show the chief of the workers the dog.'

10 See Selkirk (forthcoming) for a recent re-analysis of the Tsonga phrasal tone domain data discussed here which takes up these and other criticisms of the Edge-based approach.

Focus also often plays an important role in conditioning prosodic phrasing. An especially detailed demonstration of this is found in Kanerva's (1990) study of Chewa. We saw in (25), above, that in Chewa the entire VP (verb plus complements) forms a single tone realization domain under broad focus. As Kanerva shows, narrow focus within the VP interferes with syntactically motivated phrasing: a phrase boundary must follow the focused element, and each subsequent XP constituent of the VP is parsed into a separate prosodic phrase. As a result, the VP is parsed into more, smaller prosodic phrases – indicated with parentheses – under narrow focus:

(29) *Focus and phrasing in Chewa* (Kanerva 1990: 98)

a. anaményá nyumbá ndí mwáála 'He hit the house with a rock.'
   he hit house with rock
b. What did he do? (VP focus)
   (anaményá nyumbá ndí mwáála)
c. What did he hit the house with? (Oblique PP focus)
   (anaményá nyumbá ndí mwáála)
d. What did he hit with the rock? (Object NP focus)
   (anaményá nyuúmba) (ndí mwáála)
e. What did he do to the house with the rock? (V focus)
   (anaméenya) (nyuúmba) (ndí mwáála)

While Bickmore (1990), Kanerva (1990), and Truckenbrodt (1995, 1999, 2007) show that the effect of branchingness and focus on phrasing can be accommodated in the Edge-based approach, it is also clear that neither that approach nor the direct reference approach predicts that such non-syntactic factors should interact with syntactic ones in defining the contexts for phrasal processes.

5 Conclusion

As we have seen, Bantu tone has played a role in several areas of phonological theory, notably in the areas of phonological representation and the role of prosodic factors in conditioning featural realization. The accentual properties of many Bantu tone systems have raised the question of whether some form of prosodic representation – accent and/or prosodic constituents – mediates tone realization, or whether the accentual properties are best considered the result of non-prosodic processes which conspire to produce an accentual output. Depressor tones raise questions of how to formalize the interaction of tone with other laryngeal features. The opacity and transparency of depressor consonants to High tone displacement processes raise the same representational problems as opacity and, especially, transparency in harmony systems. The phrasal realization of High tone raises again the question of whether prosodic constituency best defines the domains for High tone realization, or whether independently motivated syntactic constituents and principles provide all that phonology needs to know about syntax. The Bantu tone data illustrating these issues are complex enough that phonological theories will surely be wrestling with these questions for some time to come.

ACKNOWLEDGMENTS

I would like to thank the editors and two anonymous reviewers for useful comments. The usual disclaimers apply.

REFERENCES

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Bickmore, Lee. 1990. Branching nodes and prosodic categories. In Inkelas & Zec (1990), 1–17.

Bickmore, Lee. 1995. Tone and stress in Lamba. *Phonology* 12. 307–341.

Bradshaw, Mary M. 1999. A crosslinguistic study of consonant–tone interaction. Ph.D. dissertation, Ohio State University.

Byarushengo, Ernest Rugwa, Larry M. Hyman & Sarah Tenenbaum. 1976. Tone, accent, and assertion in Haya. In Larry M. Hyman (ed.) *Studies in Bantu tonology*, 185–205. Los Angeles: Department of Linguistics, University of Southern California.

Cassimjee, Farida. 1998. *IsiXhosa tonology: An Optimal Domains Theory analysis*. Munich: Lincom Europa.

Cassimjee, Farida & Charles W. Kisseberth. 1992. The tonology of depressor consonants: Evidence from Mijikenda and Nguni. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 18(2). 26–40.

Cassimjee, Farida & Charles W. Kisseberth. 1998. Optimality Domains Theory and Bantu tonology: A case study from Isixhosa and Shingazidja. In Hyman & Kisseberth (1998), 33–132.

Cassimjee, Farida & Charles W. Kisseberth. 2001. Zulu tonology and its relationship to other Nguni languages. In Shigeki Kaji (ed.) *Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese accentology and other topics*, 327–359. Tokyo: ILCAA.

Chen, Matthew Y. 1990. What must phonology know about syntax? In Inkelas & Zec (1990), 19–46.

Cheng, Lisa L. & Laura J. Downing. 2009. Where's the topic in Zulu? *The Linguistic Review* 26. 207–238.

Clements, G. N. & John A. Goldsmith. 1984a. Introduction. In Clements & Goldsmith (1984b), 1–17.

Clements, G. N. & John A. Goldsmith (eds.) 1984b. *Autosegmental studies in Bantu tone*. Dordrecht: Foris.

de Lacy, Paul. 2002. The interaction of tone and stress in Optimality Theory. *Phonology* 19. 1–32.

Donnelly, Simon. 2007. Aspects of tone and voice in Phuthi. Ph.D. dissertation (revised), University of Illinois at Urbana-Champaign.

Downing, Laura J. 1996. *The tonal phonology of Jita*. Munich: Lincom Europa.

Downing, Laura J. 2005. On the ambiguous segmental status of nasals in homorganic NC sequences. In Marc van Oostendorp & Jeroen van de Weijer (eds.) *The internal organization of phonological segments*, 183–216. Berlin & New York: Mouton de Gruyter.

Downing, Laura J. 2009. On pitch lowering not linked to voicing: Nguni and Shona group depressors (with an appendix by Thilo C. Schadeberg). *Language Sciences* 31. 179–198.

Fromkin, Victoria A. (ed.) 1978. *Tone: A linguistic survey*. New York: Academic Press.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT. Published 1979, New York: Garland.

Goldsmith, John A. 1984a. Meeussen's Rule. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*, 245–259. Cambridge, MA: MIT Press.

McCawley, James D. 1978. What is a tone language? In Fromkin (1978), 113–131.
Myers, Scott. 1987. Tone and the structure of words in Shona. Ph.D. dissertation, University of Massachusetts, Amherst.
Myers, Scott. 1997. OCP effects in Optimality Theory. *Natural Language and Linguistic Theory* 15. 847–892.
Myers, Scott & Troi Carleton. 1996. Tonal transfer in Chichewa. *Phonology* 13. 39–72.
Nurse, Derek & Gérard Philippson. 2003a. Introduction. In Nurse & Philippson (2003b), 1–12.
Nurse, Derek & Gérard Philippson (eds.) 2003b. *The Bantu languages*. London & New York: Routledge.
Odden, David. 1986. On the role of the Obligatory Contour Principle in phonological theory. *Language* 62. 353–383.
Odden, David. 1987. Kimatuumbi phrasal phonology. *Phonology Yearbook* 4. 13–36.
Odden, David. 1988. Predictable tone systems in Bantu. In Harry van der Hulst & Norval Smith (eds.) *Autosegmental studies on pitch accent*, 225–251. Dordrecht: Foris.
Odden, David. 1990. Syntax, lexical rules and postlexical rules in Kimatuumbi. In Inkelas & Zec (1990), 259–277.
Odden, David. 1995a. Tone: African languages. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 444–475. Cambridge, MA & Oxford: Blackwell.
Odden, David. 1995b. Phonology at the phrasal level in Bantu. In Francis Katamba (ed.) *Bantu phonology and morphology*, 40–68. Munich: Lincom Europa.
Odden, David. 1999. Typological issues in tone and stress in Bantu. In Shigeki Kaji (ed.) *Cross-linguistic studies of tonal phenomena: Tonogenesis, typology, and related topics*, 187–215. Tokyo: ILCAA.
Odden, David. 2007. The unnatural tonology of Zina Kotoko. In Tomas Riad & Carlos Gussenhoven (eds.) *Tones and tunes*, vol. 1: *Typological studies in word and sentence prosody*, 63–89. Berlin & New York: Mouton de Gruyter.
Peng, Long. 1992. A unified theory of tone-voice. Ph.D. dissertation, University of Arizona.
Philippson, Gérard. 1998. Tone reduction *vs.* metrical attraction in the evolution of Eastern Bantu tone systems. In Hyman & Kisseberth (1998), 315–329.
Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology*. Dordrecht: Reidel.
Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371–405.
Selkirk, Elisabeth. 1995. The prosodic structure of function words. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 439–469. Amherst: GLSA.
Selkirk, Elisabeth. 2000. The interaction of constraints on prosodic phrasing. In Merle Horne (ed.) *Prosody: Theory and experiment*, 231–261. Dordrecht: Kluwer.
Selkirk, Elisabeth. Forthcoming. The syntax–phonology interface. In Goldsmith *et al.* (forthcoming).
Stevick, Earl W. 1969. Tone in Bantu. *International Journal of American Linguistics* 35. 330–341.
Truckenbrodt, Hubert. 1995. Phonological phrases: Their relation to syntax, focus, and prominence. Ph.D. dissertation, MIT.
Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30. 219–255.
Truckenbrodt, Hubert. 2007. The syntax–phonology interface. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 435–456. Cambridge: Cambridge University Press.
Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.
Zerbian, Sabine. 2006. Expression of information structure in the Bantu language Northern Sotho. Ph.D. dissertation, Humboldt-Universität, Berlin. Published as *ZAS Papers in Linguistics* 45.
Zerbian, Sabine. 2007. Phonological phrasing in Northern Sotho (Bantu). *The Linguistic Review* 24. 233–262.

115 Chinese Syllable Structure

San Duanmu

1 Introduction

The title of this chapter may seem both too broad and too narrow. On the one hand, Chinese dialects are probably as diverse as Romance languages. Should one not treat them separately? However, I shall show that there are considerable similarities among them and that a single treatment is useful. On the other hand, the study of syllables is closely related to the studies of phonemes, stress, and tone. Therefore, the discussion cannot be limited to syllable structure alone, but will cover some of its relations to other areas as well.

The Chinese writing system is not alphabetic, but Chinese scholars have studied syllables for a long time. For example, Sun Yan of the Three Kingdoms period (about 200–280) invented a method, known as *Fanqie* "reverse cut," to indicate the pronunciation of written graphs (characters), each of which represents a monosyllabic word. The method uses two familiar graphs, where the first has the same onset as the target graph and the second has the same rime as the target. Similarly, there were riming books that grouped Chinese graphs into different sets, based on their initial consonants and whether they rime with each other.

Despite the traditional scholarship, and the fact that Chinese syllables seem less complicated than those in English, disagreements exist over most major issues. For example, what is the maximal size of the Chinese syllable? Does every syllable have a vowel? How does the analysis of syllables interact with the analysis of phonemes? What is the interaction between syllable weight, stress, and tone? What are major phonotactic restrictions that rule out non-occurring syllables? What is the cause of the massive loss of syllable types in Chinese? In this chapter I address such questions. Following the editors' guidelines, I shall focus on major empirical facts, rather than arguing for my own analysis.

However, the distinction between facts and theories is not always clear. For example, the word 'outside' in Standard Chinese, which sounds the same as the English word *why*, has been transcribed as [waj] (Hartman 1944), [uai] (Cheng 1966), and [ua^{i}] (You *et al.* 1980). For Hartman (1944) the syllable has two glide phonemes and a vowel phoneme, for Cheng (1966) it has three vowel phonemes, and for You *et al.* (1980) it has two vowel phonemes. Accordingly, the syllable analyses are GVG (or CVC), VVV, and VV respectively. One could see that even basic terms like glides, diphthongs, and triphthongs are not always transparent or meaningful. Such problems exist in English and other languages, too. Therefore, our discussion is, inevitably, also of general theoretical interest.

2 The maximal syllable structure

A maximal Chinese syllable, regardless of the dialect, is often thought to contain four positions, or CGVX, where C is a consonant, G a glide, V a vowel, and X either a consonant or the second part of a long vowel or diphthong. Syllables with a syllabic consonant will be discussed later, as will syllables whose rime seems to be longer than VX. In CGVX, a diphthong takes two positions and cannot be followed by a consonant coda. Some examples from Standard Chinese are shown in (1). For simplicity, I enclose phonetic transcription in square brackets, regardless of the level of analysis. As noted above, some analyses would transcribe a pre-nuclear glide as a high vowel (e.g. Cheng 1966; T. Lin and Wang 1992).

(1) *The* CGVX *analysis of the maximal syllable in Standard Chinese*

[kʰwai]	'fast'	[kwaŋ]	'light'
[kwaː]	'melon'	[tswan]	'diamond'
[tʰjan]	'day'		

Different Chinese dialects can use different phonemes to fill the four slots of a syllable. Three dialects, Standard Chinese, Cantonese, and Shanghai, are shown in (2).

(2) *Dialect variation: Phonemes which can fill each position in* CGVX

	C	G	V	X
Standard Chinese	most Cs	[j w ɥ]	any V	[i u n ŋ ɚ]
Cantonese	any C	[w]	any V	[i u n m ŋ p t k]
Shanghai	any C	[j w ɥ]	any V	([ʔ ŋ])

In all dialects, the C position can be filled by almost any consonant, with occasional exceptions; for example, similarly to English, [ŋ] is not used in the C position in Standard Chinese. The G position can be filled by one of three glides in Standard Chinese and Shanghai. In Cantonese, there are two glides [w j] which can occur without C, but when the C is filled only [w] can be used. The V position can be filled by any V (and sometimes by a syllabic consonant, such as [n] 'fish' in Shanghai, to be discussed later). For the X position, Cantonese is among the most conservative dialects, which has kept a full set of nasals and a full set of unreleased stops. In contrast, Shanghai is among the most advanced dialects with regard to the X position, which only allows a glottal stop or a nasal; in addition, these two sounds often combine with the preceding vowel to form a single sound, i.e. a glottalized V or a nasalized V. Therefore, all syllables in Shanghai behave like open syllables. I shall return to this issue below.

There is a fair amount of disagreement on the analysis of CGVX, especially with regard to the affiliation of G. Three proposals are shown in (3).

Standard Chinese ends in [m], but in casual speech such syllables are found, for example [wom], shown in (13).

(13) *New syllable created by syllable merger*

wo mən → wom 'we'
I PL

Similarly, devoicing of non-low vowels often happens in syllables that have an aspirated onset (including voiceless fricatives) and a low tone. Some examples are shown in (14). The transcription is based on Duanmu (2007). HL, H, LH, and L are four lexical tones in Standard Chinese. When a sound is devoiced, the tone cannot be heard, which is indicated by Ø.

(14) *Devoicing of non-low vowel with a low tone in Standard Chinese*

	L-LH		Ø-HL	
[ɤ] → [x]	kʰɤ-nəŋ	→	kʰx-nəŋ	'possible'
	HL-L		HL-Ø	
[i] → [ɕ]	ji-tɕʰi	→	ji-tɕʰɕ	'together'
	H-L		H-Ø	
[y] → [ɕʷ]	ʈʂəŋ-tɕʰʷy	→	ʈʂəŋ-tɕʰʷɕʷ	'strive for'
	L-H		Ø-H	
[y] → [ɕʷ]	ɕʷy-tʷo	→	ɕʷɕʷ-tʷo	'many'
	L-HL		Ø-HL	
[u] → [xʷ]	ʂʷu-tɕa	→	ʂʷxʷ-tɕa	'summer vacation'
	H-L		H-Ø	
[u] → [xʷ]	ɕin-kʰʷu	→	ɕin-kʰʷxʷ	'working hard'
	L-HL		Ø-HL	
[u] → [xʷ]	tʰʷu-tʲi	→	tʰʷxʷ-tʲi	'land'

Devoicing can affect syllables in any position (initial, medial, or final). Devoiced [i ɤ u y] sound like [ɕ x xʷ ɕʷ], respectively. A reviewer suggests that [ʷ] would imply [+back] for [ɕʷ] and recommends [ɕᶣ] instead. There are two reasons not to follow the suggestion. First, while [w] is indeed an IPA symbol for labial-velar, [ʷ] is a diacritic for "labialized," without implying velar or [+back]. Second, [ɥ] has two articulatory components, palatal and labial. Since the palatal component is already present in [ɕ], [ɕᶣ] has internal redundancy, while [ɕʷ] does not.

Devoiced syllables have durations similar to the originals (although rime length is not indicated in the above transcription), and therefore they still sound like separate syllables. If one assumes that every syllable must have a vowel, one would propose devoiced vowels [i̥ ɤ̥ u̥ ẙ]. Syllabic consonants can be devoiced, too. Some examples are shown in (15).

(15) *Devoicing of syllabic consonants in Standard Chinese*

	HL-L		HL-Ø	
[ʋ] → [f]	təu-fʋ	→	tou-ff	'tofu'
	HL-L		HL-Ø	
[z] → [s]	ʂaŋ-tsʰz	→	ʂaŋ-tsʰs	'last time'

Duanmu (1990, 2007) proposes an analysis that maintains both a small syllable size and a small phoneme inventory. The maximal syllable is CVX, similar to that in Ao (1992, 1993). However, unlike Ao, who assumes that C is filled by a single phoneme, Duanmu assumes that C can be filled by two phonemes, a consonant and a glide, which will merge into a complex sound. In Duanmu's analysis of Standard Chinese, the phoneme inventory is slightly smaller than that of traditional analyses, and about half the size of Ao (1992).

Hartman (1944) proposes to minimize the phoneme inventory even further. For example, while Ao (1992) assumes [k k^h k^w k^{hw}] to be four phonemes in Standard Chinese, Hartman assumes the representation [k kh kw khw], in which there is only one velar stop [k], and [kh khw kw] are clusters of two or three phonemes each, where [w] and [h] are independently needed. The cost, of course, is a more complicated syllable structure, which has three positions before the vowel, and it vastly overpredicts the possible syllables of Chinese.

In (17) I compare the analysis of some syllable-initial units in Standard Chinese by Hartman (1944), Cheng (1966), Ao (1992), and Duanmu (2007), where "onset" refers to the maximal size of the given units and "stops" refers to the number of velar stop phonemes.

(17) *Analyses of four syllable-initial units in Standard Chinese*

Author	Underlying	Surface	Onset	Stops
Hartman	[k kh kw khw]	[k kh kw khw]	CCC	1
Cheng	[k k^h ku k^hu]	[k k^h kw k^hw]	CC	2
Ao	[k k^h k^w k^{hw}]	[k k^h k^w k^{hw}]	C	4
Duanmu	[k k^h ku k^hu]	[k k^h k^w k^{hw}]	C	2

Hartman (1944) represents an extreme position in which phonemic economy overrides syllable structure. At the opposite extreme, Ladefoged (2001: 170) believes that phonemes have no place at all. Instead, he suggests that syllables are the basic units of analysis and that "consonants and vowels are largely figments of our good scientific imaginations."

8 Missing syllables

I use the term "missing syllables" to refer to those that do not occur in the language, although their structure fits the maximal size. In Chinese, the majority of conceivable syllables are missing. Consider an example from Nantong Chinese, which allows CVC syllables. According to Ao (1993), Nantong has 15 vowels and 38 consonants. Given that the initial or final C can be absent, there are 39 choices for the initial C and 39 for the final C, and the total number of possible syllables is 39 × 15 × 39 = 22,815, of which only about 400 occur in Nantong, or barely 2 percent. Ao's illustration is perhaps intended to be dramatic. Given the fact that the final C in Nantong can only be [k] or [ŋ], a more reasonable estimate is 39 × 15 × 3 = 1,755 possible syllables, of which 23 percent occur. Still, most conceivable syllables are missing, and an explanation is needed.

Let us consider Standard Chinese in some detail. The phonemes of Standard Chinese are shown in (18) and (19), based on Duanmu (2007).

18 × 3 = 54 CG combinations. However, only 29 (54 percent) are found. The restrictions turn out to be quite systematic: labial Cs do not combine with [w ɥ], and velar and retroflex Cs do not combine with [j ɥ]. In traditional feature terms, the former shows a restriction against two labials, and the latter shows a restriction against [+back, −back] or [−palatal, +palatal]. The details are shown in (22). Given the restrictions, we expect 32 CG forms, of which just three are missing.

(22) *Expected* CG *forms in Standard Chinese (missing numbers in parentheses)*

		j	w	ɥ
Labials (3)	[p p^h f m]	4 (−1)	–	–
Dentals (7)	[t t^h ts ts^h s n l]	7	7	7 (−2)
Velars (3)	[k k^h x]	–	3	–
Retroflexes (4)	[tʂ $tʂ^h$ ʂ ʐ]	–	4	–
Expected	32			
Occurring	29			
Missing	[fj tɥ t^hɥ]			

Next, let us consider GVX combinations. In Standard Chinese there are 100 possible GVX forms, calculated in (23). I have ignored tonal contrasts. In addition, I have omitted the vowel [ɚ] and syllabic consonants.

(23)

Position	*Choices*	*Notes*
G	4	One of [j u ɥ], or no G
V	5	One of five vowels
X	5	One of [i u n ŋ], or no X
Total	100	

The 100 possible GVX forms are shown in (24). The first column indicates choices for X, the top row indicates choices for G, and 0 indicates lack of G or X. High vowels are written as glides before the nuclear vowel.

(24)

	0-	j-	w-	ɥ-		
-0	i	+	(+)	–	–	ji = i
	u	+	–	(+)	–	wu = u
	y	+	–	–	(+)	ɥy = y
	ə	+	+	+	+	
	a	+	+	+	–	
-n	in	+	(+)	–	–	jin = in
	un	–	–	–	–	
	yn	+	–	–	(+)	ɥyn = yn
	ən	+	–	+	–	
	an	+	+	+	+	
-ŋ	iŋ	–	–	–	–	
	uŋ	+	+	(+)	–	wuŋ = uŋ
	yŋ	–	–	–	–	
	əŋ	+	+	+	–	
	aŋ	+	+	+	–	

(25) *Stability of citation tones in Standard Chinese*

Surface	H	H	H	LH	HL	H	HL	LH
Citation	H	H	H	LH	HL	H	HL	LH
	san	pei	san	pʰan	sz	pei	sz	pʰan
	'three cups'		'three plates'		'four cups'		'four plates'	

(26) *Instability of citation tones in Shanghai Chinese*

Surface	H	L	H	L	L	H	L	H
Citation	HL	HL	HL	LH	LH	HL	LH	LH
	se	pe	se	pø	sz	pe	sz	pø
	'three cups'		'three plates'		'four cups'		'four plates'	

It can be seen that while citation tones are stable in Standard Chinese, they all split in the Shanghai examples. A longer example is shown in (27), in fairly broad transcription, where | indicates a boundary between tonal domains in Shanghai.

(27) *Tonal patterns in Shanghai and Standard Chinese*

Shanghai	ku-poʔ	lu		laʔ-laʔ		tʰi-se	lu		pã-pi
Citation	LH-LH	LH	|	LH-LH	|	HL-HL	LH	|	LH-HL
Surface	L-H	0	|	L-H	|	H-L	0	|	L-H
Standard	ku-pei	lu		zai		tʰian-şan	lu		pʰaŋ-pʲan
Citation	L-L	HL	|	HL	|	H-H	HL	|	LH-H
Surface	LH-L	HL	|	HL	|	H-H	HL	|	LH-H
	Gubei	road		be-at		Tianshan	road		vicinity

'Gubei Road is in the vicinity of Tianshan Road.'

We see again the stability of citation tones in Standard Chinese (except for one rule, which changes L to LH before L, as seen on the first syllable). In contrast, citation tones are lost in Shanghai unless they occur in the initial position of a domain; in addition, each surviving citation tone is split between the first two syllables of a domain.

There are two approaches to the difference between the dialects. The first is typological. For example, Yue-Hashimoto (1987) suggests that Shanghai has left-dominant tonal domains but Standard Chinese does not. Similarly, Chen (2000b) suggests that tonal domains in Shanghai are determined by left-headed stress, whereas those in Standard Chinese are not. Moreover, Yip (1989) proposes that there are two kinds of tones: those in Standard Chinese are units that cannot be split, and those in Shanghai are clusters that can. However, the typological approach in effect restates the difference and offers no explanation why Shanghai behaves one way and Standard Chinese behaves another way.

In the second approach, proposed by Duanmu (1990, 1999), the difference in tonal behavior is related to an independent difference in rimes: dialects that are like Shanghai in tonal behavior (with unstable citation tones) have no diphthongs or true codas, while dialects that are like Standard Chinese in tonal behavior (with stable citation tones) have diphthongs and/or true codas. In other words, Shanghai has only "simple rimes" (V or C) while Standard Chinese has many "complex rimes" (VC or VG). Consider the data in (28) and (29).

(28) *Rime types in the basic vocabulary of 2,500 morphemes in Standard Chinese*

VC or VG	1,519	61%	e.g. [man] 'slow', [mai] 'sell'
V or C	981	39%	e.g. [ma] 'scold', [sz] 'four'
Total	2,500	100%	

(29) *Rimes in Shanghai Chinese* (Duanmu 2008)

[m n ɚ z u ø ɔ y i o ɤ e a ỹ ĩ ø̃ ɤ̃ ã iˀ aˀ oˀ]

The nasal vowels in Shanghai can be represented as [Vŋ] underlyingly, and the glottalized vowels can be represented as [Vʔ] underlyingly. Both [Vŋ] and [Vʔ] can merge into a single sound without loss of underlying segmental features.

Given the difference in rime structure between Shanghai and Standard Chinese, it is possible to explain their difference in tonal behavior. First, most rimes in Standard Chinese are inherently heavy. In contrast, syllables in Shanghai have no inherent weight: they can be heavy or light, depending on the prosodic environment. They are long when spoken in isolation or in a stressed position, such as the first syllable of a disyllabic word or compound, which is a trochaic foot; otherwise the syllables are short. Phonetic studies confirm the predictions (Zhu 1995).

The relation between syllable structure and tone is mediated by stress: heavy syllables have stress and stressed syllables can carry tone (see CHAPTER 42: PITCH ACCENT SYSTEMS). The principles are stated in (30) and (31).

(30) *Weight–Stress Principle*

a. Stressed syllables are heavy (or light syllables are unstressed).
b. Unstressed syllables are light (or heavy syllables are stressed).

(31) *Tone–Stress Principle*

a. Stressed syllables can be accompanied by a lexical tone (pitch accent).
b. Unstressed syllables are not accompanied by a lexical tone (pitch accent).

The Weight–Stress Principle has been proposed in various forms in the literature (e.g. Prokosch 1939; Kager 1989; Prince 1990; Hammond 1999; CHAPTER 57: QUANTITY-SENSITIVITY). Kager (1999) refers to (30a) as the Stress-to-Weight principle, and Prince (1990) refers to (30b) as the Weight-to-Stress principle. The Weight–Stress Principle is a bidirectional requirement on the relation between weight and stress, regardless of which comes first.

The Tone–Stress Principle has also been proposed in various forms in the literature (e.g. Liberman 1975; Clements and Ford 1979; Pierrehumbert 1980; Goldsmith 1981). In Chinese, the Tone–Stress Principle is evidenced by the fact that unstressed syllables lose their lexical tones. In English, it is evidenced by the fact that only stressed syllables are assigned a pitch accent.

If a language has many complex rimes, then many syllables will remain heavy and stressed. And because they are stressed, they will keep their lexical tones. The chance of tone split will be low, because most syllables have their own tones and cannot take those from others. In contrast, if a language has no complex rime, then many syllables can become light and unstressed (unless they occur in prosodically strong positions, which are initial positions in trochaic feet). And because they are unstressed, they will lose their lexical tones. The chances for tone split will be high then, because many syllables are toneless and can take a piece from another syllable. A detailed analysis is offered in Duanmu (2008: ch. 7).

Leben (1973) observes a similar pattern in the African language Mende. For example, in a compound made of two monosyllabic words, the second word will lose its underlying tone, and the tone of the first word will spread over both syllables. The Mende case would be expected if it lacks CVX syllables and if its compounds form trochaic feet.

10 Suffixation and rime changes

Many Chinese dialects have suffixes that merge with the preceding syllable. The most common case is the diminutive suffix. In some dialects the rime change is quite simple, whereas in others it can be quite complicated. Let us begin with Chengdu, shown in (32). The process is quite straightforward: in the diminutive form, the rime is replaced by [ɚː].

(32) *Diminutive suffix in Chengdu*

Word	*Diminutive*	
[kən]	[kɚː]	'root'
[kau]	[ka˞ː]	'cake'
[kou]	[kɚː]	'dog'
[pʲan]	[pʲa˞ː]	'side'
[kʷan]	[kʷɚː]	'hall'
[ja]	[jɚː]	'bud'
[wan]	[wɚː]	'bowl'
[yan]	[ya˞ː]	'yard'

Next we consider diminutive forms in Standard Chinese, which are more complicated. Some examples are shown in (33).

(33) *Diminutive forms in Standard Chinese*

Word	*Diminutive*	
[pʰai]	[pʰaɚ]	'plaque'
[pʰiː]	[pʰʲɚː]	'skin'
[pʰan]	[pʰaɚ]	'dish'
[jaŋ]	[jãɚ̃]	'lamb'
[nʲau]	[nʲauᵊ˞]	'bird'
[tʰuː]	[tʰuᵊ˞ː]	'rabbit'

In the first three cases the suffix replaces [n] and [i], although in the third case [i] survives as a glide on the onset. In the fourth case the suffix replaces the coda, but part of the coda is preserved as nasalization on the rime. In the last two cases the suffix is realized as a retroflex color on the original vowel. A number of

(37) *Truncation of polysyllabic morpheme to monosyllabic ones*

Non-truncated	*Truncated*	
bian-fu chao	*bian chao*	'bat nest'
蝙蝠巢	蝙巢	
jia-na-da yuan	*jia yuan*	'Canadian dollar'
加拿大元	加元	
jia-li-fu-ni-ya zhou	*jia zhou*	'California State'
加利福尼亞州	加州	

The compound *bian chao* 'bat nest' was made up by me, but seems quite acceptable. The compounds *jia zhou* 'California State' and *jia yuan* 'Canadian dollar' are real ones.

Overall, it is fair to say that every syllable is a morpheme (or a perceived morpheme) in Chinese. In addition, since most morphemes are also words in Chinese, it is fair to say that most Chinese syllables are words, or that most Chinese words are monosyllabic.

The close relation between syllables and words in Chinese raises an interesting question. If we find a generalization for the Chinese syllable, how do we know whether it is indeed a generalization for syllables and not a generalization for words? For example, should we say that every syllable in Chinese has a tone, or should we say that every word in Chinese has a tone?

12 Statistical data on syllables in Standard Chinese

Compared to English, Standard Chinese has a fairly small inventory of syllables. Consider the data in (38) and (39). A simplex word is one that contains one morpheme.

(38) *Monosyllabic simplex words in English in the CELEX lexicon* (Baayen *et al.* 1995)

All simplex words	7,401
Monosyllabic simplex words	3,834
Different pronunciations	3,219

(39) *Syllables in Standard Chinese*

Vocabulary type	*All*	*Common*
Characters	12,041	2,500
Syllables (with tones)	1,334	1,001
Syllables (without tones)	413	386

English has over 3,000 monosyllables in simplex words. In contrast, Standard Chinese has about 1,000 syllables including tones, or 400 excluding tones.

Although Standard Chinese has just 1,334 syllables (including tones), only about 1,000 are commonly used. Therefore, up to 300 syllables are unfamiliar to the average speaker. This is confirmed by the study of Myers and Tsay (2005), who found that acceptability judgments on possible syllables in Standard Chinese are gradient and influenced by lexical frequency.

Ao, Benjamin. 1993. Phonetics and phonology of Nantong Chinese. Ph.D. dissertation, Ohio State University.

Baayen, R. Harald, Richard Piepenbrock & Hedderik van Rijn. 1995. *The CELEX lexical data base* [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Bao, Zhiming, Jianguo Shi & Debao Xu. 1997. *Shengcheng yinxixue lilun jiqi yingyong* [*Generative phonology: Theory and usage*]. Beijing: Zhongguo Shehui Kexue Chubanshe.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45–86.

Borowsky, Toni. 1989. Structure preservation and the syllable coda in English. *Natural Language and Linguistic Theory* 7. 145–166.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Chao, Yuen-Ren. 1933. Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology, Academia Sinica* 4. 121–134.

Chao, Yuen-Ren. 1934. The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica* 4. 363–397.

Chao, Yuen-Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chen, Matthew Y. 1992. The chameleon r-suffix in Yanggu. *Journal of East Asian Linguistics* 1. 197–213.

Chen, Matthew Y. 2000a. From tone to accent. *Cahiers de Linguistique – Asie Orientale* 29. 179–203.

Chen, Matthew Y. 2000b. *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.

Cheng, Robert L. 1966. Mandarin phonological structure. *Journal of Linguistics* 2. 135–158.

Cheung, Kwan-Hin. 1986. The phonology of present-day Cantonese. Ph.D. dissertation, University of London.

Clements, G. N. & Kevin C. Ford. 1979. Kikuyu tone shift and its synchronic consequences. *Linguistic Inquiry* 10. 179–210.

Coetzee, Andries W. 2006. Variation as accessing "non-optimal" candidates. *Phonology* 23. 337–385.

Coleman, John. 2001. The phonetics and phonology of Tashlhiyt Berber syllabic consonants. *Transactions of the Philological Society* 99. 29–64.

Dong, Shaoke. 1985. Yanggu fangyan de er hua [Er affixation in the Yanggu dialect]. *Zhongguo Yuwen* 1985. 273–276.

Dong, Shaowen. 1958. *Yuyin changtan* [*Introduction to phonetics*]. Beijing: Wenhua Jiaoyu Chubanshe.

Duanmu, San. 1990. A formal study of syllable, tone, stress and domain in Chinese languages. Ph.D. dissertation, MIT.

Duanmu, San. 1999. Metrical structure and tone: Evidence from Mandarin and Shanghai. *Journal of East Asian Linguistics* 8. 1–38.

Duanmu, San. 2007. *The phonology of Standard Chinese*. 2nd edn. Oxford: Oxford University Press.

Duanmu, San. 2008. *Syllable structure: The limits of variation*. Oxford: Oxford University Press.

Feng, Aizhen. 1996. Introduction to *Dictionary of Fuzhou dialect*. *Fangyan* 1996(2). 81–99.

Fidelholtz, James L. 1975. Word frequency and vowel reduction in English. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23. 200–213.

Frisch, Stefan A., Nathan R. Large & David B. Pisoni. 2000. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42. 481–496.

Goldsmith, John A. 1981. English as a tone language. In Didier L. Goyvaerts (ed.) *Phonology in the 1980's*, 287–308. Ghent: E. Story-Scientia.

Hall, T. A. 2001. The distribution of superheavy syllables in Modern English. *Folia Linguistica* 35. 399–442.

Hall, T. A. 2002. The distribution of superheavy syllables in Standard German. *The Linguistic Review* 19. 377–420.

116 Sentential Prominence in English

Carlos Gussenhoven

1 Introduction

In the linguistic expressions of many languages, words vary in prominence. In the English compound (1a), there is some sense in which *tax* has more prominence than *relief*, and in the sentence (1b) the same can be said about the subject NP *My bike* as compared to the verbal phrase *has been stolen*. Example (1c) illustrates how the perception of prominence can be more differentiated than this: *English* appears to be more prominent than *teacher*, which in turn may seem more prominent than *Old*.

(1) a. tax relief
  b. My bike has been stolen!
  c. an Old English teacher

The above observations touch on two aspects of the phonological structure of English which have been the topic of intense debate over the past decades. The first concerns the phonological nature of this prominence: how is it represented in the structure? There have been widely different answers to this question. The main division is between views that see prominence as a single dimension and views that separate word stress from intonational pitch accents. A number of representations that fall in the first class of views are discussed in §2. §3 introduces the second class and explains the difference between word stress and accentuation. §4 then returns to some of the examples discussed in §2, offering an account for them first by assuming that the sentence prominences are pitch accents, and second by identifying the accent assignment or deletion rules that are held responsible for the distribution of the pitch accents in each case. In doing this, §4 also takes a position on the second main controversy: why are the various levels of prominence to be found where they are? What determines that in (1a) through (1c) *tax*, *bike*, and *English* have the most prominence within their structures and what determines that *teacher* may seem more prominent than *Old* in (1c)? Finally, §5 considers the question of whether there are residual differences in prominence that are not covered by the description in §3 and §4. It concludes by suggesting that such differences arise during phonetic implementation rather than being due to differences in representation.

2 The phonological representation of prominence

A useful simplification of the changing perspective on the nature of sentential "stress" is to distinguish two views. The first view, to be referred to as the Infinite Stress View (ISV), takes the impression of multiple degrees of prominence as a starting point and translates these into a representation of gradient stress. This is the older of the two views, and was the basis of an early treatment of stress above the level of the word in generative phonology (Chomsky and Halle 1968). The other view, the Pitch Accent View (PAV), takes sentential prominence to be due to intonational tones that are associated with specific syllables. This view was first expressed by Bolinger (1958), who introduced the term "pitch accents" for these tones (see CHAPTER 32: THE REPRESENTATION OF INTONATION; CHAPTER 50: TONAL ALIGNMENT for related discussion). Today, there are probably no linguists who adhere to the ISV in its original form. Many, however, adopt some version within a newer conception of sentential prosodic structure, which is also, or even largely, composed of an intonational structure.

2.1 *The infinite stress view*

The earliest views saw stress as an increase in loudness due to greater amplitude of the sound wave (CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE). This implies that stress is a continuous phonetic variable, which can have different values within words as well as across words. The idea that word stress and prominence differences among words are essentially the same phenomenon is expressed in Bloomfield (1933: 111), for instance. He distinguished "emphatic stress," as on the syllable *my* in (2) (where I have reproduced his original transcription of *This is my birthday present*), from "high stress," as in the syllables *this* and *birth-*, which is different from "low stress" on *pres-*, which in its turn is distinct from "no stress," on all other syllables.

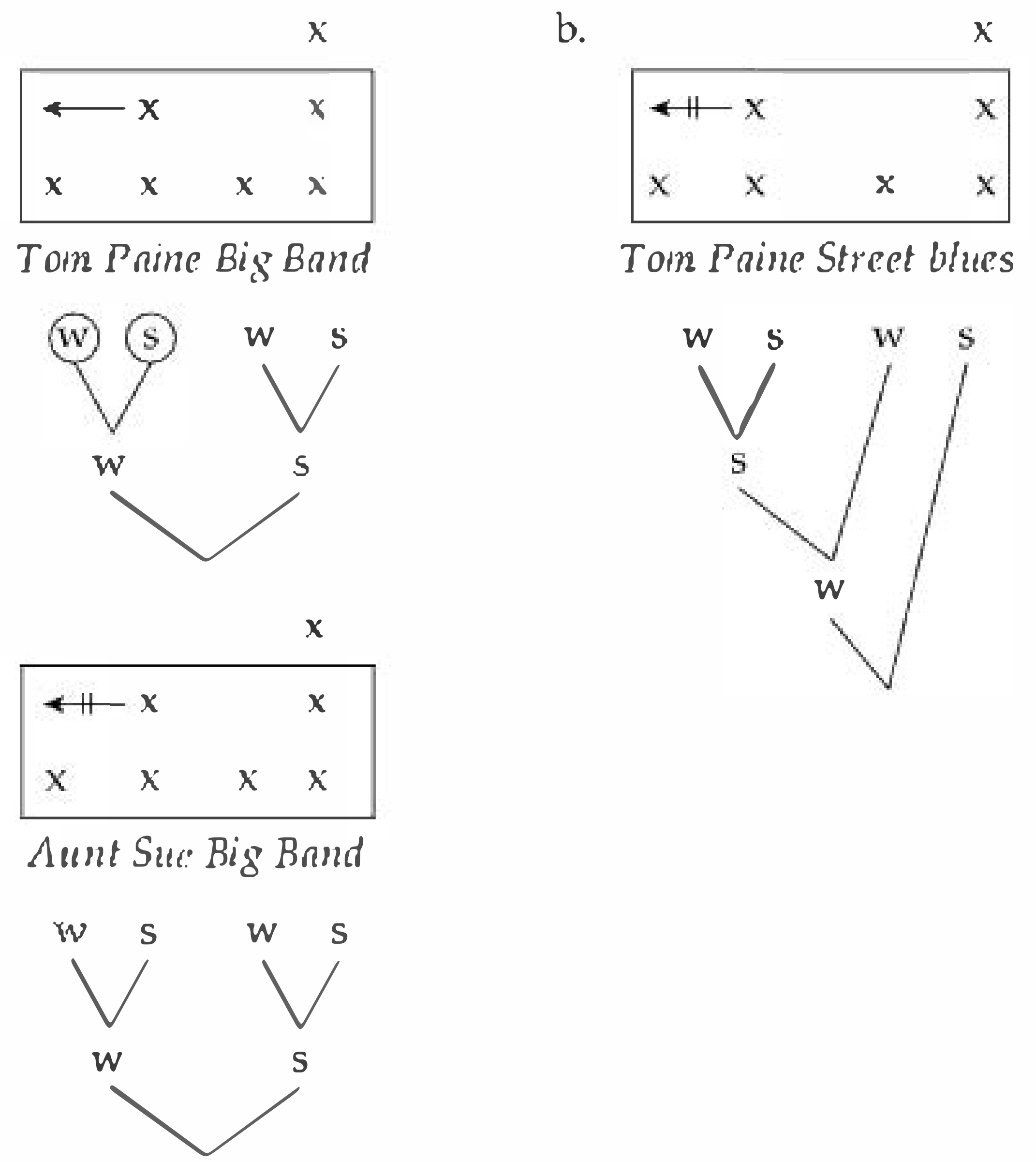
(2) ˈðis iz ˈˈmaj ˈbəːθdej ˌpreznt ("i.e. not yours")

Chomsky and Halle (1968: 20) incorporated the ISV in their rule-based grammar of English stress. Example (3a) has the most prominence on *board*, but *black* can be felt to have more prominence than *eraser*. The stress pattern is therefore 2 1 3. In (3b), *black* is most prominent, while *eraser* has more prominence than *board*, giving 1 3 2. Finally, (3c) again has the most prominence on *board*, but of the other two words it is *eraser* that seems more prominent than *black*, giving 3 1 2. The stress levels 2 and 3 are best compared across the three structures, instead of within them, for instance by comparing *black* in (3a) and the same word in (3c).

(3)	a.	*a black board-eraser*	'a board-eraser that is black'	2 1 3
	b.	*a blackboard eraser*	'an eraser for a blackboard'	1 3 2
	c.	*a black board eraser*	'an eraser of a black board'	3 1 2

Chomsky and Halle accounted for this three-way distinction by postulating two rules, reproduced in (4a) and (4b), plus a convention, (4c). The order of

(10) a. b. c.



2.2 *The problem with the ISV*

The deeper problem lies in the conception of phonological prominence, more specifically in the hypothesis that prominence levels above the word are of the same kind as prominence levels within the word. The next section will argue that word prosodic structure is essentially different from sentential prosodic structure. Part of the reason why they have been conflated is that acoustic measurement techniques have only been readily available since the last quarter of the twentieth century and, where they existed before, the implications of the data they provided were not always adequately incorporated into the way linguists thought about stress. As we will see in §3, when a speaker produces even a single word, we do not only observe the word prosody, but also its sentence prosody. While the connections between these two representations are strong, they are separate components of the phonological structure of English.

3 The pitch accent view

3.1 *Pitch and stress as independent components in syllable prominence*

The new conception of phonological prominence followed various demonstrations that differences in stress are not expressed in acoustic intensity differences, or even auditory loudness differences. Mol and Uhlenbeck (1956) showed that no

If we abstract away from the presence of a pitch accent, the difference between stressed and unstressed syllables is phonetic in many languages (CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE). The stressed syllables may be longer and have greater intensity, as in Spanish (Ortega-Llebaria 2006), or the segments in them may additionally be more precisely articulated, as in Catalan (Astruc and Prieto 2006; cf. Sluijter and van Heuven 1996). In English, such differences have been phonologized, meaning that the greater duration and the articulatory precision have led to constraints on the segmental composition of stressed and unstressed rhymes. A stressed rhyme in English must consist of a long vowel or diphthong or a short vowel plus a consonant. An unstressed rhyme cannot contain any vowel other than [ə i u] (Bolinger 1986: 347).[2] Most importantly from the point of view of our topic, only stressed syllables can be accented.

3.3 *Accent*

Accent is a place marker in the phonological structure where tones are to be inserted (Goldsmith 1977; Hyman 1978; CHAPTER 42: PITCH ACCENT SYSTEMS). The location of these accents, as well as their presence, may be lexically determined, as in Japanese (see also CHAPTER 120: JAPANESE PITCH ACCENT). For instance, (12a) and (12b) are accented words, but differ in the location of the accent, while (12c) is an unaccented word.

(12) a. hási 'chopsticks' b. hasí 'bridge' c. hasi 'edge'

In other languages, accent is assigned on the basis of phonological or morphological information, as in Nubi, a creole spoken in Uganda and Kenya. All Nubi words have one accented syllable, but a verb loses its accent when it combines with an object into a nominalized VP. In (13a) and (13b), a minimal pair of verbs is illustrated in combination with an object; in (13c) the nominalization is shown, and since both 'rent' and 'rent out' now lose their accent, the sentence is ambiguous (Gussenhoven 2006).

(13) a. ána gí pángisa júa 'I will rent out a house.'
I will rent-out house
b. ána gí pangisa júa 'I will rent a house.'
c. pangisa júa séme má 'Renting (out) a house is good.'
rent(-out) house good be

How does English differ from Nubi? The first difference concerns the existence of a choice among the tone(s) to be inserted in any accent position. In Nubi, an accented syllable must have an H* pitch accent. In English, the requirement is merely that a pitch accent be inserted, but which one, out of the many that could be

[2] The final vowel in words like *fellow* was taken to be an allophone of unstressed [u] by Bolinger. Before consonants, [ɪ] may appear, but the situation varies across varieties. In mainstream American English, [ə] and [ɪ] are allophones, with [ɪ] appearing before velars, as in *secure* [sɪ'kjur], and [ə] elsewhere, as in *Sinatra* [sə'natrə]. In a number of varieties, including ones spoken in the USA, [ə] varies with [ɪ] in some contexts (e.g. *peerless* is ['pɪɹləs] or ['pɪɹlɪs], while contrasting them in other contexts, as in *roses vs. Rosa's* or, for non-rhotic dialects, *offices vs. officers*. In unstressed syllables, a syllabic consonant may take the place of a historical [ə] (e.g. *listen* /'lɪsn̩/).

English has a rather large number of accentuation and deaccentuation rules. First, words may have more than one accent, specifically words that have stressed syllables before the main stress, as in *CALiFORnia* (Selkirk 1984). Second, in addition to two rules that resemble the two Nubi rules, it has a rule deleting initial accents and, importantly, a rule that distributes accents to mark information structure. §4 is devoted to accentuation in English.

4 Accent in English

There are two ways in which the description of accentuation in English could be approached. We could either assume that accents are absent and assign them where they are needed, or we could assume that syllables are accented by default and assume they are removed where they do not occur. The second approach is in tune with descriptions of information structure in which accents are removed for "givenness," instead of assigned for "newness" (Schwarzschild 1999) and was adopted for other rules in Gussenhoven (1991, 2005). In the deaccentuation option, accents are first assigned to all syllables that can ever be accented, and then removed where they are not needed. In §4.1, I define the accented syllables of English in two steps, one for primary stressed syllables and one for secondary stressed syllables. §4.2 then describes three deaccentuation rules, the compound deaccentuation rule (henceforth the "Compound Rule"), the Initial Accent Deletion rule, and a rule I will refer to as the "Inverse Compound Rule." §4.3 is devoted to the Rhythm Rule. §4.4 contains three sections dealing with meaningful phrase-level effects: Schmerling's generalization (§4.4.1), phatic elements (§4.4.2), and information structure (§4.4.3). §4.5 deals with the status of pre-nuclear accents. Finally, §5 discusses prominence distinctions that cannot be attributed to stress or accent, and §6 concludes the chapter.

4.1 Lexical generalizations in English

Taking the position of main stress for granted, the first accentuation rule is straightforward: place an accent on the primary stress of every word. The words in (12) come out as in (17) as a result.

(17) aLASka, sarDINE, cigaRETTE, HELicopter, marihuAna, MELancholy

Secondary stressed syllables before the primary stress are also provided with pitch accents. As a result, the representations of *sardine, cigarette,* and *marihuana* are as in (18).

(18) SARDINE, CIGaRETTE, MARiHUAna

Two comments are required here. First, English has "cyclic" stress (Kiparsky 1979). This means that the primary stress of a base survives a morphological derivation that adds a primary stress to the right of the base. The old primary stress will now be a pre-primary secondary stress. In (19a) and (19b), the secondary stress is on the second syllable of the word, because that is where the base has its primary stress; in (19c) it is on the first syllable, for the same reason. In terms of their accentability, there is no difference between the primary stressed syllables

4.2.3 Reverse compounds

"Reverse compounds" are structures in which the first constituent is deaccented. When the first constituent in the NP denotes the category and the second constituent is the name of the item belonging to that category, the first constituent is deaccented. Example (25b) is from Selkirk (1984: 221), who pointed out the contrast with the phrasal *LAKE HILL* 'a hill called after Mr or Mrs Lake', where no deaccenting takes place. The example is representative of geographical names of this type, like *Mount Kilimanjaro*; (25c) illustrates a person category.

(25)	a.	the BOOK JOSHua	→	the book JOSHua
	b.	LAKE HILL	→	lake HILL
	c.	AUNT AGatha	→	aunt AGatha

4.2.4 Initial Accent Deletion

Initial Accent Deletion causes all accents except the last to be deleted in a class of morphological formations, most strikingly compounds. In (26), examples are given in which the first constituent has two accents underlyingly. In the compound noun (26a), the double-accented word *marihuana* loses its pre-final accent, and in (26b) *Tom* undergoes the same fate, as it is the pre-final accent of the name *Tom Paine*, a noun phrase included as the first member of a compound noun. Similarly, *Old* loses its accent in the noun phrase *Old English*, because it too is embedded in a noun compound, while (26d) shows that multiple pre-nuclear accents will all be deleted. Example (26c) forms a minimal pair with the noun phrase *old English teacher* 'an English teacher who is old', where the head noun is the compound *English teacher*: here *old* retains its accent. Similarly, *a Second Language Conference* ("a conference on the study of L2") has an accent just on *Language*, while *a second Language Conference* ("the second of a series of conferences on language") has accents on *second* and *Language*.

(26)	a.	MARiHUAna cigarette	→	mariHUAna cigarette
	b.	TOM PAINE street	→	tom PAINE street
	c.	OLD ENGlish teacher	→	old ENGlish teacher
	d.	i-COULDn't-CARE-LESS attitude	→	i-couldn't-care-LESS attitude

Initial Accent Deletion also applies to English words formed with the help of suffiixes that leave the stress pattern of their base intact. These are known as "stress-neutral" suffixes, and include *-ish*, *-ist*, *-ly*, *-ness*, and *-less*. ("Stress-changing" suffixes include *-ian*: compare *ALbuquerque* – *ALbuquerqueish*, but *ALbuquerque* – *ALbuQUERquian*.) The suffixes *-based*, *-fast*, *-free*, *-like*, *-proof*, *-prone*, *-style*, *-tight*, *-type*, *-wise*, and *-worthy* also belong here, as does the accented suffix *-esque*. The examples in (27) are presented without a "walk-through," but notice the case of (27d), where the base, *REMbrandt*, starts out with only a single accent. Here it is the accent on the suffix that is the last accent in the derived word.

(27)	a.	UNKINDness	→	unKINDness
	b.	TAIPEI-based	→	taiPEI-based
	c.	NORTH-koREa-style	→	north-koREa style
	d.	REMbrandTESQUE	→	rembrandTESQUE

Reverse compounds, too, undergo Initial Accent Deletion. If the second member has two accents, only the last of these survives. In fact, the Reverse Compound Rule could be subsumed by Initial Accent Deletion, but I have preferred to treat the reverse compounds separately, in order to identify them as a group that requires further research.

(28) a. mount KIlimanJAro → mount kilimanJAro
b. route SIXty-ONE → route sixty-ONE

4.3 A post-lexical phonological generalization in English

Once words are inserted into phrases, further deletions of accents may take place. These "post-lexical" generalizations are in part phonological, meaning they are triggered by specific phonological contexts, and in part morphosyntactic, in which case the deletion of the accent is the direct expression of a meaning distinction. The most widely discussed phonological generalization, earlier illustrated in (8), concerns "stress shift," more properly "rhythm-induced accent deletion," also known as the Rhythm Rule. In the PAV, this rule deletes medial accents in the phonological phrase (Gussenhoven 1991; Shattuck-Hufnagel *et al.* 1995). Consider the alternants of *Japanese* in (29). In (29a), we have an isolated pronunciation, in which both accents are retained. In (29b) the first syllable loses its accent, as it is medial in the phonological phrase, while in (29c) the third syllable loses its accent, because *it* is medial. The phonological phrase (ϕ) is a prosodic constituent above the phonological word, and corresponds in the default case with a syntactic phrase.

(29) a. $(\text{JAPaNESE})_{\phi}$
b. $(\text{GOOD japaNESE})_{\phi}$
c. $(\text{JAPanese GOODS})_{\phi}$

The Rhythm Rule makes it clear that it is important to describe the presence of pre-primary accents in words. For instance, (30a) shows that words like *sepTEMber* and *senSAtional* – cf. (20) – really lack an accent on the word-initial foot, while (30b), from Jones (1967), shows that disyllables like *DUNDEE* do have one, just as words like *inTERpreTAtion* and *asSOciAtion*, shown in (32c); cf. (19). Example (30d) shows the effect of *-esque*, due to Initial Accent Deletion. In (30e) (cf. (12a) from Prince 1983), we again see the effect of that rule. *Street* has lost its accent through the Compound Rule and *Tom* lost its accent through Initial Accent Deletion, meaning that there is no medial accent left for the Rhythm Rule to delete. In (30f), there are two medial accents to be deleted by the Rhythm Rule. Next, in (30g)-(30i), we see the effect of Initial Accent Deletion on reverse compounds, and the subsequent inability of the Rhythm Rule to contribute to their prosodic shape.

(30) a. $(\text{sepTEMber STORMS})_{\phi}$ *not* $(\text{SEPtember STORMS})_{\phi}$
b. $(\text{DUNDEE MARmalade})_{\phi}$ → $(\text{DUNdee MARmalade})_{\phi}$
c. $(\text{asSOciATion FOOTball})_{\phi}$ → $(\text{asSOciation FOOTball})_{\phi}$
d. $(\text{rembrandTESQUE LIGHT})_{\phi}$ *not* $(\text{REMbrandtesque LIGHT})_{\phi}$
e. $(\text{tom PAINE street BLUES})_{\phi}$ *not* $(\text{TOM paine street BLUES})_{\phi}$

f.	(TOM PAINE BIG BAND)$_{\varphi}$	→	(TOM paine big BAND)$_{\varphi}$
g.	(aunt SUE BIG BAND)$_{\varphi}$	→	(aunt SUE big BAND)$_{\varphi}$
		not	(AUNT sue big BAND)$_{\varphi}$
h.	(mount kilimanJAro BLUES)$_{\varphi}$	*not*	(MOUNT kilimanjaro BLUES)$_{\varphi}$
		not	(mount KIlimanjaro BLUES)$_{\varphi}$
i.	(route sixty-ONE BLUES)$_{\varphi}$	*not*	(ROUTE sixty-one BLUES)$_{\varphi}$
		not	(route SIXty-one BLUES)$_{\varphi}$

4.4 *Meaningful distinctions at the phrasal level*

The motivation for having pitch accents on particular words in English sentences has traditionally been looked at from a syntactic perspective. The Nuclear Stress Rule of Chomsky and Halle (1968), given in (4b), applied inside phrases, as illustrated in (5a) and (5c), as well as to combinations of phrases. As a result, (31a) was interpreted as having the primary stress on the object NP *Mary*, a pattern known as "normal" stress. Other prosodic versions of the same sentence, like (31b), were said to have "contrastive stress," due to contextual constraints that more recently have been collectively called "information structure." In (31b), a contrastive stress rule, not formalized, has retained the primary stress on *John*, demoting the stress levels on *kissed* and *Mary* as a result.

(31) a. [[*John* $]_{NP}$ [*kissed* [*Mary* $]_{NP}]_{VP}]_{S}$

1	1	1	Starting point
	1	1	(4b)
	2	1	
1	2	1	(4b)
2	3	1	Output

b. A: *Who kissed Mary?*

1 3 2

B: [[*John* $]_{NP}$ [*kissed* [*Mary* $]_{NP}]_{VP}]_{S}$

Schmerling (1976) and others have pointed out that the distinction between "normal" and "contrastive" is not easily maintained. "Normal" stress may seem an unproblematic concept in (31a): this is how the sentence would be produced when read from a contextless written presentation. In many situations, however, such a reading may strike the listener as quite marked. In (32a), the prominence on *two* seems both "contrastive," in the sense that the two-year-old is compared with some older individual, and "normal," in the sense that this is how any native speaker would read it. Schmerling (1976) also pointed out that the way her husband announced the death of President Johnson to her was as (32b), an out-of the-blue statement, with no accent on *died*. As it happened, some weeks earlier, her mother had announced the death of another American president, Truman, as (32c). Unlike Johnson, Truman had been critically ill for some time and his death came as no surprise. I here give Schmerling's (1976) prominence patterns as accentuations. It is hard to see which utterance has "normal stress"; if a choice must be made, it would be (32b), unlike what the Nuclear Stress Rule predicts.

(32) a. Even a TWO-year-old can do that
b. JOHNson died
c. TRUman DIED

(41) I'm sitting here very NICEly, thank you very much.

Hearer-appeal markers intend to engage the listener, like vocatives, "softening" expressions like *don't you think, eh,* and equal polarity tags (42).

(42) It's SUCH a cute CHILD, don't you think, Peter?

Textual markers include comment clauses, as in (43a), and reporting clauses, as in (43b). Reporting clauses may be complex, and still show a sustained absence of accents. Unaccented reporting clauses may be integrated in the intonational structure in two ways. They may be included in the intonational phrase, just like comment clauses, in which case any final boundary tone appears after the reporting clause, or be cliticized, in which case the final boundary tone occurs before the reporting clause, to close off the reported direct speech, as well as after it. In the latter option, probable for (43a), there is a noticeable boundary between the reported direct speech and the reporting clause.

(43) a. It's tomorrow, I think.
b. 'Then why not GO there?', he interrupted, no longer feeling he didn't care.

Approximatives are expressions that indicate that information is less than precise. This is a quite varied group, with items like *or more, the way he did, or something, and all that, kind of thing, in a way.*

(44) I suppose we all ARE in a way.

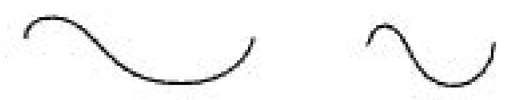
Epithets are appositive descriptions, often derogatory, as illustrated in (45). Example (45a) is from Bing (1979). In Gussenhoven (1986: 128), I erroneously argued, contra Bing (1979), that epithets are accented. My confusion was due to the frequent occurrence of a final H% boundary tone in this kind of sentence, added to the fact that, like many reporting clauses, epithets are typically set off from the preceding clause by an intonational phrase boundary. Thus, if *neighbors* in (45a) is giving a falling–rising contour (H*L H%: high pitch on *neigh-* and low-to-high pitch on *-bors*), the pitch on *the finks* will repeat L H%, the low-to-high pitch of *-bors*. To see how the pattern goes without H%, (45b) naturally ends in L%, once after *is* (H*L L%) and once after *boy* (L L%). As Bing pointed out, this pattern contrasts with the repetition of the pitch accent: if the H*L H% is repeated on *the finks*, the expression changes into an appellative apposition, *Fink* now being their surname; see (46b). Similarly, a polar tag is accented, as shown in (46b), where *IS* repeats the contour of *ILL*, H*L L%.

(45)

a. My nextdoor NEIGHbors, the finks, have been coming over EVery NIGHT.


b. Here he IS, the stupid boy.

(46)



a. My nextdoor NEIGHbors, the FINKS, will be OUT tonight.

b. He ¡ISn't ILL then, IS he?

4.4.3 *(De)accenting for "newness" and "givenness"*

Deaccenting for "givenness" is illustrated in (47a), where the phrase *in the city* is a stand-in for *Toulouse*. The absence of a pitch accent on *city* causes the predicate *set foot* to have the last pitch accent. By contrast, in (47b) *in the city* refers to a specific district in London, and therefore conveys different information from *London*. Pitch accents thus express "information structure," the way the information in the sentence relates to the state of understanding. The lack of accent on *city* in (47a) indicates that reference is made to existing information, just as the presence of the accent on the same word in (47b) indicates that some referent other than "London" is to be understood, i.e. the financial district known as "the City."

(47) a. They can't have been seen in Toulouse. They NEVer set FOOT in the city.
b. They haven't really seen much of London. They NEVer set foot in the CITy.

The relation between accentuation and information structure in English and other European as well as Asian languages is a widely studied phenomenon. Recent volumes are Bosch and van der Sandt (1999), Molnár and Winkler (2006), and Lee *et al.* (2007). Here I discuss two issues that have played a role in the discussion over the past decades: (i) highlighting and (ii) the focus constituent.

(i) *Highlighting*: Dwight Bolinger was the main proponent of the position that a pitch accent lends significance to the word it is used on, even in a compound or a predicate–argument combination. This "highlighting view" was originally launched in reaction to the syntax-based derivation of "normal stress" illustrated in (31a) (Bolinger 1972), but later repeated in response to Schmerling (1976) (Bolinger 1977) and Gussenhoven (1983a) (Bolinger 1985, 1987). The highlighting role of pitch accents is evident not only in the expressive use of pitch accents in (48a) – cf. Bolinger (1987: example (145)), which goes against Predicate Deaccentuation – but also in the unusual locations of pitch accents, "distortion," as in (48b) (Bolinger 1985: example (32)). Counterpresuppositional accentuations like that in (48c), which can be analyzed as representing negative polarity focus, must thus be motivated by the meaning of *in* in Bolinger's analysis (Bolinger 1985).

(48) a. Your FEET STINK.
b. A: You've overlooked the possibility of dealing with him personally.
B: Not at all. There never really was a POSsibility.
c. They can't have been seen in Toulouse. They were never IN the city.

The expressive power of pitch accents and the expressive motivation for some pitch accent locations is undeniable (cf. also Ladd 2008: 246), but the consensus in *A and* [*B and* C] than in [*A and B*] *and* C. Second, Katz and Selkirk (2009) produce evidence that a contrastive accent like that on *Modigliani* in (52) has greater acoustic prominence than the phonetically less salient thematic accent in *Moma*, which Zubizaretta (1998: 84) might call a "miniature" accent. Katz and Selkirk (2009) argue that there is a binary contrast here.

(51) Gary is a really bad art dealer. He gets attached to the paintings he buys. He acquired a few Picassos and fell in love with them. The same thing happened with a Cezanne painting. So he would *only* offer that $[\text{Modigliani}]_{FOC}$ to $[\text{MoMA}]_{New}$. I bet the Picassos would have fetched a much higher price.

The third case concerns "second occurrence focus" (SOF). It refers to the unaccented repetition of a constituent in the scope of a "focus operator" like *also, even, only* which was accented in the preceding sentence. Féry and Ishihara (2009) found small but significant differences between postnuclear SOF, as illustrated in the B response in (52a), and non-focus, as illustrated in the B response (52b), even though both instances of *song* are unaccented. Their experiment was on German; (52) improvises on the kind of data they used (cf. also Beaver *et al.* 2007).

(52) a. A: It was a fun party. Michael even sang a SONG.
B: And WALdemar even sang a [song].
b. A: Did MANy people sing a song at the party?
B: Yes! WALdemar even sang a song.

Fourth, there may be durational differences and possibly pitch differences between unaccented primary stresses and unaccented secondary stresses. For instance, if we placed the compound *SPANish student* and the phrase *a SPANish STUdent* in a reporting clause (where their accents must disappear), are these structures still different phonetically? And is the answer the same in the case of a minimal pair of single words, like the noun *INterchange* and the verb *INterCHANGE* (cf. Schmerling 1976: 24; Gussenhoven 1991)? Fifth, even though listeners may not use the information, there are reports that narrow focus, shown in (53a), is pronounced differently from broad focus, shown in (53b). Depending on the language, the narrow focus pronunciation may have a longer duration, a steeper slope, a later peak, an earlier peak, or a higher peak (Avesani and Vayra 2004; Smiljanić 2004; Sityaev and House 2004; Baumann *et al.* 2007; Hanssen *et al.* 2008).

(53) a. A: Were you going to drive to Bristol, you said?
B: We're going to drive to [WINDsor].
b. A: What are you going to do today?
B: We're going to [drive to WINDsor].

There are two possible responses to these five (real and potential) effects. One is that the PAV needs to be enriched with a representation of gradient prominence, thus returning to the position in Chomsky and Halle (1968). In the context of the differences in phrasal branching referred to above, Ladd (2008) suggested that this representation should be *tonal*, rather than be in terms of levels of stress. Accordingly, he proposed a tree structure much like the metrical tree

ACKNOWLEDGMENTS

I thank Haruo Kubozono and Hubert Truckenbrodt for their productive discussion of some of the issues in this chapter and two anonymous reviewers for their useful comments on an earlier version. I am grateful to Luise Dorenbusch for spotting a number of errors at proof stage.

REFERENCES

Adger, David. 2007. Stress and phrasal syntax. *Linguistic Analysis* 33. 238–266.

Astruc, Lluïsa & Pilar Prieto. 2006. Stress and accent: Acoustic correlates of metrical prominence in Catalan. In Antonis Botinis (ed.) *Proceedings of the ISCA Tutorial and Research Workshop on Experimental Linguistics*, 73–76. Athens: University of Athens.

Avesani, Cinzia & Maria Vayra. 2004. Broad, narrow and contrastive focus in Florentine Italian. In Solé *et al.* (2004), 1803–1806.

Baumann, Stephan, Johannes Becker, Martine Grice & Doris Mücke. 2007. Tonal and articulatory marking of focus in German. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 1029–1032. Saarbrücken: Saarland University.

Beaver, David, Brady Zack Clark, Edward Flemming, T. Florian Jaeger & Maria Wolters. 2007. When semantics meets phonetics: Acoustical studies of second occurrence focus. *Language* 83. 245–276.

Beckman, Mary E. 1986. *Stress and non-stress accent*. Dordrecht: Foris.

Bing, Janet M. 1979. Aspects of English prosody. Ph.D. dissertation, University of Massachusetts, Amherst.

Bing, Janet M. 1984. A discourse domain identified by intonation. In Gibbon & Richter (1984), 10–19.

Birch, Stacy & Charles Clifton, Jr. 1995. Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech* 33. 365–391.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Bolinger, Dwight L. 1958. A theory of pitch accent in English. *Word* 14. 109–149.

Bolinger, Dwight L. 1972. Accent is predictable (if you're a mind-reader). *Language* 48. 633–644.

Bolinger, Dwight L. 1977. Review of Schmerling (1976). *American Journal of Computational Linguistics* 14. 1–23.

Bolinger, Dwight L. 1985. Two views of accent. *Journal of Linguistics* 21. 79–123.

Bolinger, Dwight L. 1986. *Intonation and its parts: Melody in spoken English*. London: Edward Arnold.

Bolinger, Dwight L. 1987. More views on "Two views of accent." In Carlos Gussenhoven, Dwight L. Bolinger & Cornelia Keijsper (eds.) *On accent*, 124–146. Bloomington: Indiana University Linguistics Club.

Bosch, Peter & Rob van der Sandt (eds.) 1999. *Focus: Linguistic, cognitive, and computational perspectives*. Cambridge: Cambridge University Press.

Bresnan, Joan. 1971. Sentence stress and syntactic transformations. *Language* 47. 257–281.

Büring, Daniel. 2006. Focus projection and default prominence. In Molnár & Winkler (2006), 321–346.

Chomsky, Noam. 1972. Deep structure, surface structure, and semantic representation. In Danny D. Steinberg & Leon A. Jakobovitz (eds.) *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology*, 183–216. London: Cambridge University Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Selkirk, Elisabeth. 1995. Sentence prosody: Intonation, stress, and phrasing. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 550–569. Cambridge, MA & Oxford: Blackwell.

Shattuck-Hufnagel, Stefanie, Mari Ostendorf & Ken Ross. 1995. Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics* 22. 357–388.

Siraj, Pasha. 2008. Stress-dependent word tone in Singaporean English. Poster presented at TIE3 (Tone and Intonation in Europe), Lisbon.

Sityaev, Dmitry & Jill House. 2004. In Solé *et al.* (2004), 1819–1822.

Sluijter, Agaath M. C. & Vincent J. van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100. 2471–2485.

Smiljanić, Rajka. 2004. *Lexical, pragmatic, and positional effects on prosody in two dialects of Croatian and Serbian.* London: Routledge.

Solé, M. J., D. Recasens & J. Romero (eds.) (2004). *Proceedings of the 15th International Congress of Phonetic Sciences.* Barcelona: Causal Productions.

Steedman, Mark. 2000. Information structure and the syntax–phonology interface. *Linguistic Inquiry* 31. 649–689.

Swerts, Marc, Emiel Krahmer & Cinzia Avesani. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30. 629–654.

Truckenbrodt, Hubert. 1995. Phonological phrases: Their relation to syntax, focus, and prominence. Ph.D. dissertation, MIT.

Truckenbrodt, Hubert. 2006. Phrasal stress. In Keith Brown (ed.) *Encyclopedia of language and linguistics*, 2nd edn., vol. 9, 572–579. Oxford: Elsevier.

Turk, Alice & Mariko Sugahara. 2009. Durational correlates of English sublexical constituent structure. *Phonology* 26. 477–524.

Warner, Natasha, Allard Jongman, Joan A. Sereno & Rachèl Kemps. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32. 251–276.

Winkler, Susanne. 1997. *Focus and secondary predication.* Berlin & New York: Mouton de Gruyter.

Zubizaretta, Maria Luisa. 1998. *Prosody, focus, and word order.* Cambridge, MA: MIT Press.

117 Celtic Mutations

S. J. Hannahs

1 Introduction: Initial consonant mutation

Initial consonant mutation is one of the best known, although not best understood, phonological characteristics of the Celtic languages: the initial consonant of a word form undergoes a systematic alternation with a phonetically distinct consonant in specific contexts. (On mutation in languages other than Celtic see CHAPTER 65: CONSONANT MUTATION.)

As an example of such a systematic alternation in the initial consonant of a word, consider Welsh. The Welsh word for 'father', *tad*, has four distinct forms: the citation form *tad* [taːd], the "soft mutation" form *dad* [daːd], the "aspirate mutation" form *thad* [θaːd], and the "nasal mutation" form *nhad* [n̥aːd].

(1) *Welsh initial consonant mutations*

citation form	*tad*	[taːd]	'father'
soft mutation	*dy dad*	[də daːd] (2SG father)	'your father'
aspirate mutation	*ei thad*	[i θaːd] (3FEM SG father)	'her father'
nasal mutation	*fy nhad*	[ə n̥aːd] (1SG father)	'my father'

Which of these four forms occurs in any particular instance depends on the wider context in which the word appears, as discussed below. Here we see that following the 2nd person singular possessive pronoun *tad* mutates to *dad*, following the 3rd person feminine singular possessive pronoun *tad* mutates to *thad*, and following the 1st person singular possessive pronoun *tad* mutates to *nhad*. The number and nature of the mutations differ between the modern Celtic languages, though there are some overarching similarities across the family, such as mutations following specific prepositions, mutations associated with particular grammatical contexts, and so forth. Within each of the two branches of the Celtic family, the Brythonic languages (also known as "P-Celtic"), Breton and Welsh, have very similar patterns to each other, as do the Goidelic languages (or "Q-Celtic"), Irish and Scottish Gaelic.[1]

[1] Note the conventional difference in pronunciation between Irish Gaelic [geilɪk] and Scottish Gaelic [gælɪk], often called simply *Gaelic* [gælɪk], reflecting the pronunciation of each in its own language.

Although initial consonant mutation has its origin in phonetics as a series of phonetically triggered external sandhi alternations, the overwhelming bulk of initial consonant mutation in the modern languages is not phonetically triggered. Indeed, it is not phonologically transparent. Rather, it is lexical and/or grammatically conditioned. Nonetheless, it is possible to sketch an outline of the historical phonetic bases of initial consonant mutation, and to discuss the general absence of phonological transparency. Moreover, despite this general absence of phonological transparency, there are a number of recent accounts (syntactic, phonological, and morphological) of productive instances of initial consonant mutation, for instance analyses of "soft mutation" (i.e. lenition) in Welsh. Beyond such productive occurrences of initial consonant mutation there is a large "residue," that is, phonologically non-productive (typically lexicalized or grammaticalized) aspects of mutation. This residue – which may well be syntactically or morphologically productive – nevertheless has phonological implications: even in the absence of productive phonological processes, initial consonant mutation still requires some sort of phonological representation, along with a phonological characterization of the alternations themselves.

A brief word on terminology is in order here before looking at the languages and the data involved. The statement above that initial consonant mutation has its origins in phonetics means simply that the alternations were originally phonetically transparent, e.g. lenition involved voicing and/or frication of intervocalic consonants; nasal mutation involved contact with a nasal segment. In the modern languages many of the alternations have become grammaticalized and phonologically opaque in various ways. The object of a specific preposition may undergo a particular mutation, and the complement of a particular pronoun may show a specific mutation, in both cases in the absence of a transparent phonetic context triggering the alternation. This in turn leads to an interesting analytical dilemma: although the mutations affect the phonological shape of word forms, the alternations themselves may be triggered by non-phonological structures in a phonetically opaque way. That is, they may be subject to grammatical (e.g. syntactic or morphological) considerations outside the phonology proper. Indeed, some (e.g. Stewart 2004; Green 2006, 2007) have taken the lack of a unitary cause of mutation and an absence of unitary effect as evidence that it is not a unitary phenomenon and not therefore properly phonological.

After a brief discussion of the languages involved, the diachronic origins of initial consonant mutation, and the problem of determining whether the mutations are best analyzed in terms of phonology, morphology, or syntax (or some combination thereof), this chapter will look at the data of initial consonant mutation in the Celtic languages, discuss the theoretical interest in initial mutation, and survey some of the more influential analyses of initial consonant mutation.

2 Background

2.1 *The Celtic languages*

There are four living Celtic languages, belonging to two branches of the Celtic family of Indo-European. The Brythonic (or Brittonic) branch includes Breton (spoken in Brittany in France) and Welsh (spoken in Wales and parts of Patagonia), whilst the Goidelic branch includes Irish (spoken in Ireland) and Scottish Gaelic (spoken in Scotland and on Cape Breton Island, Canada). The family also includes the relatively recently extinct Cornish (Brythonic) and Manx (Goidelic), which are both subject to current revival movements, but which will not be considered here. (Pre-)historically, the Celtic family also included Gaulish, Cumbric, and Celtiberian, and the evidence suggests that they also displayed initial mutations (see for instance Fleuriot 1975 and Gray 1944; though Jackson 1953: 548 is more skeptical).

2.2 *Origins of initial consonant mutation*

The Celtic mutations were originally phonetically triggered external sandhi forms. For example, the initial mutation variously referred to as "soft mutation" in Welsh and Breton, "lenition" in Breton or Scottish Gaelic, and "aspiration" in Irish, is traceable to phonetically conditioned intervocalic lenition (see CHAPTER 66: LENITION). Writing about lenition in "British," i.e. Brythonic, Jackson (1953: 543) notes that it occurred "intervocally, whether in the interior of a word or initially when preceded in a close speech-group by a proclitic ending in British in a vowel . . ." It affected the British single voiceless stops [p t k], which became voiced [b d g]. [b d g], in turn, along with [m], spirantized to [β ð ɣ] and [ṽ].[2] The [ṽ] subsequently denasalized to [v], while the voiced velar fricative [ɣ] later deleted in Welsh (in certain cases becoming [j]); it became [x] in Cornish and Breton.

Generally, then, in British (and in Celtic more widely) intervocalic stops lenited word-internally. Word-initially, single consonants lenited when preceded by a vowel-final word. According to Falc'hun (1951: 84f.), "Originally, mutations were purely phonetic. In words closely linked with each other, the ending of one word, depending on whether it was vowel final or consonant final, triggered the alternation of particular consonants at the beginning of the following word."[3] As Hickey (1997: 146f.) points out, "Any consideration of the origin of mutation must start with phenomena which are purely allophonic. . . . Low-level phonetic lenition is a necessary precondition to the development of morphological lenition at a much later point." Thus, the process that started out as a phonetically motivated allophonic alternation became morphologized.

This distinction between allophonic lenition and morphological lenition goes to the heart of the difficulty of trying to account for initial lenition – or any of the other initial consonant mutations – in the Celtic languages as solely phonological. Although initial consonant mutation does have its origins in phonetics, it has long since ceased to be a transparent phonological phenomenon.

Falc'hun (1951: 85), for example, notes that certain phonological changes led to the grammaticalization of initial consonant mutation. Specifically, stress shift and the subsequent loss of endings in Celtic removed the phonetic context for across-the-board initial mutations, yet the mutations themselves persisted, having become grammaticalized. Given that older forms of Celtic had large numbers of

[2] Jackson uses the symbols ƀ, đ, ʒ, and μ, respectively. I have replaced these here with the standard IPA symbols.

[3] "A l'origine, les mutations étaient des faits purement phonétiques. Dans les mots intimement liés entre eux, la désinence du premier mot, suivant qu'elle finissait par une voyelle ou par une consonne, entrainait l'alternance de certaines consonnes au début du mot suivant." [Translation SJH]

Syntactic soft mutation in Welsh has long been alleged to mark the direct object (see Zwicky 1984, and more recently Roberts 2005). However, as Tallerman (2006) shows unambiguously, and as Borsley *et al.* (2007) discuss at length, an account of soft mutation based on direct objects (or abstract accusative case marking) fails to account for the data. Even considering just the data in (4), the mutated word doesn't have a single grammatical function nor does it belong to a single syntactic category – in (a) it's a direct object, in (b) it's the head of a verb phrase, in (c) it's the verb belonging to an embedded clause, in (d) it's the verb in an embedded infinitival clause, in (e) it's the subject of the sentence. So, even leaving aside the absence of phonological conditioning and restricting the focus to grammatical conditioning, the grammaticalization of initial consonant mutation is both highly complex and multidimensional.

It is important to note, too, that although the examples above are of a single initial mutation in a single language, the lenition process known as "soft mutation" in Welsh, these examples are nevertheless indicative of the variety of triggers for initial consonant mutation in the Celtic languages in general. Taking the example of Irish, the Christian Brothers' grammar (n.d.: 15) includes the following under the "rules for eclipsis" (i.e. a type of nasalization, discussed below):

(5) *Eclipsis in Irish occurs following*

a. the plural possessive adjectives *ár* 'our', *bhur* 'your', and *a* 'their'
b. the article when followed by the genitive plural (both genders)
c. a simple preposition followed by the article and a noun in the singular
d. the numeral adjectives *seacht* (7), *ocht* (8), *naoi* (9), and *deich* (10)
e. the particles *cha* 'not', *an* 'whether', *cá* 'where', *nach* 'whether . . . not', *go* 'that', *muna* 'unless', *dá* 'if', and after the relative particle *a* when it is preceded by a preposition, or when it means 'all that' or 'what'.

So, although initial consonant mutation was originally phonetically motivated, it is now overwhelmingly phonologically opaque. Moreover, even while various mutations are robust and productive, they may also be associated with a number of different contexts, making reasonably comprehensive accounts of initial consonant mutation very difficult to articulate. Nonetheless, over the years there have been attempts to account for initial consonant mutation within various linguistic frameworks and to analyze it from various perspectives, including phonological, morphological, syntactic, or some combination of these, such as morphophonological or morphosyntactic. Furthermore, some success can be achieved through narrowly focusing on a particular aspect of a specific type of mutation in a single language, though this is often not easily generalizable across mutations or necessarily extensible to other languages.

While it is clear that initial consonant mutation in the modern Celtic languages is not primarily phonological, at least five relevant questions still remain for phonology:

(i) What phonological alternations are involved in the mutations?
(ii) How can the alternations be characterized phonologically?
(iii) Where do the alternations occur, i.e. (how) can the alternation site be defined in phonological terms?
(iv) What triggers the alternations?
(v) How are the alternations represented phonologically?

These questions will be referred to again below with reference to various accounts of mutation. Before looking more closely at these accounts, the next section presents the primary data of initial consonant mutation in Irish, Scottish Gaelic, Welsh, and Breton. Note that the data presented are representative of the initial consonant mutations in each of the languages. However, given the wide and significant dialect variation within and across the Celtic languages, coupled with the absence of a spoken standard, the data given will not match the specifics of every dialect.[5]

4 The data

The tables in the following subsections summarize the alternations involved in the principal initial consonant mutations in the four living Celtic languages. The symbols are broadly phonetic rather than orthographic. In each case the top row indicates the radical or underlying form, typically matching the initial consonant in a citation form or dictionary entry; the changes shown are those indicated by the type of mutation in the left-hand column. For example, an Irish word beginning with /b/, e.g. *bó* [bo] 'cow', shows up as *bhó* [vo] under "aspiration" and as *mbó* [mo] under "eclipsis."

(6) *Irish initial consonant mutations* (cf. Grijzenhout 1995: 8)

citation form	*bó*	[bo]	'cow'
aspiration	*a bhó*	[ə vo]	'his cow'
eclipsis	*a mbó*	[ɔ mo]	'their cow'

Note, too, that not all consonants are subject to mutation. Those that do not mutate are not included in the tables in the "Radical" row.

4.1 Q-Celtic (Irish and Scottish Gaelic) mutations

4.1.1 Irish

For an overview of Irish mutation, see Ní Chiosáin (1991) and Grijzenhout (1995: 9), for example. Note that the Irish and Scottish Gaelic consonant systems include a contrast between palatalized ("slender") consonants and non-palatalized ("broad" or "velarized") consonants. The palatalized consonants are indicated by an apostrophe. Both types of consonant undergo mutation, as indicated in the tables. Note also the sonorants represented with a capital letter, [N], [R], and [L], again in both Irish and Scottish Gaelic. Historically, these appear to have been geminate sonorants and their behavior in the modern languages is different from other sonorants, particularly with respect to mutation; synchronically, their phonetic nature

[5] It is of importance to phonological research to note that none of the modern Celtic languages recognizes a single spoken standard. Therefore, unlike analyses of English, in which accounts of specific dialects can refer for example to Received Pronunciation, General American, or some other standard English, there is no standard point of reference for the sound systems of Irish, Scottish Gaelic, Welsh, and Breton.

varies from dialect to dialect, though current research seems to indicate that – at least for Skye Gaelic – they are associated with Advanced Tongue Root, as compared with "normal" sonorants (see Archangeli 2009). Finally, note that Irish [f] is traditionally a voiceless bilabial fricative; this is now giving way to a voiceless labiodental fricative (cf. Christian Brothers, n.d.: 9 and Mac Eoin 1993: 110). For both /f/ and /f′/ in Irish and Scottish Gaelic, the "aspirated" reflex is loss of the consonant (hence "zero" in the table); for /f′/ this results in palatalization of a preceding consonant.

(7) *Mutation table for Irish, showing the radical consonant, aspiration* (a.k.a. lenition), séimhiú, *and eclipsis* (a.k.a. nasalization), urú

Radical	p	p′	t	t′	k	k′	b	b′	d	d′	g	g′
Aspirated ("lenited")	f	f′	h	h′	x	x′	v~w	v′	ɣ	ɣ′~j	ɣ	ɣ′~j
Eclipsed ("nasalized")	b	b′	d	d′	g	g′	m	m′	N	N′	ŋ	ŋ′

Radical	f	f′	s	s′	R	R′	L	L′
Aspirated ("lenited")	zero	zero	h	h′	r	r′	l	l′
Eclipsed ("nasalized")	v	v′	s	s′	R	R′	L	L′

4.1.2 *Scottish Gaelic*

For Scottish Gaelic mutations, see MacAulay (1992b: 238ff.) and Rogers (1972), for example. In addition to the observations above concerning the palatalized *vs.* non-palatalized contrast and the sonorants represented with capital letters, relevant for both Irish and Scottish Gaelic, there is another representational issue with respect to Scottish Gaelic. The stop contrast in Scottish Gaelic is typically represented as a [±voice] contrast, i.e. using the symbols p/b, t/d, k/g. However, the phonetic reality is that in virtually every major dialect the contrast is actually [±aspiration]. Traditionally, the p/b, t/d, k/g notation has been used for typographical convenience – the visual contrast is maintained and it is easier to read, e.g. [b] *vs.* [p] *vs.* [p′] in lieu of [p] *vs.* [pʰ] *vs.* [p′ʰ]. The principal allophones of the stops in Scottish Gaelic are in fact voiceless; the phonological contrast in the stop series is really one of aspiration.

In the mutation table below I have followed the typographic tradition and used the voiced and voiceless symbols. Thus *b* here represents plain unaspirated [p], while *p* represents aspirated [pʰ], *p′* represents [p′ʰ], etc. Absence of a cell in the table, e.g. at the intersection of /s/ and "Eclipsed," indicates that the radical consonant appears in that particular environment.

(8) *Mutation table for Scottish Gaelic, showing eclipsis* (a.k.a. nasalization), uradh, and *lenition* (a.k.a. aspiration), sèimheachadh

Radical	p	p′	t	t′	k	k′	b	b′	d	d′	g	g′
Lenited	f	f′	h	h′	x	x′	v	v′	ɣ	ɣ′	ɣ	ɣ′
Eclipsed ("nasalized")	mp	mp′	nt	nt′	ŋk	ŋk′	mb	mb′	nd	nd′	ŋg	ŋg′

Radical	f	f′	s	s′	L	L′	m	m′	N	N′	R
Lenited	zero	zero	h	h′	L	l′	ṽ	ṽ′	n		r
Eclipsed ("nasalized")	mf	mf′									

4.2 P-Celtic (Welsh and Breton) mutations

4.2.1 Welsh

For Welsh mutations, see for example Morris-Jones (1913: 176) and Borsley *et al.* (2007: 19–26). The radical consonants shown typically have their IPA values, though phonetically /p t k/ tend to be aspirated and /b d g/ tend toward devoicing. /ɬ/ (orthographic <ll>) is realized as a voiceless lateral fricative. /ʰr̥/ represents a voiceless pre-aspirated trill (which either contrasts with a voiced [r], cf. Ball and Williams 2000: 16, or is allophonically related to it, cf. Awbery 1986: 13); this is variously realized as voiceless pre-aspirated [ʰr̥], voiceless post-aspirated [r̥ʰ], a sequence of trill followed by aspiration, [rh], or a sequence of aspiration followed by a trill, [hr]. The soft mutation reflex of /g/ is absence of the segment (zero). The voiceless nasals, [m̥ n̥ ŋ̊], occur only as the nasal mutation reflexes of /p t k/.

(9) *Mutation table for Welsh, showing soft mutation,* treiglad meddal, *nasal mutation,* treiglad trwynol, *and aspirate mutation,* treiglad llaes

Radical	p	t	k	b	d	g	m	ɬ	ʰr̥
Soft	b	d	g	v	ð	zero	v	l	r
Nasal	m̥	n̥	ŋ̊	m	n	ŋ			
Spirant	f	θ	χ						

4.2.2 Breton

For Breton mutations, see for example Press (1986), Hemon (2007: 1), and Jackson (1967: 311). As with the Welsh, the Breton consonants also typically have their IPA values. The only issue of note is the differing behavior of the singleton /g/ *vs.* the complex onset /gw/: in both the soft and mixed mutations the reflex of /g/ is [x], whereas the reflex of /gw/ is [w].

(12) *Illustration of the Old Irish morphophonemes L, G, and N; base form* [kna:ˈṽʹ][7]

mo chnáim	[mo xna:ˈṽʹ]	moL kna:ṽʹ	'my bone'
do chnáim	[do xna:ˈṽʹ]	doL kna:ṽʹ	'your bone'
a chnáim	[a xna:ˈṽʹ]	aL kna:ṽʹ	'his bone'
a cnáim	[a kna:ˈṽʹ]	aC kna:ṽʹ	'her bone'
a gcnáim	[a gna:ˈṽʹ]	aN kna:ṽʹ	'their bone'

(Hamp 1951: 235)

Thus, Hamp's account, very much within the American structuralist tradition, relies on diacritic association of morphophonemes with the specific items which trigger mutation. By listing the effect of each of the morphophonemes, the correct phonetic form derives from spelling out the effects of a specific morphophoneme on a particular target, i.e. a particular initial consonant.

The emergence of generative phonology (following Chomsky and Halle 1968) led to a re-evaluation of the morphophonemic approach to initial consonant mutation. Rogers (1972), for example, explicitly rejects Hamp's morphological approach to initial consonant mutation. Instead, Rogers develops a set of *SPE*-type phonological rules to account for the alternations associated with Scottish Gaelic mutations.

Although Rogers acknowledges the morphological nature of initial mutation,[8] he views the morphophonemic approach of Hamp (1951) and Oftedal (1962) as arbitrary. Instead, he "attempts to describe the conditioning in terms of either natural phonological segments or morphological features" (1972: 64). In order to do this, Rogers develops a set of 14 distinct *SPE*-type phonological rules. In fact, these rules essentially address the second of our five questions relevant for phonology: How can the alternations be characterized phonologically? In addition to his phonological rules, Rogers also needs a morphological trigger for the specific mutation in question. For example, the rule leniting /f/ to zero includes in its context of application the morphophoneme for lenition [+Lnt] at a morpheme boundary:

(13) $f \rightarrow \emptyset$ / [+Lnt] + __

So even though Rogers's explicit goal is to take the mutations out of the morphology and put them in the phonology, the attempt is only partly successful. Although he is able to develop a set of phonological rules to account for the changes involved in initial mutation in Scottish Gaelic, the rules themselves refer crucially to morphological information. Furthermore, the "lenition feature" [+Lnt] appears to enjoy a life of its own, rather than necessarily being anchored to a lexical item, as Hamp's morphophonemes are. From the vantage point of current phonology, it appears that while Rogers avoided the American structuralist practice of listing employed by Hamp, his own account is not free of crucial dependency on morphology and thus is only a partially phonological account.

7 Hamp also includes *kna:Lmʹ* as a possible base form to show that the *L* diacritic can also appear word-internally. This form is omitted here for clarity.

8 "The fact that the possessive pronouns for 'her', 'his', and 'their' are homophonous shows the morphological conditioning of the mutations" (Rogers 1972: 63).

One area of initial consonant mutation that has drawn a great deal of syntactic attention in recent years is "syntactic soft mutation" in Welsh (see §3 above). Growing from initial work by Harlow (1989), and further developed in a number of subsequent publications, including Tallerman (1990, 2006, 2009), Borsley and Tallerman (1996), Borsley (1997, 1999), and Borsley *et al.* (2007), the environment of Welsh syntactic soft mutation is shown to be analyzable in terms of phrasal syntax. Specifically, the central idea behind this work is that a mutable initial consonant immediately preceded by a syntactic phrase, i.e. an XP, undergoes soft mutation. This is known as the XP Trigger Hypothesis, or XPTH. In the examples in (15), the XP in question is an NP, though other maximal projections also trigger the mutation (see Borsley *et al.* 2007 and Tallerman 2009).

(15) a. Dechreuodd [Huw]NP *olchi* 'r llestri (golchi)
begin.PAST.3SG Huw wash.INF the dishes
'Huw began to wash the dishes.'

b. Mae yn [yr ardd]NP *gi* (ci)
be.PRS.3SG in the garden dog
'There's a dog in the garden.'

Thus, the XPTH provides a structural account of Welsh soft mutation distinct from attempts to characterize soft mutation as marking direct object or abstract case marking (cf. Zwicky 1984; Roberts 1997, 2005) mentioned earlier. Clearly in the cases in (15) the object of mutation is not a direct object: in (15a) it's an infinitive, in (15b) it's a subject.

In terms of comparing the phonological account with the XPTH, it is difficult to decide between them. On the one hand, a phonological account of a phonological phenomenon might be preferred, yet if the conditioning environment for syntactic soft mutation really is syntactic structure – the XP rather than the Phonological Phrase – the XPTH might be preferable.

5.3 *Morphological accounts*

Recently, both Stewart (2004) and Green (2006) have returned to the argument initiated by Hamp (1951) that mutation is morphology. A difference, however, lies in the fact that Hamp's approach is morphophonological, where Stewart and Green both argue for a non-phonological morphological position. Stewart (2004) reviews the primary literature on Celtic mutation, coming back to the conclusion that the Celtic mutations are entirely morphological, not phonological. As he puts it:

> once it is admitted that the mutations are not subject to natural class behavior and that their distribution crucially must make reference to non-phonological information . . . it seems evident that what we are dealing with is not live synchronic phonology . . . but is now a set of surface correspondences that has been quite fully morphologized. (Stewart 2004: 63)

In dealing with these "quite fully morphologized" correspondences in Scottish Gaelic initial mutation, Stewart proposes three specific types of morphological domains: "intralexemic morphology," dealing with paradigms of inflected forms on the basis of stems of individual lexemes, "interlexemic morphology," and "supralexemic morphology." According to Stewart, each of the various mutations in Scottish Gaelic falls into one of these domains.

In addition to these domains of morphology, Stewart proposes "meta-redundancy-rules" to describe each mutation type, thereby defining a "Lenition constellation" and a "Nasalization constellation," a "constellation" being "the collection of rules which instantiate the meta-rule" (2004: 125). By removing non-automatic (i.e. non-phonological) alternations from the phonology and relieving the syntax of the burden of morphological marking (at least with respect to initial mutation), his stated goal is to allow for more constrained theories of phonology and syntax. Despite the interesting possibilities raised, Stewart explicitly avoids "a rigorous formalization of the concepts involved" (2004: 145).

In an examination of initial consonant mutations and the evidence they provide for phonological theory, Green (2006, 2007) argues that the mutations lie "entirely outside the phonology" (2006: 1948). Instead, Green maintains that the mutations "are properties of the lexicon, which consists . . . of whole words listed in all their actual surface forms and connected to each other through their shared properties" (2006: 1948). Thus, rather than viewing initial consonant mutation as the instantiation of some change in an initial consonant, Green takes a Bybeean view (cf. Bybee 1985, 2001), assuming that a mutating word form will belong to a set of forms related in the lexicon (see also CHAPTER 1: UNDERLYING REPRESENTATIONS). For example, Irish *bróg* [bro:g] 'shoe' will be related in the lexicon to *bhróg* [vro:g] 'shoe' (lenited) and *mbróg* [mro:g] 'shoe' (eclipsed). Crucially, the mutated forms are not derived from an underlying /bro:g/. Rather, it is proposed that various determiners, complementizers, and other proclitics in the Celtic languages subcategorize for specific mutations, thus selecting from the lexicon the appropriate shape for the lexical item in question. Recalling Lieber's (1987) prespecification analysis, Green's proposal shifts the burden from the phonological implementation of a morphological diacritic firmly to subcategorization and lexical selection.

Maintaining the position that initial consonant mutation is morphological in nature, Green (2006: 1958–1959) adduces four primary arguments against the processes (as distinct from the environments) of Celtic mutation being phonological: (i) no feature or bundle of features can produce the wide variety of effects found within a single mutation; (ii) syntactically triggered mutation requires a mutation-triggering feature, yet there is no independent evidence, apart from mutation itself, for such a feature; (iii) mutations are sometimes triggered by non-adjacent proclitics; and (iv) the variety of exceptions and irregularities in mutation is inconsistent with a phonological analysis.

Turning to theory-internal issues, Green also points out that a phonological account of Celtic mutation is incompatible with the strong version of Optimality Theory (Prince and Smolensky 2004 and a large body of subsequent work), in that the Celtic mutations do not reflect the interaction of faithfulness constraints with universal principles of markedness. In OT terms, a mutated word-form is "unfaithful" to its input, the citation form. OT predicts that an unfaithful form will be selected as optimal only if markedness is improved by the unfaithful form. In the case of the initial mutations in Celtic, however, the selection of an unfaithful, mutated word form does not improve markedness, and thus runs counter to the predictions of OT. This is taken by Green as further evidence that initial mutation is not phonological.

6 Theoretical issues informed by initial consonant mutation

Accounts of initial consonant mutation are necessarily connected with the theoretical frameworks within which they have been proposed. Nonetheless, just as phonological theory informs the analysis of phonological data, the accounts themselves may inform various theories. Apart from the questions and issues surrounding the analysis of mutation itself, whether in accounting for the changes, the triggers, the contexts, or the representations, initial consonant mutation has played an interesting role in various aspects of the development of linguistic theory. As we have seen, these include Hamp's (1951) use of mutation to argue for the role of morphological marking in linguistic analysis, Lieber's (1983) use of mutation to support autosegmental theory, Gnanadesikan's (1997) use of mutation to argue for a system of ternary feature scales in phonology, and Green's (2006) rejection of phonological mutation to argue for the strong version of Optimality Theory, i.e. that phonological processes result only from the interaction of markedness and faithfulness.

The phenomenon of initial mutation has also been used as evidence for the relationship between linguistic operations. Hannahs and Tallerman (2006), for example, use the evidence of initial mutation and its interaction with the shape of the Welsh definite article (which appears variously as *yr* [ər], *y* [ə], or *r* [r]) as a diagnostic for determining the sequence of lexical insertion in the syntax. The definite article may trigger mutation (e.g. of a feminine singular noun), thereby altering the shape of a following noun. The shape of the definite article, on the other hand, is conditioned (in part) by whether it is followed by a consonant-initial or a vowel-initial word. Recalling that the Welsh soft mutation reflex of /g/ is zero, a feminine singular noun such as *gorsaf* 'station' will appear as *orsaf* after the definite article; the definite article itself appears as *yr* [ər], since *orsaf* is vowel initial. Masculine singular *gosteg* 'silence', however, does not mutate after the definite article and so retains its initial [g].

(16)	gorsaf	'station (FEM SG)'	gosteg	'silence (MASC SG)'
	yr orsaf	'the station'	y gosteg	'the silence'
	*y gorsaf, *yr gorsaf		*yr osteg, *yr gosteg	

The examples in (16) show that in this context *gorsaf* mutates to *orsaf* and the article surfaces as *yr* [ər]. In the same context, though, masculine singular *gosteg* 'silence' does not mutate, as shown, and the article appears as *y* [ə], since it precedes a consonant. Since the shape of the noun depends on the presence of the definite article, and since the shape of the definite article depends on the initial segment of the following noun, the interaction provides evidence relative to lexical insertion and distinctions between function words and content words.

7 Conclusion

Celtic initial consonant mutation has its origin in phonetics as a series of phonetically triggered external sandhi phenomena. The overwhelming bulk of initial consonant mutation in the modern languages, however, is not transparently phonological. Rather, initial consonant mutation in the living Celtic languages – Irish, Scottish Gaelic, Breton, and Welsh – is a complex set of lexically, syntactically, morphologically, and/or structurally triggered alternations reflected in the phonology. Moreover, the alternations themselves are not phonologically unitary. As such, a full account of initial consonant mutation necessarily involves reference to a number of different levels of linguistic description and analysis.

Nonetheless, there are several phonologically relevant questions (as outlined above in §3), some of which have been addressed in various ways in the literature. For example, among the accounts considered above, questions (ii) (How can the alternations be characterized phonologically?) and (v) (How are the alternations represented phonologically?) have been addressed by a number of linguists, including Hamp (1951), Rogers (1972), Ewen (1982), Lieber (1987), Ni Chiosáin (1991), Grijzenhout (1995), Pyatt (1997), Stewart (2004), and Green (2006, 2007). Fewer have addressed the third question (Where do the alternations occur, i.e. (how) can the alternation site be defined in phonological terms?), among them Hannahs (1996) and Neeleman (2005). Question (i) (What phonological alternations are involved in the mutations?) is superficially addressed by all – in that some mention must be made of the segmental alternations involved in mutation – but few have looked at this explicitly in detail, Ní Chiosáin (1991) being the exception. Question (iv) (What triggers the alternations?) has received various answers; Hamp (1951) and Oftedal (1956), for example, attribute the trigger to a morphophoneme, while Rogers (1972) and Ni Chiosáin (1991) cite phonological rules.

Clearly, there is still work to be done to develop a comprehensive account of initial consonant mutation in Celtic. Just as clearly, the account will need to consider the grammar as a whole – syntax, morphology, and phonology – in order to provide the full picture of this intriguing but complex phenomenon.

ACKNOWLEDGMENTS

Many thanks to Anna Bosch, Maggie Tallerman, Marc van Oostendorp, Keren Rice, and two anonymous reviewers. Their comments have greatly improved this chapter.

REFERENCES

Anderson, John M. & Charles Jones. 1974. Three theses concerning phonological representations. *Journal of Linguistics* 10. 1–26.

Anderson, John M. & Charles Jones. 1977. *Phonological structure and the history of English.* Amsterdam: North-Holland.

Archangeli, Diana. 2009. The sonorants of Skye Gaelic. Paper presented at the Formal Approaches to Celtic Linguistics conference, University of Arizona.

Awbery, Gwenllian M. 1975. Welsh mutations: syntax or phonology? *Archivum Linguisticum* 6. 14–25.

Awbery, Gwenllian M. 1986. *Pembrokeshire Welsh: A phonological study.* Cardiff: National Museum of Wales.

Ball, Martin J. & Briony Williams. 2000. *Welsh phonetics.* Lampeter: Edwin Mellen Press.

Borsley, Robert D. 1997. Mutation and case in Welsh. *Canadian Journal of Linguistics* 42. 31–56.

Jones, D. M. 1949. Written communication. In *Actes du 5e congrès international des linguistes, Paris, juillet 1948*, 240–242. Paris: Klincksieck.

Lieber, Rochelle. 1983. New developments in autosegmental morphology: Consonant mutation. *Proceedings of the West Coast Conference on Formal Linguistics* 2. 165–175.

Lieber, Rochelle. 1987. *An integrated theory of autosegmental processes*. Albany: SUNY Press.

MacAulay, Donald. 1992a. The Celtic languages: An overview. In MacAulay (1992c), 1–8.

MacAulay, Donald. 1992b. The Scottish Gaelic language. In MacAulay (1992c), 137–248.

MacAulay, Donald. 1992c. *The Celtic languages*. Cambridge: Cambridge University Press.

Mac Eoin, Gearóid. 1993. Irish. In Martin J. Ball & James Fife (eds.) *The Celtic languages*, 101–144. London: Routledge.

Martinet, André. 1952. Celtic lenition and Western Romance consonants. *Language* 28. 192–217.

Morris-Jones, John. 1913. *A Welsh grammar: Historical and comparative*. Oxford: Oxford University Press.

Neeleman, Ad. 2005. Soft mutation at the interface. In Hans Broekhuis, Norbert Corver, Riny Huybregts, Ursula Kleinhenz & Jan Koster (eds.) *Organizing grammar: Linguistic studies in honor of Henk van Riemsdijk*, 447–457. Berlin & New York: Mouton de Gruyter.

Nespor, Marina & Irene Vogel. 1982. Prosodic domains of external sandhi rules. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part I, 225–255. Dordrecht: Foris.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Ní Chiosáin, Máire. 1991. Topics in the phonology of Irish. Ph.D. dissertation, University of Massachusetts, Amherst.

Ó Dochartaigh, Cathair. 1979. Lenition and dependency phonology. *Éigse* 17. 457–494.

Oftedal, Magne. 1956. *A linguistic survey of the Gaelic dialects of Scotland*, vol. 3: *The Gaelic of Leurbost, Isle of Lewis*. Oslo: Aschehoug.

Oftedal, Magne. 1962. A morphemic evaluation of the Celtic initial mutations. *Lochlann* 2. 93–102.

Pedersen, Holger. 1897. *Aspirationen i Irsk*. Copenhagen: Spirgatis.

Press, Ian. 1986. *A grammar of modern Breton*. Berlin: Mouton de Gruyter.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Pyatt, Elizabeth J. 1997. An integrated model of the syntax and phonology of Celtic mutation. Ph.D. dissertation, Harvard University.

Rice, Keren. 1993. A re-examination of the feature [sonorant]: The status of "sonorant obstruents." *Language* 69. 308–344.

Roberts, Ian G. 1997. The syntax of direct object mutation in Welsh. *Canadian Journal of Linguistics* 42. 141–168.

Roberts, Ian G. 2005. *Principles and parameters in a VSO language: A case study in Welsh*. Oxford: Oxford University Press.

Rogers, Henry. 1972. The initial mutations in modern Scots Gaelic. *Studia Celtica* 7. 63–85.

Selkirk, Elisabeth. 1980. Prosodic domains in phonology: Sanskrit revisited. In Mark Aronoff & Mary-Louise Kean (eds.) *Juncture*, 107–129. Saratoga: Anma Libri.

Selkirk, Elisabeth. 1981. On prosodic structure and its relation to syntactic structure. In Thorstein Fretheim (ed.) *Nordic Prosody II*, 111–140. Trondheim: Tapir.

Steriade, Donca. 1993. Closure, release, and nasal contours. In Marie K. Huffman & Rena A. Krakow (eds.) *Nasals, nasalization, and the velum*, 401–470. Orlando: Academic Press.

Stewart, Thomas W. 2004. Mutation as morphology: Bases, stems, and shapes in Scottish Gaelic. Ph.D. dissertation, Ohio State University.

Tallerman, Maggie. 1987. Mutation and the syntactic structure of modern colloquial Welsh. Ph.D. dissertation, University of Hull.

Tallerman, Maggie. 1990. VSO word order and consonantal mutation in Welsh. *Linguistics* 28. 389–416.

Tallerman, Maggie. 2006. The syntax of Welsh "direct object mutation" revisited. *Lingua* 116. 1750–1776.

Tallerman, Maggie. 2009. Phrase structure vs. dependency: The analysis of Welsh syntactic soft mutation. *Journal of Linguistics* 45. 167–201.

Zwicky, Arnold M. 1984. Welsh soft mutation and the case of object NPs. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 20. 387–402.

118 Turkish Vowel Harmony

Barış Kabak

1 Introduction

Turkish vowel harmony is a cover term for a set of well-formedness conditions that dictate which vowels can co-occur within what we will simply call a non-compounded word, irrespective of its morphosyntactic complexity. This chapter aims to bring forth various issues surrounding the mechanism that realizes these vowel co-occurrence conditions in Turkish, and discuss the ways in which they contribute to the understanding of the nature and dynamics of phonological representations and processes. We will explore Turkish vowel harmony in terms of three sets of issues, each of which has proved highly instructive for theory building and testing, and provoked decade-long debates in the field of phonology. These are: (i) what is it that vowel harmony targets and what are the surface effects that it achieves?, (ii) how does it propagate in roots and suffixes?, and (iii) how can vowel harmony be obstructed, and what are the consequences of this for the nature of phonological representations? We will touch upon relevant theories and proposals in the literature in so far as they shed light on the interpretation of empirical facts, and also introduce new data and insights where necessary.

2 The substance of Turkish vowel harmony

Vowel harmony could be defined in at least two different ways, with fundamentally different assumptions. First, along the lines of the structuralist tradition (e.g. Bloomfield 1933), one can approach it through the surface effects it achieves, forming a set of observations about a pattern where an individual vowel may only co-occur with certain vowels, hence, by inference, the co-occurrence of the same vowel with certain others is assumed to yield ill-formed structures. In this sense, vowel harmony is only a set of vowel co-occurrence (phonotactic) constraints, the motivation of which may not be obvious. Vowel harmony could also be described as an operation whose purpose is to achieve some form of "identity" among certain types of vowels that co-occur within a word. In this sense, vowel harmony is viewed as a process that has the power to actively manipulate structures with the aim of neutralizing vowels on its way that meet its criteria. The crucial ingredients of this approach to vowel harmony are then a trigger, a target, and a context. Below, we will use both approaches to describe the two different types of vowel harmony patterns observed in Modern Turkish.

2.1 Front–back harmony

With perfectly symmetrical front–back, high–low, and round–unround oppositions, Turkish has eight phonemic vowels. (1) displays the vowel phonemes of Turkish and their phonological characterization based on features that are largely agreed upon.[1]

(1) *Turkish vowel phonemes*

	front		back	
	unround	round	unround	round
high	i	y	ɯ	u
low	e	ø	ɑ	o

Within a non-compound word, including both the root and following suffixes, vowels from the set /ɑ ɯ o u/ (i.e. "back" vowels) cannot combine with those from the set /e i ø y/ (i.e. "front" vowels).[2] As such, forms like */dɑniz/ and */demuɪz/ are ruled out, but /deniz/ 'sea' and /dɑmɯz/ 'animal shed' are predicted to be well-formed. As early as Trubetzkoy (1939), and later on in the works of Jakobson (1941) and Jakobson *et al.* (1952), such surface co-occurrence patterns in Turkish vowels played a significant role in motivating "phonological contrasts" or "oppositions" as valid notions for phonological operations instead of individual phonemes themselves (See CHAPTER 17: DISTINCTIVE FEATURES). The set of vowels is not a random set of individual phonemes; rather, it forms a phonetically motivated set of oppositions: front vowels as opposed to back ones, not, say, /e/ as opposed to /ɑ/. Each distinct vowel is then viewed as a bundle of features, which is composed of features characterizing the frontness, height, and roundness of the vowel in question (/e/: "front," "low," "unround"); /ɑ/: "back," "low," "unround"). That /ɑ/ co-occurs with /o/, as well as /ɯ/ and /u/ on the surface, is not random, then, but principled: it falls out straightforwardly from the nature of the feature that the vowels in question share, namely back.

In Turkish, all eight vowels can occupy the initial syllable of a root, while the quality of the vowels in non-initial syllables is constrained by the initial vowel. Accordingly, front–back harmony (also known as palatal harmony) is viewed as a process whereby the frontness/backness (henceforth frontness) of a non-initial root vowel or any subsequent suffix vowel (with some exceptions) is obligatorily determined by the preceding vowel. If that vowel is back, the following vowel,

[1] Instead of adhering to the conventions of Turkish orthography, we will use IPA symbols throughout the chapter, ignoring allophonic variation (e.g. palatalization in the context of front vowels). The letter known as "soft g" (ğ) is represented as a velar glide (/ɰ/), which may be realized in careful speech but is generally deleted in casual speech in standard Istanbul Turkish, the variety of the author and most of the data discussed in this chapter. See Kabak (2007: 1381) and the references therein for details on the phonological behavior of this phoneme.

[2] This does not entail that vowels from a single set can freely combine with one another, due to the effects of rounding harmony discussed below.

and consequently all subsequent vowels within a word, must be back. Likewise, if an initial vowel is front, then all the subsequent vowels are front. This is demonstrated in (2) by polysyllabic morphologically simplex roots which meet the description of vowel harmony – what is often referred to as "internal/root harmony" – and by two regular Turkish suffixes in (3) – "external/suffiix harmony." The suffixes in question are the plural and the genitive, whose height features are invariable (i.e. they contain a non-high and a high vowel, respectively), and they are attached to the roots /dɑl/ 'branch' and /jer/ 'place', containing an invariable back and a front vowel, respectively. Compounded words do not constitute a single vowel harmony domain, as indicated by the failure of the second member of the compound to undergo harmony with the first, or vice versa in (3c, 3d).

(2) *Front-back harmony in roots* (morphologically simplex words)

a.	jɑnlɯʃ	'wrong'	*jɑnliʃ, *jenlɯʃ
b.	josun	'moss'	*josyn, *jøsun
c.	zengin	'rich'	*zengɯn, *zɑngin
d.	kømyr	'coal'	*komyr, *kømur

(3) *Front-back harmony in morphologically complex non-compounded words*

a.	dɑl-lɑr-ɯn	'of branches'	*dɑl-ler-in, *dɑl-ler-ɯn,*dɑl-lɑr-in, *dɑl-ler-ɯn
b.	jer-ler-in	'of places'	*jer-lɑr-ɯn, *jer-lɑr-in, *jer-ler-ɯn, *jer-lɑr-in
c.	ɑk#deniz	'the Mediterranean'	*ɑkdɑnɯz, *ekdeniz
d.	ɯzgɑrɑ#køfte	'grilled meatball'	*ɯzgɑrɑkoftɑ, *izgereкøfte

2.2 *Rounding harmony*

Apart from front–back harmony, the rounding of non-initial vowels within a word must also agree with that of the preceding vowel. However, this condition only holds for high vowels (irrespective of the height of the preceding vowel), low vowels remaining invariably unrounded unless they are underlyingly round. Rounding harmony is illustrated in (4) with roots (see also examples (2b, 2d) above), and in (5) with the genitive suffix that contains a high vowel attached to roots with a round vowel.

(4) *Rounding harmony in roots* (morphologically simplex words)

a.	torun	'grandchild'	*torɯn, *tɑrun
b.	kuru	'dry'	*kurɯ, *kɯru
c.	køty	'bad'	*køti, *kety
d.	dyɰyn	'wedding'	*dyɰin, *diɰyn

(5) *Rounding harmony in morphologically complex words*

a.	jol-un	'of a/the road'	*jol-ɯn
b.	søz-yn	'of a/the word'	*søz-in
c.	buz-un	'of a/the ice'	*buz-ɯn
d.	jyz-yn	'of a/the face'	*jyz-in

It is not the case that the genitive suffix in (5) must surface with a round vowel. The same suffix appears with a non-round high vowel following a non-round root vowel (6a), or a non-round suffix vowel (6b).

(6) a. jɑʃ-ɯn 'of an/the age' *jɑʃ-un
age-GEN
b. jol-lɑr-ɯn 'of (the) roads' *jol-lɑr-un, *jol-lor-un, *jol-lor-ɯn
road-PL-GEN

Despite the fact that it appears after a rounded vowel, the plural morpheme in (6b) obligatorily surfaces as a non-rounded vowel (i.e. */lor/), since it is low, demonstrating that only a non-initial high vowel, be it a root or a suffix vowel, can undergo rounding harmony.[3] Furthermore, we see that the roundness of the first vowel does not determine the roundness quality of the vowel of the genitive suffix, although it is high. Instead, whether the vowel in question will be round or unround is determined by the preceding vowel, a point we will come back to below.

For completeness, it should be noted that roundness and frontness do not always come from the same segment. This is observed in cases where certain root-final consonants, most notably the palatal lateral, are assumed to determine the frontness of the suffix vowel while the rounding property comes from the preceding vowel (e.g. /petroʎ-dy/ (petrol-PAST) 'it was petrol'), which we will discuss in more detail in §4.

In summary, both harmony processes impose a number of restrictions on the phonological shape of Turkish words, and hence entail a number of *expected* patterns and gaps:[4]

(7) a. All vowels must agree in terms of frontness or backness.
b. High vowels must also agree in roundness with the immediately preceding vowel, hence no high rounded vowels can be found after non-round vowels within a word.
c. No low round vowels (i.e. [o] and [ø]) may be present in a non-initial syllable of a word.[5]

3 The mechanism of Turkish vowel harmony

3.1 *Earlier treatments*

So far, we have demonstrated that vowel harmony in Turkish is an agreement operation within an extended domain, such as a morphologically complex word. What is not obvious, however, is the way in which a particular harmony feature

[3] See Charette and Göksel (1996, 1998) for an analysis that accounts for the blocking of rounding harmony on low vowels by the use of licensing constraints that are also argued to derive the vowel inventory of Turkish.

[4] Note that there are several words in the Turkish lexicon, mostly borrowings, that violate these patterns. These are known as disharmonic words, which will be discussed in §4. Despite the presence of disharmonic words, Kabak *et al.* (2010) show a facilitatory effect of vowel disharmony in Turkish (but not in French) in signaling word boundaries, suggesting that Turkish listeners can exploit vowel harmony regularities in speech segmentation.

[5] Low round vowels are allowed to surface in non-initial syllables if they are underlyingly present, as in some loanwords, e.g. [kɑmjon] 'truck', [fritøz] 'deep fryer', or in native roots due to the presence of a labial consonant preceding the vowel, a phenomenon known as the Labial Attraction Rule (e.g. /ʧɑbuk/ 'quick'; see §4 for details).

(i) the rounding value, just like the frontness of non-initial vowels, is predicted to originate from some source on the left, and (ii) rounding can be realized only on high vowels in this position. This, in and of itself, speaks for the highly special character of this suffix compared to other exceptional suffixes, and furthermore suggests that it comes from a once-independent lexical item, which Turkish vowel harmony has not been pervasive enough to domesticate (see Kabak and Revithiadou 2009 and the references therein for the emergence of similar suffix complexes). Due to its invariability, the second vowel of this suffix is arguably fully specified for its phonological content, blocking any phonological operations from manipulating it. Nevertheless, the variable portion of the suffix, namely the initial high vowel, does not get its harmonic properties from its right neighbor, but from what precedes. The apparent mono-directionality here could also be explained by notions such as "root control," which refers to a widely attested principle that root vowels have an effect on suffix vowels, rather than the other way around. Notice that while the root-control analysis also straightforwardly accounts for the so-called "epenthetically driven vowel harmony," a *regressive* harmony process observed in roots where the quality of epenthetic vowels is determined by the following vowel, the mono-directionality analysis encounters problems with these cases. However, as we will see in §5 below, this sort of harmony runs on different principles than regular vowel harmony. As such, there are reasons to keep the two processes distinct, and hence continue to regard Turkish vowel harmony proper as a strictly progressive assimilation process.

These observations have provoked skepticism about the very nature of vowel harmony and raised questions as to whether it operates by means of a fundamentally different principle than other assimilatory processes such as umlaut (see Anderson 1980 for a thorough discussion). Indeed, a famous "assimilation" treatment to vowel harmony was proposed in Lees (1961), where the first vowel of a root, being fully specified for frontness and roundness, acts as a trigger by the use of linear rules of the type made prominent in Chomsky and Halle (1968; henceforth *SPE*). These rules ensure that those vowels unspecified for the same features are filled in iteratively. Similar iterative rules, though only reserved for suffix harmony, have been put forth in the works of Kardestuncer (1982, 1983). A simplified version of a linear vowel harmony rule, taken from van der Hulst and van de Weijer (1991), is illustrated in (9).

(9) SPE-*type linear rules of vowel harmony*

a. *Front–back harmony*
$V \rightarrow [+\text{back}] \ / \ [V_{+\text{back}}]\ C_0 \_\_$

b. *Rounding harmony*
$[V_{+\text{high}}] \rightarrow [+\text{round}] \ / \ [V_{+\text{round}}]\ C_0 \_\_$

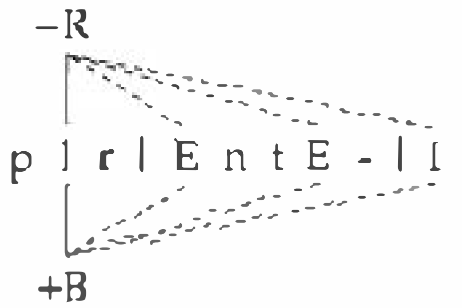
This sort of treatment handles all cases of harmony in Turkish, including the problematic ones for the root marker approach in a straightforward way. As for disharmonic roots, all that has to be stipulated is that the vowels that do not undergo harmony are marked with a diacritic, such as [–Vowel Harmony], which blocks the effect of the preceding root vowels.

SPE-type rule notations have been rejected on empirical and conceptual grounds, which significantly weakened the liability of linear rules for explaining phonological phenomena in the field. In the case of Turkish vowel harmony, however, the major objection to the type of linear rule-based treatments came from observations that it does not achieve unity with other harmonic systems across the languages of the world, where vowel harmony is bidirectional and extend to both prefiixes and suffiixes (Crothers and Shibatani 1980). This problem is trivial in the case of Turkish, where one can assume that the directionality is determined on a language-specific basis.

The core of the problem that Crothers and Shibatani (1980) raised against linear rules had rather to do with the nature of harmony regularities and irregularities in roots, and their incompatibility with rules like (9) above that crucially depict assimilation. Essentially, there is no observable assimilation in root vowels, which suggests that front–back harmony in (2) and rounding harmony in (4) above constitute at best "static" regularities, hence assimilation-type rules cannot account for them. We will discuss harmony in roots in §4 below.

Furthermore, what linear rules failed to capture was in fact more than the issue of directionality. It was rather the autonomy of specific vocalic features (backness, roundness) from other vocalic features, and most importantly from consonantal features. Turkish vowel harmony was important for another theory brewing at the time, Autosegmental Phonology, influenced largely by the Firthian school of Prosodic Analysis, but more elaborate about the way and the reason vowel features can or cannot propagate within a word (see CHAPTER 14: AUTOSEGMENTS). In this approach, which is thoroughly extended to Turkish vowel harmony in the seminal work of Clements and Sezer (1982), distinctive features are believed to be present on independent tiers specifying, among others, the frontness, height, and rounding properties of segments. Harmonizing root and suffix vowels are left unspecified. This can be observed in (10), where –R and +B stand for non-round and back features respectively that are prespecified on the initial vowel of the root.

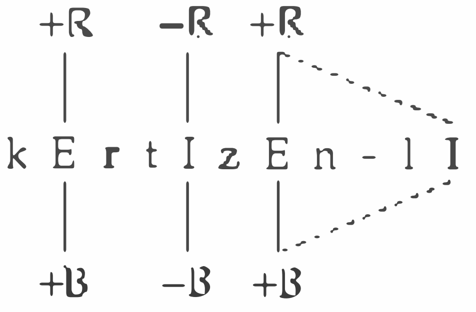
(10) *Autosegmental representation of front–back and rounding harmony*
(from Kabak and Vogel 2010)



[pɯrlanta-lɯ] 'with brilliants'

The subsequent root and suffix vowels are not specified for these two features, as indicated by archiphonemic representations such as E and I. As such, vowel harmony is described as the association of these feature specifications to different vowels lacking the same specifications within an extended domain. Whenever harmony is blocked, be it in roots or suffixes, autosegmental analyses resort to prespecification to indicate the invariable harmonic feature. Since association lines are not allowed to cross, only the rightmost prespecified feature in disharmonic roots prevails, thereby opening up its own harmony domain for the suffixes that follow (11).

(11) *Lexical prespecification of disharmonic root vowels*
(from Kabak and Vogel 2010)



[kortizon-lu] 'with cortisone'

While straightforwardly capturing the autonomy of vowel features, and hence making phonological operations less abstract, autosegmental approaches also have to make a number of assumptions about the precise shape of phonological representations, without which the theory would fail to explain the spreading of harmony features, their blocking, and the inactivity of vowel harmony operations in the face of static harmony regularities in roots, which we turn to below.

4 Vowel harmony and the shape of phonological representations

4.1 *Separating root (dis)harmony from suffix harmony*

Despite the overwhelming number of harmonic roots in Turkish,[7] no vowel *alternation* can be observed in them. In other words, front–back harmony in (2) and rounding harmony in (4) above are nothing more than static surface regularities, hence assimilation-type rules such as the ones given in (9), as well as autosegmental spreading-type analyses we have seen above, cannot properly account for them. However, the constraints on vowel co-occurrences in nearly all roots of native origin are almost the same as those in suffixes in Turkish. As such, we are dealing with one single phenomenon, realized, however, in two different ways: (i) harmony as a co-occurrence restriction, and (ii) harmony as an active phonological process. To remedy this problem, Kiparsky (1973), in his "two-pronged" analysis of vowel harmony, suggested that vowel harmony on the surface must be handled by two different mechanisms, one reserved for purely static harmony patterns in roots/stems, and the other for suffix harmony that is purely assimilatory in nature. While the earlier is handled by Morpheme Structure Constraints (henceforth MSCs; see also CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS), the other is the job of phonological rules.

Three such MSCs are proposed by Zimmer (1969b) for Turkish roots, two of which are essentially the same as the front–back and rounding harmony for suffixes (the *palatality* MSC and the *labiality* MSC, respectively). The third MSC, the *labial consonant* MSC (also known as the Labial Attraction Rule; Lees 1966), is specifically reserved for *disharmonic* occurrences of round vowels after non-round vowels in roots: a high vowel following an /ɑ/ with an intervening labial consonant, /p b f v m/, will be round (e.g. /ɑvutʃ/ 'palm of hand', /tʃɑbuk/ 'quick').

[7] Kabak *et al.* (2008), based on the first two vowel occurences of a carefully selected 8944 set of polysyllabic roots (morphologically simplex) from the TELL database, reports that 29 percent of all such roots are disharmonic with respect to front–back harmony.

Notice that this condition does not apply across morpheme boundaries: /av-ɯ/ (hunt-ACC), not */av-u/, suggesting that MSCs are not always the replication of active phonological rules.[8]

The separation of root harmony from suffix harmony in Turkish fits well with the Derived Environment Condition, developed most notably in Kiparsky (1973) (see CHAPTER 88: DERIVED ENVIRONMENT EFFECTS). This condition essentially blocks neutralization processes, vowel harmony being one of them, to take place in non-derived (morphologically simplex) environments, and reserves such rules to only apply in derived environments.[9] The details of such an analysis do not concern us here. Two empirical and theoretical issues that fall out from such a differential treatment of root and suffix harmony, however, must be highlighted, since they yield serious consequences for the nature of lexical representations and the way phonological processes operate.

First, it entails that vowel harmony must be "inactive" in roots, and hence automatically raises questions with respect to the lexical representation of Turkish roots. Are root vowels fully specified despite harmony? How can our representational device insightfully separate disharmonic roots from harmonic ones? What are the psycholinguistic implications of this separation for the language learner and user?

Second, it fails to capture the formal similarity and functional redundancy between pervasive harmony patterns (handled by MSCs) and phonological rules, which is known to be the "Duplication Problem" in the literature, most notably addressed by Kenstowicz and Kisseberth (1977). As McCarthy puts it, "if MSCs and rules really are distinct components of linguistic theory, then they should be clearly differentiated in form and function, but they are not" (1998: 123). The question here is how can we combine seemingly two, but essentially the same, phenomena in a single grammar? We will unpack these issues in the context of root harmony and disharmony below, which have also received a great amount of attention in recent history of phonological theorizing, most notably in Optimality Theory, where the duplication problem is rendered irrelevant (see below).

4.2 *The limits of harmony and disharmony in roots*

From a scientific point of view, investigating harmony is not just to account for what is possible as a harmonic sequence, but also for what is possible as a disharmonic sequence. Any theory should hence be able to explain both harmonic and disharmonic patterns, ideally using the very same apparatus.[10] This has been the prevalent force behind much work done in the last two decades. Various restrictions on disharmony are stated in a seminal paper by Clements and Sezer (1982). Their observation is that the vowels from the set /i e ɑ o u/ may combine freely in roots, leading to the violation of both types of harmony. This is illustrated in (12), where all examples violate front–back harmony, and additionally rounding harmony in (12b) and (12c), where the non-initial vowel is expected to be round

[8] There are several exceptions to the Labial Attraction Rule in Turkish and the rule does not act as magnet for analogy (see Inkelas *et al.* 1997 for details).

[9] It should be noted that, rather than a structure-changing process, if vowel harmony is viewed as a feature-filling process, it cannot be blocked in underived environments.

[10] The source of disharmony in roots is primarily and predominantly loanwords, which are well integrated into Turkish, and perhaps except for a few recently borrowed items, an average Turkish speaker may not even know that they are foreign.

since it is high, and in (12d), where the non-initial vowel is round despite the fact that it is non-high (i.e. there is no harmony-related reason why rounding should appear on the second vowel).

(12) *Root disharmony: /*i e ɑ o u*/ are freely combined*

			Type of violation
a.	gɑrip	'strange'	Front–back harmony
b.	polis	'police'	Front–back and rounding harmony
c.	butik	'boutique'	Front–back and rounding harmony
d.	pilot	'pilot'	Front–back and rounding harmony

On the contrary, vowels from the set /y ø ɨ/ are shown not to occur in disharmonic roots, even though this observation is not without exceptions (e.g. /byro/ 'office'). Goldsmith (1990: 304–309) argues that /i e ɑ o u/ is in fact not a random set of vowels, but coincides with the cross-linguistically favored five-vowel system. In such a system, the specification for the frontness and backness (i.e. [back] according to Goldsmith) is redundant since it is fully predictable from rounding. As such, [back] is underspecified in roots that combine these five vowels, consequently leading to no violation of front–back harmony. The specification of [back] is required, however, when the root contains any of the other three vowels, /y ø ɨ/, since they involve marked combinations of frontness and rounding. This argument finds further support from Clements and Sezer's (1982) documentation of harmony regularizations on disharmonic loanwords, where the same set of vowels tends to be changed to fit at least into the five-vowel system, as can be seen in (13) below. Note that this process does not necessarily turn disharmonic roots into harmonic ones (e.g. (13e)).

(13) *Regularization of /*y ø ɨ*/ in disharmonic roots*

a.	mersørize	~	merserize	'mercerized'	
b.	kupyr	~	kypyr	'clipping'	(also kipyr)
c.	biskyvit	~	byskyvyt	'biscuit'	(also bisky:t/bysky:t)
d.	pyro	~	puro	'cigar'	
e.	komynist	~	kominist	'communist'	(also komunist)
f.	motør	~	motor	'motor'	
g.	ʃøvɑlje	~	ʃovɑlje	'knight'	

Yet another observation from Clements and Sezer (1982) is that while /ɯ/ does not occur in roots which are disharmonic with respect to rounding, the other two marked vowels, /ø/ and /y/ do, as can be seen in (14). However, it should be noted that disharmonic sequences containing /ø/ and /y/ also undergo regularization of the type shown above, as indicated in (14), suggesting that rounding harmony may be equally effective on /ø/ and /y/.

(14) */ø/ and /y/ in disharmonic roots*

a.	virys	~	vyrys	'virus'
b.	myhim	~	myhym	'important'
c.	døviz	~	døvyz	'foreign currency'
d.	minør	~	mynør	'minor'

A number of accounts have been proposed to handle such patterns, mostly on the basis of Clements and Sezer's (1982) observations in different frameworks (e.g. van der Hulst and van de Weijer 1991, and for optimality-theoretic treatments, see Kirchner 1993, Polgárdi 1999, and Kiparsky and Pajusalu 2003). Although the details of these accounts are not of concern here, what should be pointed out is that accounts couched within the framework of Optimality Theory are uniformly capable of accounting for both root and suffix harmony, whereby the duplication problem discussed above becomes a moot point. More specifically, the fact that harmony is optimal regardless of the marked or unmarked status of the vowels in a sequence and that disharmony is tolerable if it comes from unmarked vowels (*à la* Clements and Sezer 1982) are handled within the same grammar by using a combination of (i) featural markedness constraints, essentially banning marked segments from surfacing (e.g. *ɯ, *y, *ø), (ii) a constraint that achieves harmony (e.g. AGREE(F), HARMONY), and (iii) some faithfulness constraint that ensures that the fully specified vowels are realized the way they underlyingly are (e.g. IDENTROOT(F)).

These proposals, however, must be given a serious reconsideration in light of data that have not been considered so far. Here I present a set of data that will amount to the observation that any vowel combination may become legitimate in Turkish as long as the donor language permits them and may remain stable despite overwhelming vowel harmony regularity in the language. Let us start with the sequence /y–ɑ/ (or vice versa), which is not expected to be tolerated. Closer inspection, however, reveals that they are surprisingly common in Modern Turkish. In fact, they constitute the seventh most common *disharmonic* sequence in a recent corpus study (Kabak *et al.* 2008). Some words of this sort are illustrated in (15). Notice that they are highly stable disharmonic patterns, exhibiting no regularization or variation of the sort illustrated in (13) and (14).

(15) /y–ɑ/ *combinations in some common disharmonic roots*

a.	kyrtɑj	'abortion'
b.	syrɑt	'speed'
c.	mɑːmyl	'product'
d.	kɑktys	'cactus'
e.	ɑky	'battery'

As for /ɯ/, it should be remembered that the claim has been that, first of all, it does not occur in front–back disharmonic roots since it is from the marked vowel set, and is the only one from that set that cannot escape rounding harmony. If we consider that disharmonic words come primarily from other languages into Turkish, the chances of finding /ɯ/ in such combinations should be remarkably low, since many languages from which Turkish imported its foreign lexical stock do not contain such a vowel. Furthermore, there is no reason to assume that /ɯ/ is the most marked vowel of all, as is assumed via high-ranked constraints such as *ɯ in a number of optimality-theoretic analyses of harmony (e.g. Kirchner 1993; Kiparsky and Pajusalu 2003), since the same vowel emerges to repair syllable structure violations, disproving its alleged "marked" status in the language. Aside from these external and internal factors, what is perhaps more important to note is that

(17) */ɯ/ substitution going against lexical diffusion and regularization*

a. kɯrlent 'pillow' from *ghirlanda* (Italian)
b. kɯdem 'rank' from *kidem* (Arabic)
c. kɯble 'mecca' from *kible* (Arabic)

Furthermore, the English schwa in borrowings is almost always mapped onto /ɯ/, as evinced by the fact that the subsequent suffix is realized as back. This can give rise to perfectly acceptable front vowels co-occurring with, presumably a non-epenthetic /ɯ/, as illustrated in (18). Since these are recent borrowings, time will show whether they will prevail as disharmonic.

(18) *English schwa is mapped onto Turkish /ɯ/, yielding disharmony*

a. [pirintɯr-a] 'to the printer'
printer-DAT
b. [bilendɯr-dan] 'from the blender' (also as [bilendir], but rarely)
blander-ABL
c. [kɯlintɯn-lar] 'the Clintons'
Clinton-PL
d. [tikɯt-lar] 'tickets' (especially in the context of 'ticket restaurants')
ticket-PL

The same generalization can be made in the context of /ø/. It is arguably not a common vowel in languages which Turkish has borrowed from, except for some French borrowings, where we see their frequent occurrence in disharmonic sequences (19).

(19) */ø/ in disharmonic words of French origin*

a. røtar 'delay' (sometimes also as rotar)
b. dansøz 'belly dancer'
c. firitøz 'deep frier'
d. ordøvr 'hors d'œuvre' (sometimes also as ordevr, ordev)
e. rømork 'trailer' (sometimes also as romørk)

Notice that /ɯ/ in combination with /ø/ and /y/ (as opposed to /i/, /e/, or /o/) is not attested. However, the fact that there are such lexical gaps should not be readily taken to mean that there are constraints, dormant or active, that militate against such sequences. A more plausible explanation lies in the unlikelihood of donor languages that have /ɯ/, /ø/, and /y/ or contain words where these vowels co-occur.

The above data cast doubts on models that try to account for disharmony by compartmentalizing vowels into different sets based on notions such as markedness. Furthermore, they suggest that we may need to change our approach toward disharmony: any combination of vowels seems to be legitimate as long as the donor languages permit them, whose likely adaptation may additionally but not necessarily be determined by how often the item is used and the acoustic/phonetic properties of the vowel in question. Variation is perhaps the most challenging phenomenon to account for in regularization, which makes it apparent which emerges when no lexically specified feature is able to fill in an unspecified feature.[12] More research is necessary here.

4.4 *Is vowel harmony active in roots?*

This is perhaps one of the most controversial issues in Turkish vowel harmony, the answer to which has had serious consequences for the shape of phonological representations in the lexicon, and consequently the whole mechanism for handling harmony.[13] Notice that the same question is equally valid in the context of disharmonic roots since their shape has also been shown to be determined, at least partially, by the dictates of vowel harmony. As stated before, no alternation could be detected in root vowels; hence, there may be no reason to assume harmony operations there. This is perhaps the most important reason to resort to full specification of vowel features in roots. Indeed, the full specification of root vowels is essentially the approach taken in some treatments couched within Optimality Theory, where given two competing input forms, one fully specified (i.e. prespecified as containing a particular harmony property), and one fully or partially underspecified, Lexicon Optimization (Prince and Smolensky 1993) will ensure that the fully specified alternative will be preferred (see Inkelas 1995 and Kabak and Vogel 2010 for discussion; also CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION). On the other hand, resorting to underspecification for harmonic roots,[14] and hence assuming that harmony is still in operation there, typically stem from the following reasons.

The first has to do with issues such as economy and simplicity in representations, and keeping redundancies to a minimum. Vowel harmony is then regarded as an economical device since it gives harmonic vowels a "free ride" in roots (e.g. van der Hulst and van de Weijer 1991). Second, vowel harmony can be argued to be active in roots since it seems to impose certain restrictions on the type of disharmonic vowel combinations, as shown in (13). Notice, however, that regularizations do not always achieve optimal harmonic sequences. Rather, they seem to be dictated by the five-vowel set, which are the five most frequent vowels of Turkish. The fact that there are many disharmonic roots that are well integrated into the Turkish lexicon, as indicated for instance by their resistance to harmonization due to their high frequency of occurrence, further suggests that regularization is not necessarily a strong indicator of harmony governing the underlying shape of Turkish roots.

12 Here it could be assumed that the relevant feature is blocked from spreading via truncation, as in tonal phenomena. That is, while the relevant harmonic feature is pronounced on the vowel that it is associated with, it is not allowed to spread onto the following vowels. When the backness feature is blocked from further spreading, the underspecified suffix vowels surface as front (Coronal), which is argued to be the default place feature for Turkish vowels (see Kabak 2007 and Kabak and Vogel 2010 for details).

13 Some of the treatments of Turkish vowel harmony with some form of full specification can be found in the work of, for example, Kiparsky (1973), Clements and Sezer (1982), Kirchner (1993), Inkelas (1995), Inkelas *et al.* (1997), and Polgárdi (1999). Analyses that assume underspecification of harmonic root vowels are presented in van der Hulst and van de Weijer (1991), Lahiri (2000), Harrison and Kaun (2001), Kabak (2007), and Kabak and Vogel (2010).

14 See Kabak (2007) and Kabak and Vogel (2010) for Maximal Underspecification, which assumes the presence of underspecified vowels also in disharmonic roots.

following vowel. The roundness triggering vowel is typically a high one (23a)–(23c), as suggested by the unacceptability of rounded epenthetic vowels before low round vowels in (23d)–(23e). There are cases, however, where a low vowel also triggers rounding albeit with perhaps additional impact from the rounding of the first consonant (23f)–(23g).

(23) *Epenthesis-driven regressive rounding harmony*

a.	gurup	'group'	
b.	kulyp	'club'	
c.	fylyt	'flute'	
d.	sɯpor	'sport'	*supor
e.	sɯmokin	'smoking'	*sumokin
f.	fylørt	'flirt'	(also as filørt)
g.	buroʃyr	'brochure'	(also as bɯroʃyr)

On the basis of about 100 loanwords containing word-initial consonant clusters, Kaun (1999) reports that Turkish speakers consistently provide rounded epenthetic vowels when the trigger is high round. When it is non-high, however, Kaun finds a lot of individual variation, but there is a stronger tendency toward regressive rounding harmony when the low trigger vowel is front than when it is back (see also Yavaş 1980c). Kaun (1999, 2004) takes these observations to indicate that, although epenthesis-driven vowel harmony does not much resemble native vowel harmony processes, it follows cross-linguistic patterns of rounding harmony. More specifically, cross-height rounding harmony is avoided in harmony systems that employ some form of rounding agreement among vowels. This may be taken as an explanation for why rounding harmony with epenthetic vowels, which are always high, is more prevalent when the trigger is also high. Furthermore, typological inspection suggests that rounding harmony systems typically prefer front rounded triggers, explaining why front low vowels tend to trigger the rounding of the epenthetic vowels more than back low vowels. The reader is referred to Kaun (2004) for various phonetic and perceptual explanations for such typological tendencies in rounding harmony.

Can we derive the properties of epenthesis-driven vowel harmony from the general properties of native vowel harmony processes? As demonstrated in (16) above, the regressive harmony process is highly sensitive to the quality of the consonants in the consonant cluster: after a velar consonant such as /k/ and /g/, the epenthetic vowel is back even in the context of a following front vowel (e.g. [kɯrem] 'cream', [kɯristal] 'crystal', [kulyp] 'club', [gɯri] 'gray', [gɯlikoz] 'glucose', etc.), yielding disharmony. This is unlike native vowel harmony, whereby consonants, except often for root-final underlying palatals (see (20) above; cf. (21)), are impartial to vowel harmony. Furthermore, the involvement of the height and frontness properties of the *trigger* vowel in epenthetically driven rounding harmony crucially deviates from the native rounding harmony process, where the tongue height specification of the *target*, not the trigger, is crucially imposed. When coupled with a great deal of individual variation, which we do not observe in native harmony patterns (at least within the same dialect), all of the above properties suggest that regressive harmony as realized on word-initial epenthetic vowels is not yet as fully phonologized as the progressive front–back and rounding harmony in Turkish.

It is insightful to compare the behavior of epenthetic vowels in word-initial consonant clusters to those occurring in word-final consonant clusters. Word-final consonant clusters are assumed to be present in the underlying representation of some native Turkish words as well as (primarily Arabic) borrowings (Yavaş 1980b). Yet again, an epenthetic vowel is inserted to prevent these clusters from surfacing in the coda position, as can be seen in the nominative singular (where no overt marker follows) and the locative forms in (24). Unlike in the case of word-initial epenthetic vowels, but just like with the suffix vowels underspecified for harmonic features, the regular front–back harmony and rounding harmony are *fully* imposed on word-final epenthetic vowels (24a)–(24c). Notice that there are commonly used borrowings where the epenthetic vowel that occurs within word-final consonant clusters is disharmonic (24d)–(24e).

(24) *Epenthetic vowels in root-final consonant clusters*

	UR	*nom sg*	*2nd pers poss*	*loc*	
a.	/bojn/	bojɯn	bojn-un	bojun-da	'neck'
		(*bojun)	(*bojun-un)	(*bojn-da)	
b.	/aln/	alɯn	aln-ɯn	alɯn-da	'forehead'
		(*alin)	(*alɯn-ɯn)	(*aln-da)	
c.	/gøɰs/	gøɰys	gøɰs-yn	gøɰys-te	'breast'
		(*gøɰis)	(*gøɰys-yn)	(*gøɰs-te)	
d.	/vakt/	vakit	vakt-in	vakit-te	'time'
		(*vakɯt)	(*vakit-in)	(*vakt-te)	
e.	/kavm/	kavim	kavm-in	kavim-de	'tribe'
		(*kavɯm)	(*kavim-in)	(*kavm-de)	

The emergence of the disharmonic epenthetic vowels in (24d)–(24e) can be taken to illustrate yet another instance of harmony blocking in Turkish. This is especially evident when the following suffixes surface as front (i.e. /vakt-in/, /kavm-in/, etc.) although the root vowel is back. Since the consonants at the right edge of these words are not palatals, it is not clear where the frontness feature comes from, providing additional data for the observation that the unexpected front variants of suffixes is not necessarily due to palatals (cf. (20) and (21) above). Instead, the disharmonic epenthetic vowel surfacing in (24d)–(24e), /vakit/ and /kavim/, can be considered as the result of the blocking of the harmonic features on the root vowel to further spread (see footnote 12 for a potential analysis), causing the epenthetic vowel (and the subsequent suffix vowels) to bear the default feature, which is arguably [Coronal] (i.e. front) in Turkish. Indeed, the fact that there are no roots where the final epenthetic vowel is back *and* disharmonic to the root vowel (e.g. */kevm/ → *[kevɯm]) provides a compelling argument in favor of this view.

In summary, there are several reasons to believe that word-final epenthetic vowels behave just like any other high vowel within roots or affixes, under the dictates of vowel harmony. The asymmetric behavior of word-initial and word-final epenthetic vowels, which are presumably identical elements, under front–back and rounding harmony suggests that what we call "harmony" in the case of word-initial epenthetic vowels runs on different principles, which nevertheless share some common properties with our classic progressive harmony, possibly due to universal tendencies and phonetic/perceptual factors.

Kabak, Barış, Kazumi Maniwa, Eva Kasselkus & Silke Weber. 2008. Vowel harmony has direction and context: A corpus study. Paper presented at the 16th Manchester Phonology Meeting.

Kabak, Barış, Kazumi Maniwa & Nina Kazanina. 2010. Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Laboratory Phonology* 1. 207–223.

Kardestuncer, Aino. 1982. A three-boundary system for Turkish. *Linguistic Analysis* 10. 95–117.

Kardestuncer, Aino. 1983. Vowel harmony and gerundive compounds in Turkish. *Acta Linguistica Hafniensia* 18. 55–64.

Kaun, Abigail. 1999. Epenthesis-driven harmony in Turkish. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 25. 95–106.

Kaun, Abigail. 2004. The typology of rounding harmony. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 87–116. Cambridge: Cambridge University Press.

Kenstowicz, Michael & Charles W. Kisseberth. 1977. *Topics in phonological theory*. New York: Academic Press.

Kiparsky, Paul. 1968. How abstract is phonology? Distributed by Indiana University Linguistics Club.

Kiparsky, Paul. 1973. Phonological representations. In Osamu Fujimura (ed.) *Three dimensions of linguistic theory*, 3–136. Tokyo: Taikusha.

Kiparsky, Paul & Karl Pajusalu. 2003. Towards a typology of disharmony. *The Linguistic Review* 20. 217–241.

Kirchner, Robert. 1993. Turkish vowel disharmony in Optimality Theory. Paper presented at the Rutgers Optimality Workshop 1, Rutgers University.

Lahiri, Aditi. 2000. Phonology: Structure, representation, and process. In Linda Wheeldon (ed.) *Aspects of language production*, 165–225. New York: Psychology Press.

Lees, Robert B. 1961. *The phonology of Modern Standard Turkish*. Bloomington: Indiana University.

Lees, Robert B. 1966. Turkish harmony and the description of assimilation. *Türk Dili Araştırmaları Yıllığı Belleten 1966*. Ankara: Ankara University.

Levi, Susannah V. 2001. Glides, laterals, and Turkish vowel harmony. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 37. 379–393.

Lightner, Theodore M. 1965. On the description of vowel and consonant harmony. *Word* 19. 376–387.

Lyons, John. 1962. Phonemic and non-phonemic phonology: Some typological reflections. *International Journal of American Linguistics* 28. 127–134.

McCarthy, John J. 1998. Morpheme structure constraints and paradigm occultation. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 32(2). 123–150.

Ní Chiosáin, Máire & Jaye Padgett. 2001. Markedness, segment realization, and locality in spreading. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 118–156. Cambridge: Cambridge University Press.

Polgárdi, Kriztina. 1999. Vowel harmony and disharmony in Turkish. *The Linguistic Review* 16. 187–204.

Postal, Paul. 1968. *Aspects of phonological theory*. New York: Harper & Row.

Prince, Alan & Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Sharaf, G. 1927. Paljatogrammy zvukov tatarskogo jazyka sravnitel'no s russkimi. *Vestnik Nauchnogo Obshchestva Tatarovedenija* 7. 65–102.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht.

Vago, Robert M. (ed.) 1980. *Issues in vowel harmony*. Amsterdam: John Benjamins.

Waterson, Natalie. 1959. Some aspects of the phonology of nominal forms of the Turkish word. *Bulletin of the School of Oriental and African Studies* 18. 578–591.

119 Reduplication in Sanskrit

ROBERT KENNEDY

Sanskrit, a classical and liturgical language of India, has been central to the advancement of linguistic analysis in a number of respects. It is notable for the support it provides for the "Indo-European hypothesis," the theory advanced by scholars such as Sir William Jones, Thomas Young, and Franz Bopp that the Indic languages are descended, along with those of the Romance, Germanic, Celtic, Slavic, and myriad others, from a common ancestral proto-language. Sanskrit has also been crucial to the development of autosegmental phonology and prosodic morphology, with its reduplicative system providing a wealth of illustrative data, such as syllable transfer effects (Steriade 1988) and the emergence of unmarked phonological structure (McCarthy and Prince 1994). The range of patterns also underscores the role of morphological conditioning in reduplication. This chapter provides an overview of reduplication in Sanskrit, portraying the range of formations, and their implications for modern theories of representation and derivation.

Sanskrit is an ancient language spoken on the Indian subcontinent. It was spoken as a living language in the third millennium BCE, and has since survived as a liturgical language of Hinduism. Eighteenth-century scholars noted similarities between Sanskrit and Classical Greek and Latin, giving rise to the Indo-European hypothesis and the comparative method in historical linguistics. Whitney (1885, 1889), Wackernagel (1896, 1905, 1930), and Wackernagel and Debrunner (1954) are comprehensive and authoritative sources on the language, though Wackernagel's volume on verbs was never completed. Monier-Williams (1864) is another early source. Sanskrit data are drawn from Vedic literature, a corpus of histories, ceremonies, legends, dramas, and sacrificial formulae, some of it versed, hymnal, or mantric, spanning centuries in the classical period. Written records of the language have been preserved in Devanagari script, which is phonetically transparent and thus expresses alternations. Many of the generalizations discussed in §3 and §4 are explicitly observed in Whitney (1889).

This chapter proceeds as follows. §1 introduces and summarizes Sanskrit reduplication, noting phonological and morphological alternations, while §2 is an overview of reduplicative models. §3 examines segmental alternations in reduplicated structures, with a discussion of their implicationsfor auto-segmental and output-oriented theories. §4 deals with prosodic effects, in particular the size, position, and syllable structure of reduplicative affixes. Here I consider the data from the perspectives of templatic autosegmental phonology, full-copy approaches, template constraints, and atemplatic approaches to reduplication. I conclude by suggesting a reorientation of analysis for Sanskrit reduplication along dimensions of both phonological and morphological alternation, as also suggested by Janda and Joseph (1986).

1 Overview of Sanskrit reduplication

Reduplication is a phenomenon in which part of a word is repeated to indicate a morphological function (see CHAPTER 100: REDUPLICATION). For example, the Sanskrit perfect stems [ta-tap] 'PERF-heat', [pa-pat-a] 'PERF-fly', and [sa-sarj-a] 'PERF-send forth' are derived by repeating the first consonant and vowel of the root. Reduplication bears on phonology for several reasons: (a) phonological similarity between reduplicant and base, (b) prosodic concerns such as the size and position of the reduplicant, and (c) scenarios in which the reduplicant's segmental or featural structure is not fully identical to any contiguous substring of the base.

Sanskrit reduplication comprises five prefixal patterns, each associated with a distinct stem formation: the intensive, the perfect, the aorist, the present, and the desiderative. In turn, each pattern has its own set of predictable phonological alternations. In this section I provide examples of each formation. Data are cited from Whitney (1885, 1889), Janda and Joseph (1986), and Steriade (1988); the latter two also use Whitney (1889) as a primary source. In some cases, missing glosses have been recovered from Monier-Williams (1864) or Deshpande (1997).

As a preview, we can make the following generalizations about the form of the various reduplicative affixes, using prosodic templates for the moment as a means of description (CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS). The intensive (1a) is /CVX/ or $\sigma_{\mu\mu}$, while the others are /CV/ or σ (1b) and (1c), and the size of the template determines how much of the copied string persists to the surface representation. The template can also be used to describe the fixed vowel of the reduplicant; thus, the intensive template includes a prespecification of [a] (1a), while the desiderative includes a [+high] specification (1b). The perfect template need only be CV (1c). Both the aorist and present may follow the desiderative pattern in some cases and the perfect in others.

(1)	a. *Intensive*	b. *Desiderative*	c. *Perfect*
	/C V X/	/C V/	/C V/
	|	|	
	[a]	[+high]	

1.1 *Intensive reduplication*

Display (2) provides examples of the intensive formation.[1] A variety of phonological formations is evident: CVC (2a), as in [bad-badh] 'INT-oppress'; CVCi (2b), as in [bhariː-bhr̥] 'INT-bear'; CVː (2c), as in [daː-dhjaː] 'INT-think'; and CV (2d),

[1] Data are transcribed generally in an IPA transliteration of Devanagari representations, except that aspirates are transcribed with [-h-]. Data tables note the source of each form: W = Whitney, S = Steriade, JJ = Janda and Joseph.

svar	sa-svara	'sound'	S
sjand	sa-ṣjade	'move on'	S
tap	ta-tap	'heat'	JJ
tviṣ	ti-tveṣa	'be stirred up'	S
tjaɟ	ta-tjaɟa	'forsake'	S
vac	u-vaːc	'speak'	JJ
vaç	va-vaça	'be eager'	S
vas	uːṣur (< u-vasa)	'shine'	JJ
jaɟ	iː-ɟe (< i-jaɟa)	'offer'	S
jat	je-te (< ja-jata)	'stretch'	S
ais	i-jeṣa	'seek, desire'	S
ai	i-jaja	'go'	S
aç	aːn-amç	'attaːn'	JJ
auc	uvoca	'please'	S
av	aːv	'favor'	JJ

1.3 Present stem reduplication

The third reduplicative pattern derives the present stem for the class of verbs illustrated in (4). Its formation is very similar to the perfect, except that base /ṛ/ and /a/ (in most cases) are reflected by [i] in the reduplicant, as in [ti-stha] 'PRES-stand' and [bi-bhṛ] 'PRES-bear'. Otherwise, the same properties of cluster reduction, palatalization, and loss of aspiration are evident.

(4) *Present stem formation*

root	*present*		*source*
bhiː	bi-bhiː	'fear'	W
bhṛ	bi-bhṛ	'bear'	W
çaː	çi-ça	'remove'	W
daː	da-daː	'give'	W
gaː	ɟi-gaː	'go'	W
ghraː	ɟi-ghraː	'smell'	JJ
hu	ju-huː	'sacrifice'	W
maː	mi-maː	'measure'	W
pṛc	pi-pṛc	'present'	W
sthaː	ti-ṣṭhaː	'stand'	JJ
vac	vi-vac	'speak'	W
vaç	vi-vaç, va-vaç	'eager'	W

1.4 Aorist stem reduplication

Many verbs also use reduplication in the derivation of their aorist stem (5).[3] Whitney notes an alternation in vowel length in the aorist stem: the reduplicant has a long vowel if the base contains a short vowel followed by one consonant (5a), as in [a-duː-duṣam] 'AOR-spoil'. The reduplicant has a short vowel if the base has a

[3] Whitney's examples include a prefix element /a-/ and the 1st person suffix /-am/, while other scholars provided abstractions without these elements; their examples are cited as such.

long vowel (5b), as in [a-bu-bhu:ṣam] 'AOR-adorn', or a short vowel followed by two consonants (5c), as in [a-da-dakṣam] 'AOR-be able'. Forms such as [a-ci-kṣipam] 'AOR-throw' show that vowel length is also avoided in the reduplicant if the base has an initial consonant sequence (5d), while the forms in (5e) show that vowel-initial roots also receive reduplicants with short vowels.

Typically the length alternation leaves the root's vowel length unchanged, but in some aorist forms derived from long-vowel roots, the reduplicative vowel is long and the base vowel is short, as seen in [a-si:-ṣadham] 'AOR-accomplish', from /sa:dh/. Furthermore, like the present stem formation, the reduplicant's vowel is typically [i] or [i:] where the root contains /a/, /ṛ/, or /ḷ/, as in [a-ɟi:-ɟanam] 'AOR-be born' and [a-vi:-vṛdham] 'AOR-grow'.

Note that that the aorist formation shows the same segmental effects as other forms: aspirates and velars are not copied as such, while only one consonant of an onset cluster is reduplicated.

(5) *Aorist stem formation*

	root	*aorist*		*source*
a.	bi:ṣ	a-bi:-biṣam	(no gloss)	W
	di:p	a-di:-dipam	'shine'	W
	duṣ	a-du:-duṣam	'spoil'	W
	ɟan	a-ɟi:-ɟanam	'be born, happen'	W
	ɟi:v	a-ɟi:-ɟivam	'live'	W
	kḷp	a-ci:-kḷpam	(no gloss)	W
	muc	mu:-muc	'release'	JJ
	ra:dh	a-ri:-radham	'succeed'	W
	riṣ	a-ri:-riṣam	'be hurt'	W
	sa:dh	a-si:-ṣadham	'accomplish'	W
	su:c	a-su:-sucam	(no gloss)	W
	taɲs	a-ta-taɲsam	'shake'	W
	tap	ti:-tap	'heat'	JJ
	va:ç	a-vi:-vaçam	'bellow'	W
	vṛdh	a-vi:-vṛdham	'grow'	W
b.	bhu:ṣ	a-bu-bhu:ṣam	'adorn'	W
	di:kṣ	a-di-di:kṣam	'be consecrated'	W
	dha:v	a-da-dha:vam	'run'	W
c.	dakṣ	a-da-dakṣam	'be able'	W
	kṛṣ	a-ca-karṣat, a-ci:-kṛṣat	'plough'	W
	vṛt	a-va-vartat, a-vi:-vṛtat	'be'	W
d.	krand	a-ci-kradam	'cry'	W
	çru	çu-çruv	'hear'	JJ
	krudh	a-cu-krudham	'get angry'	W
	kṣip	a-ci-kṣipam	'throw'	W
	pracch	a-pa-praccham	'ask'	W
	skand	a-ca-skandam	'leap'	W
	spṛç	a-pi-spṛçam	'touch'	W
	stu	tu-ṣṭav	'praise'	JJ
	sjand	a-si-ṣjadam	'flow'	W
	tras	a-ti-trasam	'be terrified'	W

e.	a:p	a:-pi-pan	'attain'	W
	am	a:-ma-mat	'injure'	W
	arc	a:r-ci-cam	'worship'	W
	arh	a:r-ɟi-ham	'deserve'	W
	arp	ar-pi-pam	(no gloss)	W
	i:kṣ	a:i-ci-kṣam	'see'	W
	ubɟ	aub-ɟi-ɟam	'force'	W
	ṛdh	a:r-di-dhan	(no gloss)	W

1.5 *Desiderative stem reduplication*

The last reduplicative pattern is the desiderative stem (6), whose structure resembles the present stem pattern of §1.3, and whose addition to a verb derives forms with a meaning roughly equivalent to 'to want to VERB'. The desiderative is derived with a CV reduplicative prefix and also includes a suffix /sa/, sometimes preceded by /i/. The vowel of the reduplicative element is uniformly high; it is always [i] (6a) or [u] (6b), the latter only if the base contains [u]. As with other reduplicative forms, aspirates and velars are not copied as such, and only one consonant of an onset cluster is reduplicated.

(6) *Desiderative stem formation*

	root	*desiderative*		*source*
a.	dhṛ	di-dhi:rṣa	'hold'	W
	ga:	ɟi-ga:sa	'go'	JJ
	gi	ɟi-gi:ṣa	'speech'	W
	hṛ	ɟi-hi:rṣa	'take'	W
	ɟi:v	ɟi-ɟi:viṣa	'live'	W
	ki	ci-ki:sa	(no gloss)	W
	kṛ	ci-ki:rṣa	'do, make'	W
	kṣi	ci-kṣi:sa	'possess'	W
	pa:	pi-pa:ṣa	'drink'	W
	spṛç	pi-spṛkṣa	'touch'	JJ
	sṛ	si-si:rṣa	'move, flow'	W
	stha	ti-ṣtha:sa	'stand'	W
	stṛ	ti-sti:rṣa	(no gloss)	W
	tṛ	ti-ti:rṣa	'cross'	W
	vid	vi-vidiṣa	'know'	W
	ja	ji-ja:sa	'go'	W
	jaɟ	i-jakṣa	'sacrifice'	JJ
	edh	e-di-dhiṣa	'thrive'	JJ
	arh	ar-ɟi-hiṣa	'deserve'	JJ
b.	bhṛ	bu-bhu:rṣa	'carry'	W
	çru	çu-çru:ṣa	'hear, listen'	W
	çudh	çu-çutsa	'purify'	JJ
	hu	ɟu-hu:ṣa	'sacrifice'	W
	kṣu	cu-kṣu:ṣa	(no gloss)	W
	mṛ	mu-mu:rṣa	'die'	W
	smṛ	su-smu:rṣa	'remember'	JJ

(CHAPTER 14: AUTOSEGMENTS; CHAPTER 54: THE SKELETON). Derivationally, these models fit into three general categories: the rule-ordered copy-and-associate approach, the correspondence approach, and the full-copy approach.

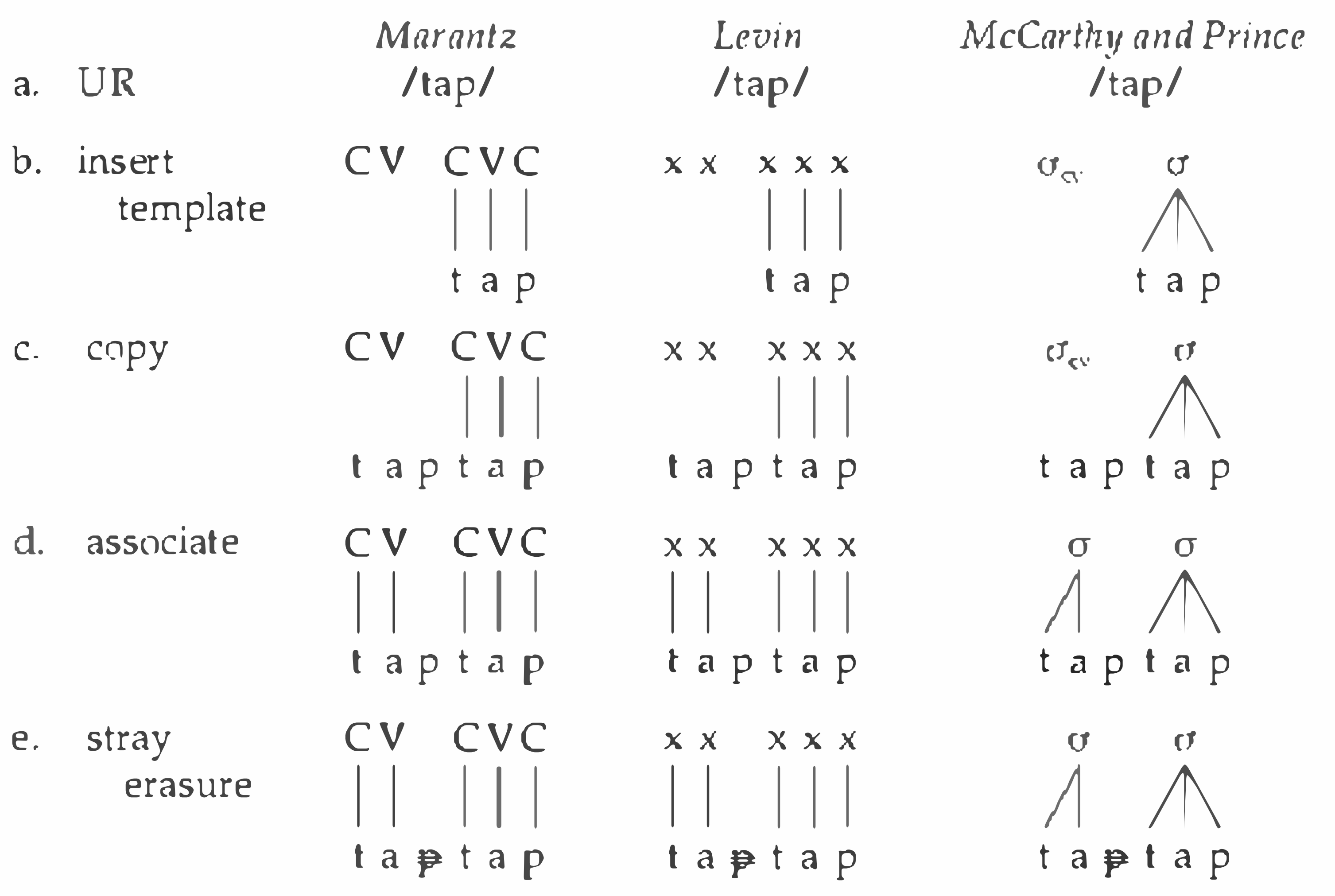
In this section, I briefly discuss how each approach would model Sanskrit reduplication, using the present stem formation of the root /tap/ 'heat'. I should note explicitly that of the models discussed in this section, only Steriade (1988) offers any detailed elaboration on the Sanskrit data. Thus, the exemplifications of how other models would handle Sanskrit reduplication are drawn from my own attempts at applying such models to these data. Further, while I note some strengths and shortcomings of each model here, I return to a fuller evaluation of all of them in §4, after a description of segmental processes in §3.

2.1 *Rule-ordered template satisfaction*

The rule-ordered approaches of Marantz (1982), Levin (1985), and McCarthy and Prince (1986) combine an autosegmental templatic morpheme with a derivational model of copy and association: the reduplicative morpheme is an empty prosodic shell, whose addition to a stem (7b) precipitates a copy operation that doubles the segmental melody of the root (7c). The doubled melody associates maximally to the template (7d), and left-over unassociated segments are erased (7e). Such models allow rules to precede or follow the copy operation, which accounts for the observation of a phonological process in one or both morphemes of a reduplicated word.

The three derivational templatic models differ from each other in the nature of the template: Marantz (1982) uses a CV skeleton, Levin (1985) uses a bare X-slot skeleton, while McCarthy and Prince (1986) use units of the prosodic hierarchy – in this case, the core syllable (one which precludes onset clusters and coda consonants).

(7)



The full-copy models are capable of capturing the syllable transfer effects discussed in §4.3; indeed such data motivate Steriade's approach in the first place. However, Steriade's model suffers from the ordering paradoxes discussed in §3.1, while Inkelas and Zoll's model may be troubled by the opacity found in the activity of the *ruki* rule discussed in §3.2, and also incurs the expense of positing an entirely distinct phonological grammar for different morphemes of the same word.

3 Segmental phonology of Sanskrit reduplication

Whitney notes a number of segmental effects which interact with reduplication:

> A non aspirate is substituted in reduplication for an aspirate . . . A palatal is substituted for a guttural or for h . . . Of two initial consonants the second if it be a non nasal mute preceded by a sibilant is repeated instead of the first. (1889: 222)

He notes elsewhere that "the dental sibilant [s] is changed to the lingual [ṣ] if immediately preceded by any vowel save [a] and by [k] or [r] unless the [s] be final followed by [r]" (1889: 61); this process, called the *ruki* rule by modern scholars, also interacts with reduplication.

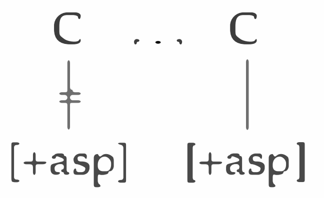
In this section I describe how these processes interact with reduplication and evaluate how well the models in §2 capture them. I first discuss deaspiration and palatalization in §3.1, processes which apply generally in the language to proximate consonants, and which affect the initial consonant of the reduplicative consonant. I then discuss the *ruki* rule in §3.2, focusing on its instantiation of reduplicative overapplication. In §3.3, I discuss cluster simplification in the reduplicants of roots with onset consonant sequences.

3.1 Consonant dissimilation

Two patterns of dissimilatory consonantal alternation are observable in Sanskrit reduplication: velar root consonants and /h/ have palatals as their reduplicative counterparts, and aspirates have unaspirated reduplicative counterparts. Both processes are observed in all CV reduplicative formations; the intensive CVC prefix sometimes allows velars and aspirated consonants.

Palatalization applies to velar consonants regardless of the nuclear vowel of the base or reduplicant (CHAPTER 71: PALATALIZATION); for example, the forms [ɟa-gam] 'PERF-go', [ci-kiːrṣa] 'DESID-do', and [a-cu-krudham] 'AOR-get angry' have palatal consonants in their reduplicants. Aspiration is not copied in the reduplicant; for example, the reduplicative consonant in [ti-sthaːsa] 'DESID-stand', [di-dhiːrṣa] 'DESID-hold', [ɟa-ghasa] 'PERF-eat', and [bi-bhṛ] 'PRES-bear' is unaspirated despite the aspiration in the base. The loss of aspiration in the reduplicant is an instance of Grassman's Law, a general process that prevents adjacent aspirates, accounting for Sanskrit exceptions to consonant correspondences in historical comparative analysis. Though a diachronic process which affects the form of roots in Sanskrit, its role in reduplication suggests a synchronic sensitivity to it. An autosegmental dissimilation rule that removes aspiration is provided in (13). Note that this rule may apply regardless of the number of intervening segments.

(13) *Grassman's Law*



Consonant dissimilation creates a paradox in derivational copy-and-associate models such as Marantz (1982) and McCarthy and Prince (1986), as well as the full-copy analysis of Steriade (1988). The fact that unreduplicated roots generally adhere to palatalization and to Grassman's Law suggests a set of morpheme structure conditions (MSCs; CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS) on the representation of underlying forms. Yet the application of these rules in reduplicated forms requires ordering them after the copying operation, which duplicates the MSCs post-lexically.

Consider for example the perfect stems [bi-bhed] 'PERF-split' (from /bhid/)and [bu-bodha] 'PERF-know' (from /baudh/). Both have one aspirated consonant; there are no roots with two proximate aspirated consonants, a fact attributed to the activity of an MSC holding over underlying representations. Following the formation of the present stem with reduplication, the intermediate form /bhi-bhaid/ contains a sequence of aspirates, to which the synchronic Grassman's Law must apply. Meanwhile, the intermediate reduplicated form /bu-baudh/ copies an unaspirated consonant, so Grassman's Law does not apply.

(14)	UR (MSC applies)	bhid	baudh
	present stem formation	bhi-bhaid	bu-baudh
	Grassman's Law	bi-bhaid	—
	other rules, output	bi-bhed	bu-bodh

One resolution is to posit a unitary post-reduplicative ordering of these processes, without any MSC. In such a scenario, without any underlying MSC, a hypothetical two-aspirate root such as /bhaudh/ is a possible lexical representation, but because of Grassman's Law, it would still surface as [-budh]. However, the rules that instantiate palatalization and Grassman's Law must follow the copy operation, as forms like [bi-bhed] show. This ordering allows velars and aspirates to occur in reduplicants as the counterparts of bases which themselves have undergone these processes. In the case of hypothetical /bhaudh/, the copying process produces /bhu-bhaudh/; Grassman's Law then yields [bhu-baudh], with aspiration remaining in the reduplicative [bh]. Thus, without the MSC, the rule-ordered approach fails to preclude a certain class of unattested reduplicated structures.

(15)	UR (no MSC)	bhaid	*bhaudh
	present stem formation	bhi-bhaid	bhu-bhaudh
	Grassman's Law	bi-bhaid	bhu-baudh
	other rules, output	bi-bhed	*bhu-bodh

Optimality Theory (Prince and Smolensky 1993) abandons MSCs for Richness of the Base, and formalizing palatalization (16a) and Grassman's Law (16b) as output constraints avoids any overgenerative result. Both constraints are respected in unreduplicated stems, and in each morpheme of reduplicated stems. This suggests that they are ranked higher than a constraint requiring faithful realization of underlying aspiration (16c).

(16) a. *[VEL][VEL]
Adjacent specifications of velar consonant place are forbidden.
b. GRASSMAN
Adjacent specifications of [+asp] are forbidden.
c. MAX-IO[+asp]
Each input [+asp] feature has a correspondent in the output.

As summarized in (17), Grassman's Law is respected in unreduplicated words by virtue of an output-oriented constraint, and the hypothetical unreduplicated form /bhaudh/ would appear as [bodh-]. Meanwhile, this alternation also motivates faithfulness constraints to target elements other than the segment.

(17)

	/bhaudh/	GRASSMAN	MAX-IO[+asp]
	a. bhodh	*!	
☞	b. bodh		*

In the reduplicative context, these consonant alternations motivate additional faithfulness constraints. For example, in [bi-bhed], the reduplicative [b] must be seen as the faithful correspondent of the base [bh]. The fact that the consonants have opposite specifications for [+asp] suggests that the constraint IDENT-BR[+asp], which requires corresponding base and reduplicative segments to have identical [asp] features, is violable. This is illustrated in (19), where IDENT-BR is ranked below MAX-IO[asp].

(18) IDENT-BR[+asp]
Segments in correspondence have identical [asp] specifiications.

The interpretation of MAX-IO needs to be quite narrow in this analysis: the aspiration feature on the reduplicative consonant is not to be taken as the output correspondent of the underlying aspiration feature. Thus, in (19b), the absence of aspiration in the base consonant triggers a violation of MAX-IO[+asp].

(19)

	/PRES+bhaid/	GRASSMAN	MAX-IO[+asp]	IDENT-BR[asp]
☞	a. bi-bhed			*
	b. bhi-bed		*!	*
	c. bhi-bhed	*!		
	d. bi-bed		*!	

Furthermore, when we return to the hypothetical problem form /bhaudh/, this additional correspondence constraint avoids the overgenerative consequence of the rule-ordered approach. IDENT ensures that the reduplicant has an unaspirated consonant to match the unaspirated consonant of the root. Thus reduplicated and unreduplicated roots both respect a unitary surface-oriented instantiation of Grassman's Law in the correspondence approach.

(28)

/PERF+smr̥/	MAX-IO	ONSET	*[+son]/ONS	*[+sib]/ONS	MAX-BR
a. smaː-smr̥			**!	**	
☞ b. saː-smr̥			*	**	*
c. maː-smr̥			**!	*	*
d. saː-sr̥	*!			**	
e. a-smr̥		*!	**		**

The correspondence approach restricts the cluster reduction pattern to the reduplicative substring; it does so through the TETU interaction between markedness and the formal distinction between reduplicative and base substrings. The base is not exceptional to the general markedness constraint, and the reduplicant is not subject to any morpheme specific process.

4 Prosodic morphology

In addition to the segmental effects discussed in §3, Sanskrit reduplication presents some particular challenges in its prosodic morphology. In particular, there is some morphological conditioning within the reduplicative prefix, both in its size and in its vocalism. In §4.1 I describe these conditions in more detail, while §4.2 shows how each of the models in §2 would handle them. Section 4.3 describes the syllable transfer effect, with an accompanying discussion again of how various models may or may not capture it.

4.1 Morphological conditioning

Despite the range of surface forms of reduplicative affixes in Sanskrit, there is sufficient regularity to attribute much alternation in form to phonological concerns, and what remains to morphological conditioning. There are two points of morphological conditioning: whether the affix indicates intensive, and whether it has a fixed vowel. CVC-, CVː-, CVG-, VC-, and CVCi- can be treated as predictable alternants of the intensive prefix: most such forms never map to the other functions, save the CVː- alternant of the aorist. Conversely, no CV- prefix ever carries an intensive function. The remaining phonological alternants are then predictable variants of a separate reduplicative formation. Vocalism is another dimension of morphological conditioning: the intensive always has [a], while the desiderative always has a high vowel. The perfect copies the base vowel, but the present and aorist vacillate between a fixed vowel and a faithful copy.

4.2 Patterns and templates

Sanskrit is a multi-pattern reduplicative system; such systems have received relatively less theoretical attention from theorists, but it is worth investigating how each of the prominent models of reduplication handles them. The templatic models could model a multi-pattern system by positing distinct inputs from different reduplicative affixes: as shown in (1) (repeated as (29)), the intensive is underlyingly

/CVX/ or $\sigma_{\mu\mu}$ (1a), while the others are /CV/ or σ (1b) and (1c). All reduplicants can then be generated by a single copying rule, but the size of the template determines how much of the copied string persists to the surface representation. The template can also be used to determine the fixed vowel of the reduplicant; thus, the intensive template includes a prespecifiication of [a] (29a), while the desiderative includes a [+high] specification (29b). The perfect template need only be CV (29c). The aorist and present formations, as we will see, both variably converge with the desiderative in some cases and the perfect in others.

(29) a. *Intensive* b. *Desiderative* c. *Perfect*

a. *Intensive*	b. *Desiderative*	c. *Perfect*
/C V X/	/C V/	/C V/
|	|	
[a]	[+high]	

However, the choice of a fixed-vowel or faithful-vowel template for the present and aorist stems must be left to allomorphic devices, as some formations require a general CV prefix, e.g. [da-daː] 'PRES-give', while others call upon a fixed high-vowel prefix, e.g. [ɟi-gaː] 'PRES-go'. In a copy-and-associate approach, a root must be specified for which of the two CV templates it takes when deriving the present or aorist stems. Further, the length alternation in the aorist motivates a rule specific to that formation; it must lengthen the reduplicative vowel under specific conditions, but it must not be allowed to apply to any of the other reduplicative stems.

Steriade's model uses no templates to distinguish reduplicative morphemes. Instead, the intensive differs from the others because of a rule which assigns a nuclear [a] just to that morpheme, along with different licensing conditions on coda consonants. Forms like [car-karṣ] 'INT-plough' and [kani-krand] 'INT-cry out' keep their postvocalic reduplicative consonants, because intensives are not subject to the same restrictions on coda consonants as perfect stems; otherwise, the intensive affix would not remain distinct from the other reduplicants. Meanwhile, though Steriade's focus is on the intensive and perfect markers, her model readily accommodates the fixed [+high] feature for the desiderative, inserting it in the same way the vowel [a] is inserted just for intensive prefixes.

Optimality-theoretic models vary in their treatment of multi-pattern systems. Template constraints such as those in McCarthy and Prince (1993) would demand that a particular reduplicative morpheme map to a specified prosodic category, for example foot or syllable. Sanskrit thus motivates two constraints: a specific INTENSIVE=$\sigma_{\mu\mu}$ and a general RED=σ. Generalized Template Theory (Urbanczyk 1996) would instead label the intensive as a root and the other reduplicants as affixes. A general prosodic constraint would force the intensive reduplicant, by virtue of its being a root, to project a foot. Conversely, another constraint would restrict other reduplicative morphemes, by virtue of their being affixes, to project at most a syllable. More general approaches (Gafos 1998; Kennedy 2003) do not use the root *vs.* affix distinction for different kinds of reduplicants, and without any morphological distinction among types of affixes, they cannot model morphologically conditioned size alternation.

To resolve this, Kennedy (2008) draws size differences from the morphological position of the affix relative to the stem: reduplicative morphemes are either stem-internal or stem-external in this model, and are thus subject to different general constraints on the alignment of morphological and prosodic elements. For example, the Sanskrit intensive would be posited as stem-external, aligning to a foot boundary, and is thus relatively larger than other reduplicants which need not align to foot boundaries.

To handle fixed vowels in reduplicants, correspondence approaches have three general strategies. One is to place the fixed vowel feature in the input representation of the reduplicant, in which case the present affix underlyingly is /Red, [+high]/, while the desiderative is underlyingly /Red, [a]/. These features are then maintained in the reduplicative substring via faithfulness constraints. Alternatively, the vowel features could be represented as part of the template constraint, in which case one constraint encodes Intensive= $\sigma_{\mu\mu}$+[a] and another encodes Desiderative=σ+[+high], while the perfect is subject to Red=σ. The reduplicant thus remains segmentally empty, and acquires its fixed vowel by satisfying the template constraint. A third approach is to leave the vowel choice to emergent markedness constraints, which would assign vowels by default (Alderete *et al.* 1995). Since some reduplicants contain fixed low vowels and others contain fixed high vowels, this approach is not appropriate for Sanskrit. Regardless, none of these approaches resolves the vacillation within the present and aorist stem affixes.

4.3 *Syllable transfer effects*

A final aspect of Sanskrit reduplication is the phenomenon of syllable transfer, discussed in depth by Steriade (1988). In short, the segment which appears in the nucleus of the reduplicant's syllable must have a base correspondent occurring somewhere in a rhyme, and a base onset segment can only have a counterpart in the reduplicant's onset. For example, in intensives like [car-karṣ] 'int-plough' and [kani-krand] 'int-cry out', the reduplicant copies a sonorant from a rhyme position in the base. In contrast, in [ɟaː-grah] 'int-seize' and [saː-smr̩] 'int-remember', the base has a sonorant as part of an onset cluster which is not copied; syllable transfer precludes forms such as *[ɟr̩-grah] or *[ɟari-grah] and *[sm̩-smr̩] or *[sami-smr̩]. The same effect is seen in perfect stems: the underlying glides in rhyme position in roots such as /baudh/ 'know' and /bhaid/ 'split' are reflected as the nuclei of the prefix syllable, as in [bu-bodha] 'perf-know' and [bi-bhed] 'perf-split'. However, the underlying onset glides in roots such as /svaɟ/[4] 'embrace' and /khjaː/ are not copied in [sa-svaɟa] 'perf-embrace' and [ca-khjau] 'perf-see'.

Of the rule-ordered models discussed here, only Steriade's full-copy approach handles this effect: the entire root is copied and syllabified, and the resulting structure is subject to a series of processes removing unlicensed material from the reduplicative substring. One process removes glides from onset clusters, turning /svaɟ/ to /saɟ/. In contrast, because copy-and-associate models generate a segment melody and associate it to an empty prosodic template, there is nothing barring the association of the reduplicant melody's /v/ in the melody to the nuclear position of the template. In other words, the copy-and-associate model predicts forms such as *[su-svaɟa] instead of [sa-svaɟa] and *[sami-smr̩] or *[sm̩-smr̩] instead of [saː-smr̩].

Nevertheless, this generalization has some limited exceptions, as in forms such [su-ṣvaːpa] 'perf-sleep', [i-jaɟa] 'perf-offer', and [u-vasa] 'perf-shine', where onset

[4] See note 2, p. 2858.

glides do vocalize as nuclear vowels in the reduplicant (see CHAPTER 15: GLIDES). To address this, Steriade argues that the adherence of a root to the syllable transfer effect is related to whether it has a distinction between its zero-grade and full-grade stems. Generally, full-grade stems include [a], and each grade is used only in specific morphosyntactic contexts, but some roots exceptionally are invariant across grades, keeping the [a] in both the zero and full grade.

For example, the full-grade and zero-grade stems for /svaɟ/ 'embrace' converge upon [sa-svaɟa], illustrating syllable transfer, while the full-grade stem in [su-svapa] 'sleep' differs from the corresponding zero-grade stem in [su-sup]. Steriade handles this by ordering perfect reduplication (30a) ahead of a rule of Syncope (30b), which removes unaccented [a] (thus deriving the zero grade from the full grade). Syncope also applies to any instances of [a] in the prefix regardless of grade, thus turning /svap-svap/ to /sup-svap/, in which the glide is necessarily vocalized – and since it is no longer part of an onset cluster, it survives the later step of removing unlicensed material (30c). In contrast, roots such as /svaɟ/ are exceptions to Syncope (30b); their intermediate forms keep their /a/, and the onset /v/ is deleted from the prefix in the removal of unlicensed material (30c).

(30)			svap (full grade)	svap (zero grade)	svaɟ (full grade)	svaɟ (zero grade)
	a.	perfect	svap-svap	svap-svap	svaɟ-svaɟ	svaɟ-svaɟ
	b.	Syncope	sup-svap	sup-sup	*exception*	*exception*
	c.	Licensing	su-svap	su-sup	sa-svaɟ	sa-svaɟ

Syllable transfer effects are exceptions to the typological claim made by Moravcsik (1978) that reduplication does not copy prosodic constituents. Sanskrit thus challenges the notion of template satisfaction (McCarthy and Prince 1986, 1993), which holds that reduplicative melodies are copied irrespective of the syllabicity of base segments. Other examples of syllable transfer have been noted, such as in West Tarangan (Nivens 1992, 1993; Takata 1992; Takata and Takata 1992; Spaelti 1997), Yaqui (Haugen 2003), and Temiar (Benjamin 1976; Broselow and McCarthy 1983). Though such effects are problematic for the copy-and-associate approach, correspondence approaches can model syllable transfer; for example, McCarthy and Prince (1993, 1994) propose the constraint ST-ROLE, which requires segments in base–reduplicant correspondence to occupy similar prosodic positions. This constraint marks structures in which base onset segments have rhyme segments as their reduplicative correspondents, and would rule out candidates such as *[su-svaɟa] instead of [sa-svaɟa] and *[sami-smr̥] instead of [saː-smr̥].

5 Summary

Reduplication in Sanskrit is a rich source of data with a variety of implications for phonological theories and models. The very breadth of the data is daunting, given the variety of formations, the phonological alternations evident in each, and the persistent exceptionality to some generalizable patterns. Even so, the exceptionality is generally principled, and the system is a fruitful source of data on reduplicative segment alternations and prosodic structure.

As Janda and Joseph (1986) note, a unitary notion of reduplication in Sanskrit is not attainable, given the larger size of the intensive marker and the appearance of fixed segmentism in only some of the remaining formations. Yet the various formations maintain a fairly consistent respect for palatalization, Grassman's Law, and cluster reduction, diverging from the former two only in certain examples of the intensive. Meanwhile, the vocalism of the present and aorist markers is not entirely predictable, but also not random: it always converges with the perfect or the desiderative vocalic pattern, and is attributable to a single point of allomorphic variation.

The multitude of prosodic shapes for reduplicative structures in Sanskrit is reducible almost to two, save the weight alternation observed in the aorist marker. Given the general trend in phonological approaches to reduplication to reduce the set of prosodic targets and generalize the relationship between reduplicative morphemes and prosodic structure, accounting for this alternation will be a necessary step in any coherent theory of multi-pattern reduplication.

The study of Sanskrit reduplication illustrates the progression of phonological theory from rule-ordered template satisfaction to generalized morphoprosodic output constraints. Yet in one respect, the rule-ordered approach is better equipped. It is only with cyclicity that the opaque interaction between reduplication and the *ruki* rule can be sensibly captured. In contrast, other aspects of Sanskrit reduplication are better modeled with optimality-theoretic constraints, such as Grassman's Law, palatalization, cluster reduction, and, ironically, the overapplication effect brought about by the interaction of reduplication and the *ruki* rule.

Clearly, no single phonological framework handles all the data flawlessly, though a constraint-based approach that incorporates intermediate representations to handle the opacity of the *ruki* rule, and allomorphic prespecification to handle variation, seems the best fit. In contrast, the two serial approaches seem less amenable to any theoretical repair to address their weak spots, most notably the segmental phenomena discussed in §3.

Regardless of their derivational character, modern generative approaches offer some additional insights into the data which grammarians such as Whitney were unable to provide. Although Whitney's own approach to reduplication in Sanskrit was to describe it thoroughly with sensible generalizations, he offered little more than a taxonomic list of subpatterns and the conditions under which they arise. In contrast, modern approaches have at their disposal a number of theoretical constructs, such as the feature, the autosegment, the template, the prosodic hierarchy, and the mora, all of which help unify this widely disparate data set, providing order to Whitney's taxonomy.

REFERENCES

Alderete, John, Jill Beckman, Laura Benua, Amalia Gnanadesikan, John J. McCarthy & Suzanne Urbanczyk. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30. 327–364.

Benjamin, Geoffrey. 1976. An outline of Temiar grammar. In Philip N. Jenner, Laurence C. Thompson & Stanley Starosta (eds.) *Austroasiatic studies*, vol. 2, 129–197. Honolulu: University Press of Hawaii.

Broselow, Ellen & John J. McCarthy. 1983. A theory of internal reduplication. *The Linguistic Review* 3. 25–88.

Spaelti, Philip. 1997. Dimensions of variation in multi-pattern reduplication. Ph.D. dissertation, University of California, Santa Cruz.

Steriade, Donca. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5. 73–155.

Takata, Masahiro & Yuko Takata. 1992. Kola phonology. In Burquest & Laidig (1992), 31–46.

Takata, Yuko. 1992. Word structure and reduplication in Kola. In Burquest & Laidig (1992), 47–68.

Urbanczyk, Suzanne. 1996. Patterns of reduplication in Lushootseed. Ph.D. dissertation, University of Massachusetts, Amherst.

Wackernagel, Jacob. 1896. *Altindischen Grammatik*, vol. I: *Lautlehre*. Göttingen: Vandenhoeck & Ruprecht.

Wackernagel, Jacob. 1905. *Altindische Grammatik*, vol. II-1: *Einleitung zur Wortlehre: Nominalkomposition*. Göttingen: Vandenhoeck & Ruprecht.

Wackernagel, Jacob. 1930. *Altindische Grammatik*, vol. III: *Nominalflexion: Zahlwort-Pronomen*. Göttingen: Vandenhoeck & Ruprecht.

Wackernagel, Jacob & Albert Debrunner. 1954. *Altindische Grammatik*, vol. II-2: *Die Nominalsuffixe*. Göttingen: Vandenhoeck & Ruprecht.

Whitney, William Dwight. 1885. *The roots, verb-forms and primary derivations of the Sanskrit language*. Leipzig: Breitkopf & Hartel.

Whitney, William Dwight. 1889. *Sanskrit grammar, including both the classical language, and the older dialects of Veda and Brahmana*. 2nd edn. Cambridge, MA: Harvard University Press.

120 Japanese Pitch Accent

Haruo Kubozono

1 Introduction

The goal of this chapter is to sketch Japanese pitch accent systems, focusing in particular on the major debates and findings in the past and the present. Many Japanese dialects, including the standard Tokyo dialect (henceforth Tokyo Japanese), have a lexical pitch accent, but pitch accent systems vary greatly from one dialect to another. This can be exemplified by (1), which shows how the recent loanword /ma.ku.do.na.ru.do/ 'McDonald's' is pronounced in different dialects. In (1) and the rest of this chapter, high-pitched portions are denoted by capital letters, while syllable boundaries are indicated by dots wherever necessary.

(1)	Tokyo	ma.KU.DO.NA.ru.do
	Kyoto/Osaka	ma.ku.do.NA.ru.do
	Nagasaki	MA.KU.DO.NA.RU.DO
	Kagoshima	ma.ku.do.na.RU.do
	Miyakonojo	ma.ku.do.na.ru.DO
	Koshikijima	MA.KU.DO.na.RU.do

Given this geographical diversity of accent patterns, it is necessary both to discuss Japanese pitch accent for each dialect, and to compare different systems. Since much previous work has centered around the pitch accent system of Tokyo Japanese, we will focus on this system in the first half of this chapter, while discussing other systems in the second half, to illustrate the diversity of pitch accent in the language.

Before going into the main discussion, let us define some basic notions (see CHAPTER 42: PITCH ACCENT SYSTEMS for more general discussion of the notion of pitch accent). The term "accent" is used in two different ways in Japanese phonology. This can be illustrated by the pitch patterns in (2) in Tokyo Japanese, where both the presence and location of a sudden pitch fall are distinctive. In many generative studies of Japanese phonology, "accent" is used to refer to phonological prominence that is assigned to a particular position of a word. According to this definition, words in (2a)–(2c) have an "accent" on a particular syllable or mora; this position is phonologically marked as "accented" since pitch abruptly falls immediately after it. On the other hand, (2d) has no accent, i.e., /sakana/ is an "unaccented word," since no abrupt pitch fall occurs even if a particle is attached to it (/ga/ is a nominative particle).

(2)	*word*	*surface pitch*	*accent representation*	
a.	inoti	Inoti-ga	í.no.ti	'life'
b.	kokoro	koKOro-ga	ko.kó.ro	'heart'
c.	otoko	oTOKO-ga	o.to.kó	'man'
d.	sakana	saKANA-GA	sa.ka.na	'fish'

In addition to this definition, the term "accent" has conventionally been used more loosely to refer to the overall pitch pattern or shape of a word. According to this definition, every word in (2) has an accent: even the so-called "unaccented word" in (2d) has an accent in the sense that it has a fixed pitch pattern with which it must be pronounced. To avoid confusion, I will adopt the first definition of the term in this chapter, while using the term "pitch pattern" or "accent pattern" to refer to the surface pitch pattern of words.

Japanese pitch accent as defined in this way is different from word stress in English, in that its phonetic correlate is primarily pitch or fundamental frequency (F0) and is independent of duration or intensity (Beckman 1986; see also CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE). This phonetic nature of pitch accent is supposedly consistent across dialects.

This chapter is organized as follows. In the next section, we will briefly consider a typology of Japanese pitch accent systems, including that of Tokyo Japanese. §3 discusses various aspects of Tokyo Japanese, focusing in particular on past debates about accent computation and representation. §4 and §5 describe other prosodic systems, to illustrate the diversity of pitch accent in the language. Kagoshima and Koshikijima Japanese are chosen in these sections as typical examples of pitch accent systems which are strikingly different from that of Tokyo Japanese. The final section summarizes the main points, as well as some issues for future work.

2 Typology of Japanese pitch accent systems

Based on the number of lexical pitch contrasts, Uwano (1999) proposes a dichotomy of Japanese pitch accent systems: multiple-pattern systems *vs.* N-pattern systems (where N can be any integer starting from one). In the former type, the number of contrastive patterns varies depending on the phonological length of the word. The accent system of nouns in Tokyo Japanese is a typical example of this type, since lexical contrasts increase in number as words become longer. On the other hand, N-pattern systems have a fixed number of lexical pitch contrasts, which is independent of the phonological length of words. For example, the Miyakonojo dialect spoken in Miyazaki Prefecture in the south of Japan has a one-pattern system where every word has a high pitch on the final mora, as illustrated in (3). In this system, the pitch feature need not be specified for each word in the lexicon, since it can be assigned by a redundancy rule (see CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION).

accent patterns for nouns of *n* syllables, where *n* represents an integer equal to or greater than one (McCawley 1968; Akinaga 1985; Shibatani 1990; Uwano 1999).[1] Thus monosyllabic words have two distinctive pitch patterns, as in (6a), disyllabic nouns have three, as in (6b), and trisyllabic nouns have four, as in (6c), etc. Each case involves a contrast in accentedness, i.e. between accented and unaccented patterns; disyllabic or longer nouns additionally involve a contrast in accent position. Phonetic patterns are shown in parentheses. In the rest of this chapter, unaccented words are marked with a superscript circle (°) for the sake of description.

(6)	a.	hí-ga	(HI-ga)	'fire-NOM'
		hi°-ga	(hi-GA)	'sun-NOM'
	b.	áki-ga	(Aki-ga)	'autumn-NOM'
		akí-ga	(aKI-ga)	'weariness-NOM'
		aki°-ga	(aKI-GA)	'vacancy-NOM'
	c.	ínoti-ga	(Inoti-ga)	'life-NOM'
		kokóro-ga	(koKOro-ga)	'heart-NOM'
		otokó-ga	(oTOKO-ga)	'man-NOM'
		sakana°-ga	(saKANA-GA)	'fish-NOM'

The *n* + 1 rule holds for nouns, but not for verbs and adjectives. Verbs and adjectives in Tokyo Japanese have only two pitch patterns, irrespective of their length. They exhibit a contrast in accentedness, just like nouns, but not in accent position, as accented verbs and adjectives are accented on the penultimate mora, as shown in (7) (see also CHAPTER 102: CATEGORY-SPECIFIC EFFECTS).

(7)	a.	*Verbs*			
		accented		*unaccented*	
		náru	'to be completed'	naru°	'to ring'
		haréru	'to clear up'	hareru°	'to become swollen'
		arúku	'to walk'	akeru°	'to open'
	b.	*Adjectives*			
		accented		*unaccented*	
		atúi	'hot'	atui°	'thick'
		umái	'tasty'	amai°	'sweet'

Research on pitch accent in Tokyo Japanese has centered around the following four questions: (i) how to derive surface pitch patterns, (ii) how to generalize accentual phenomena into accent rules, (iii) how to account for the distribution of unaccented words as opposed to accented ones, and (iv) how to compute accent patterns in a theoretical framework. Due to limitations of space, we will focus on the first three questions in the following sections.

3.1 Accent-to-pitch conversion

Traditionally, the four pitch patterns in (2) have been represented as in (8), where every mora is assigned either a high tone (H) or a low tone (L) (Haraguchi 1977).

[1] This rule does not hold for relatively long words and compound nouns, as we will see in §3.2.4.

nouns – whether native, SJ, or foreign – can remain unspecified as long as they obey the accent rule in (10a), while other accented nouns are marked in the lexicon as to their accent location. This analysis entails that native nouns primarily differ from loanwords in the extent to which lexical exceptions are permitted: the former group permits a greater number of lexically marked words than the latter group.

This idea leads to a sparsely specified representation like the one in (11), where * denotes a floating accent that is not linked to any particular position of the word in the lexicon (Kubozono 2008b). In other words, * marks a word as accented, but without saying where it is accented. This floating accent – or unlinked high tone – is subsequently anchored on a particular position by the antepenultimate accent rule.

(11) inoti*, kokóro, otokó, sakana

This representation is crucially different from those in (9), in that it treats the antepenultimate pattern as a default pattern for accented nouns. In other words, it predicts that new nouns will take this accent pattern if they are accented at all. This prediction can be readily borne out by an analysis of nonsense nouns such as /go.re.se/ and /ma.ta.go.ra.sa/, which display a strong bias towards the antepenultimate pattern (Tanaka 1996). The representation in (11) is also compatible with the fact that accent patterns are asymmetrically distributed if their frequencies are considered, namely, that the antepenultimate pattern predicted by (10a) accounts for a majority of accented nouns, irrespective of the lexical stratum they belong to (Kubozono 2006a, 2006b, 2008b).

The analysis of lexical specification in (11) implies that nouns in Tokyo Japanese permit only two productive accent patterns in addition to lexically marked ones. What this means is that nouns in this dialect basically constitute a two-pattern accent system like adjectives and verbs of the same dialect, as well as all words in Kagoshima Japanese (see (4) above and §4 below). In all these systems, the contrast between accented and unaccented patterns – equivalently, presence or absence of an abrupt pitch fall in phonetic terms – is relevant, whereas the position of accent (or pitch fall) is largely redundant. The accent system of nouns in Tokyo Japanese differs from that of adjectives and verbs in the same dialect and the system of Kagoshima primarily in the extent to which lexical exceptions are permitted: the former system permits a greater number of lexically marked words than the latter two systems.

3.2.2 *Exceptions to the antepenultimate rule*

Returning to the loanword accent in (10a), we must not overlook the fact that this rule admits a fairly large number of exceptions (CHAPTER 106: EXCEPTIONALITY). There are at least three groups of exceptions which can be accounted for by linguistic factors.

The first group consists of unaccented loanwords such as /amerika°/ 'America' and /monariza°/ 'Mona Lisa'. Generally speaking, the ratio of the unaccented pattern is relatively low in loanwords – about 10 percent, according to Sibata (1994) – but this ratio varies greatly, depending on the length and syllable structure of the word. We will consider these phonological factors in detail in §3.3.1.

A second group of exceptions to the antepenultimate accent rule in (10a) consists of trimoraic loanwords made up of a sequence of light and heavy syllables.

Loanwords of this phonological structure fall into two accent classes, those that obey the antepenultimate rule in (10a) and those that do not. Typically, the former class contains an underlying vowel in the initial syllable, while the latter contains an epenthetic vowel in the same position (Kubozono 2001, 2006a).[4] These two cases are illustrated in (12).

(12)	a.	sé.dan	'sedan'	b.	tu.ín	'twin'
		há.wai	'Hawaii'		to.rái	'try'
		dé.byuu	'debut'		do.rái	'dry'
		ká.nuu	'canoe'		gu.rée	'grey'
		gí.taa	'guitar'		bu.rúu	'blue'
		pú.rin	'pudding'		do.róo	'draw'

The examples in (12) may seem to suggest that epenthetic vowels tend to avoid bearing an accent (CHAPTER 67: VOWEL EPENTHESIS), but the facts are not that simple. An epenthetic vowel can, and actually does, bear an accent if it is followed by a sequence of light syllables, as in (13).

(13)	kú.ra.su	'class'
	dó.re.su	'dress'
	pú.ra.mu	'plum'
	dó.ra.mu	'drum'

The irregular accent pattern in (12b) emerges if the following two conditions are met: (i) the initial syllable contains an epenthetic vowel, and (ii) the following syllable is heavy. In other words, a penultimate light syllable attracts word accent more strongly than a final heavy syllable, but this strength relation is reversed if the light syllable contains an epenthetic vowel. This shows an interesting interaction between syllable weight and the nature of the vowel (epenthetic *vs.* non-epenthetic) in accent assignment.

A certain type of four-mora or longer loanword constitutes a third group of exceptions to the antepenultimate rule in (10a). These exceptions typically end in a sequence of a light syllable plus a heavy one, as exemplified in (14): σ_H and σ_L stand for heavy (bimoraic) and light (monomoraic) syllables, respectively. Words in (14a) end in a sequence of light syllables followed by a heavy syllable, whereas those in (14b) involve a heavy–light–heavy sequence in the same position. Unlike the words in (10b), these loanwords are accented not on the syllable containing the antepenultimate mora, but one syllable to the left.

(14)	a.	$\sigma_L\sigma_L\sigma_H$ #		b.	$\sigma_H\sigma_L\sigma_H$#	
		bí.gi.naa	'beginner'		ín.ta.byuu	'interview'
		tó.ro.fii	'trophy'		góo.ri.kii	'Gorky (Russian writer)'
		sú.ri.raa	'thriller'		báa.ku.ree	'Berkeley'

Statistical study has shown that this "pre-antepenultimate" accent pattern is much more common than the antepenultimate pattern in (15) for the two prosodic structures (Kubozono 1996; Katayama 1998; Shinohara 2000; Tanaka 2008).

[4] /ta.búu/ 'taboo' is one of the few words that show final accent even though the initial vowel is not epenthetic.

(15) a. bi.tá.min 'bitamin' b. ren.tó.gen 'X rays'
a.sé.an 'ASEAN' baa.bé.kyuu 'barbecue'
pi.ká.tyuu 'Pikachu' kuu.dé.taa 'a coup'

Not surprisingly, many words of the phonological structures in question fluctuate between the two accent patterns, as shown in (16). As the pre-antepenultimate pattern tends to be preferred by young speakers, the variations in (16) suggest that the pre-antepenultimate pattern in (14) represents a newer accent pattern and is replacing the traditional antepenultimate pattern shown in (15).

(16) a. do.rá.gon ~ dó.ra.gon 'dragon'
a.re.rú.gii ~ a.ré.ru.gii 'allergy'
re.bá.non ~ ré.ba.non 'Lebanon'
b. myuu.zí.syan ~ myúu.zi.syan 'musician'
en.dé.baa ~ én.de.baa '(space shuttle) Endeavor'
kaa.di.gan ~ káa.di.gan 'cardigan'

3.2.3 *Comparison with the Latin rule*

It is important to point out here a striking similarity between the Japanese antepenultimate rule in (10a) and the accent rule of Latin in (17) (Kubozono 1996, 1999). The Latin accent rule is known to be shared by English and many other languages of the world (Hayes 1995).

(17) *Latin accent rule* (Prince and Smolensky 1993; Hayes 1995)

Accent falls on the penultimate syllable if it is heavy; otherwise, it falls on the antepenultimate syllable (whether it is heavy or light).

Although the two rules might look quite different at a glance, they have in fact much in common. This can be seen from (18) and (19), where the effects of the two rules are compared with reference to the three syllables in word-final position:[5] underlined syllables are those to which an accent is assigned by each rule.

(18) *Effects of the Japanese antepenultimate accent rule in* (10a)

a. $\ldots\sigma_H\underline{\sigma_H}\sigma_H\#$ e. $\ldots\sigma_H\underline{\sigma_L}\sigma_H\#$
b. $\ldots\sigma_H\underline{\sigma_H}\sigma_L\#$ f. $\ldots\underline{\sigma_H}\sigma_L\sigma_L\#$
c. $\ldots\sigma_L\underline{\sigma_H}\sigma_H\#$ g. $\ldots\sigma_L\underline{\sigma_L}\sigma_H\#$
d. $\ldots\sigma_L\underline{\sigma_H}\sigma_L\#$ h. $\ldots\underline{\sigma_L}\sigma_L\sigma_L\#$

(19) *Effects of the Latin accent rule in* (17)

a. $\ldots\sigma_H\underline{\sigma_H}\sigma_H\#$ e. $\ldots\underline{\sigma_H}\sigma_L\sigma_H\#$
b. $\ldots\sigma_H\underline{\sigma_H}\sigma_L\#$ f. $\ldots\underline{\sigma_H}\sigma_L\sigma_L\#$
c. $\ldots\sigma_L\underline{\sigma_H}\sigma_H\#$ g. $\ldots\underline{\sigma_L}\sigma_L\sigma_H\#$
d. $\ldots\sigma_L\underline{\sigma_H}\sigma_L\#$ h. $\ldots\underline{\sigma_L}\sigma_L\sigma_L\#$

[5] Superheavy, i.e. trimoraic, syllables are excluded here, since there are severe restrictions on this structure in Japanese, just as in many other languages. See Kubozono (1999) for a full discussion.

	hime	'princess'	kaguyá-hime	'Princess Kaguya'
			sirayukí-hime	'Snow White'
	sikí	'ceremony'	sotugyóo-siki	'graduation ceremony'
			nyuugakú-siki	'entrance ceremony'
	ko°	'child'	tínomí-go	'infant'
			minasí-go	'orphaned child'
	musi°	'bug, insect'	kabutó-musi	'beetle'
			tentóo-musi	'ladybird'
b.	*Initial-accenting morphemes*			
	néko	'cat'	perusya-néko	'Persian cat'
			maneki-néko	'cat with a beckoning paw'
	gásu	'gas'	metan-gásu	'methane gas'
			puropan-gásu	'propane gas'
c.	*De-accenting morphemes*			
	iró	'color'	orenzi-iro°	'orange color'
			nezumi-iro°	'gray'
	tóo	'party'	minsyu-too°	'Democratic Party'
			kyoowa-too°	'Republican Party'
	gó	'word, language'	nihon-go°	'Japanese language'
			tyuugoku-go°	'Chinese language'

Under this analysis, all monomoraic and bimoraic morphemes are marked in the lexicon as to their accentual behavior in compounds. An alternative view to this idea of exhaustive specifications in the lexicon is that only morphemes of the type in (20c) are lexically marked, while other types of morphemes lack such lexical specifications (Kubozono 1997). This underspecification analysis is based on the general observation that all "initial-accenting" morphemes are invariably initially accented when they are pronounced in isolation (although not necessarily vice versa),[6] as well as the observation that (20a) represents the default accent pattern of compound nouns with a short N_2 (Akinaga 1985).

Adopting these two observations, the two compound accent patterns in (20a) and (20b) can be largely computed by rule or constraint interaction in an optimality-theoretic framework (Prince and Smolensky 1993). Descriptively, (20a) and (20b) can be derived by the rules in (21), which are based on two notions, the syllable and the (bimoraic) foot (Kubozono 1997).[7,8]

[6] Some bimoraic morphemes fail to keep their non-final accent in compounds and attract a default compound accent immediately before them instead: e.g. /siráyuki/ 'white snow' + /híme/ 'princess' → /sirayuki-hime/, */sirayuki-hime/ 'Snow White'. Initially accented morphemes of this type must be marked in the lexicon.

[7] Poser (1990) introduces the notion of foot into the analysis of Japanese accent, to account for the accent patterns shown by compound nouns with a long N_2, in particular. He specifically assumes that the final bimoraic foot is invisible to the compound accent rule that applies to compound nouns with a long N_2. However, the generalization of compound accent proposed in this chapter is different from Poser's in assuming that N_2's accent on its penultimate mora can be preserved in compounds unless it is on the very final syllable.

[8] The rule in (21b) is here stated descriptively, and can be formally understood as resulting from the interaction of several constraints, including NON-FINALITY and FOOTFORM, as well as accentual constraints (see Kubozono 1995, 1997, and Itô *et al.* 1996 for a proposal in this direction).

(24) a. ti[nomi]-go
in[dasú]-[gawa]
sotu[gyóo]-[siki]
b. yama-[óto]ko
minami-[áme][rika]

In terms of lexical specification, none of the N_2s in these compound nouns need to be marked in the lexicon with respect to the compound accent behavior they exhibit. Rather, their compound accent behavior can be readily computed on the basis of their phonological structure. All in all, compound nouns with a short N_2 and those with a long N_2 can be grouped together with respect to their compound accent patterns. Those with a short N_2 permit lexical exceptions to the generalization in (21), i.e. unaccented compounds like (20c), but their accent patterns are otherwise the same as those of compound nouns with a long N_2. We will develop this generalization further in §3.3.4, where unaccented compounds are discussed.

Finally, the descriptive generalization in (21) has a further advantage in allowing us to capture the fundamental similarity between the compound accent rules and the accent rule for simplex nouns, i.e. the antepenultimate rule given in (10a). In fact, the antepenultimate rule has basically the same effects as the rule in (21b), as can be seen from the foot structure of the loanwords in (25) (= (10b)). This is not accidental, but simply reflects the fact that the accent rule for accented simplex nouns and that for morphologically complex nouns have basically the same nature: they both put an accent maximally towards the end of the word (edgemostness), while avoiding the final position aligned with the right edge of the word (non-finality).

(25) [bá.na]na wa[sin][ton]
[ká.na]da [ba.do][min][ton]
o[rén]zi [pai][náp][pu.ru]

In sum, the accent rules for compound nouns and the rule for accented simplex nouns only differ in (21a): simplex words are free from this faithfulness principle, which is a natural consequence resulting from the morphological non-complexity of simplex nouns.

3.3 *Unaccented words*

I have so far deferred the discussion of unaccented words, words that are pronounced with a flat pitch pattern even when a particle is attached to them. About half of the vocabulary of Tokyo Japanese belongs to this accent category (Hayashi 1982: 331). Presence or absence of a lexical pitch accent is contrastive, and in fact accounts for most homophonous word pairs that can be accentually distinguished: e.g. /áme/ 'rain' *vs.* /ame°/ 'candy', /haná/ 'flower' *vs.* /hana°/ 'nose', /sénsi/ 'soldier' *vs.* /sensi°/ 'death in battle'. The main question in this research area then is when the seemingly peculiar class of unaccented words occurs and to what extent their distribution can be predicted by linguistic factors.

The distinction between accented and unaccented words seems arbitrary and item-specific in the native and SJ strata of the lexicon, as can be seen from the homophonous word pairs given above and below: /náru/ 'to be completed' *vs.*

(34)	a.	*Tone A*	
		A.ka	'red'
		a.KA-pen	'red pen'
		a.ka-SIN.goo	'red signal'
		aka-en.PI.tu	'red pencil'
		aka-en.pi.tu-MON.dai	'red pencil problem'
	b.	*Tone B*	
		a.O	'blue, green'
		a.o-PEN	'blue pen'
		a.o-sin.GOO	'green signal'
		a.o-en.pi.TU	'blue pencil'
		a.o-en.pi.tu-mon.DAI	'blue pencil problem'

This rule is, in a sense, a mirror-image of the compound accent rule in Tokyo Japanese. In Tokyo, the compound accent rule only considers the phonological structure of the final constituent (N_2), so that compounds basically take one and the same accent pattern if they share an N_2 (see (20) and (27), for example).

Kagoshima Japanese employs the opposite strategy, whereby it considers the initial constituent, keeping its prosodic feature in compounds. Interestingly, the tonal feature of the initial morpheme is realized at the very end of the compound in this system. Take the two morphemes /A.ka/ and /a.O/ in (34), for example. Compound nouns starting with /A.ka/ bear a high pitch on their penultimate syllable, no matter how long they may be. Similarly, compounds starting with /a.O/ are high-pitched on their final syllable. Since compound nouns can consist of three or more morphemes and hence can be very long, the speaker must remember the tonal property of the initial morpheme up to the very end of long expressions.

Finally, Hirayama's Law also applies to phrasal expressions in Kagoshima Japanese. As illustrated in (35), basic syntactic phrases (*bunsetsu*), consisting of a content word plus one or more particles, undergo the compound tone rule. Thus syntactic phrases beginning with a Tone A morpheme all take Tone A, realizing a high pitch on the penultimate syllable within the phrasal domain. Similarly, those beginning with a Tone B morpheme all take Tone B, with a high pitch on their very final syllable.

(35)	a.	A.ka	'red'
		a.KA-ga	'red-NOM'
		a.ka-KA.ra	'from red'
		a.ka-SIN.goo	'red signal'
		a.ka-sin.GOO-ga	'red signal-NOM'
		a.ka-en.PI.tu	'red pencil'
		a.ka-en.pi.tu-KA.ra	'from (the) red pencil'
	b.	a.O	'blue, green'
		a.o-GA	'blue-NOM'
		a.o-ka.RA	'from blue'
		a.o-sin.GOO	'green signal'
		a.o-sin.goo-GA	'green signal-NOM'
		a.o-en.pi.TU	'blue pencil'
		a.o-en.pi.tu-ka.RA	'from (the) blue pencil'

b.	*Tone B*		
	O.to.KO	(o.to.KO)	'man'
	NI.WA.to.RI	(ni.wa.to.RI)	'chicken'
	NI.WA.toI	(ni.wa.TOI)	'chicken (coll. form)'
	SEN.seI	(sen.SEI)	'teacher'
	HA.RU.YA.su.MI	(ha.ru.ya.su.MI)	'spring vacation'

The basic rule employed here is to start words with a high pitch and have one low-pitched portion before the second peak starts (Kubozono 2008a).[21] In autosegmental terms, this means that high tone cannot spread onto a syllable that is adjacent to another high tone.[22] This is true of both tonal types, Tone A and Tone B, which are differentiated basically by the loci of the second peak. It must be noted in this connection that the low-pitched portion is a syllable-sized unit rather than a mora, as shown by the following forms.

(42)	a.	KE.da.MOn	'wild animal'
		KE.da.MOn-ga	'wild animal-NOM'
		KE.DA.mon-KA.ra	'from (the) wild animal'
	b.	E.FU.bii.Ai	'FBI'
		E.FU.bii.Ai-ga	'FBI-NOM'
		E.FU.BII.ai-KA.ra	'from FBI'

There are two interesting facts to note about the second peak. One of them is the fact that non-syllabic moras like the moraic nasal and the second member of long vowels and diphthongs can form the second peak in Tone B words, but not in Tone A words. Thus, word-final moras in /mi.kan/ in (40b), /ni.wa.toi/ and /sen.sei/ in (41b), constitute a second peak on their own, but the penultimate moras in /ke.da.mon-ga/ and /e.fu.bii.ai-ga/ in (42) cannot.[23] In the latter case, the high pitch moves one mora to the left, so that the antepenultimate mora constitutes the second peak on its own. In both Tone A and Tone B, the second peak is always monomoraic, but the two tonal patterns exhibit an asymmetry.

This asymmetry is interesting in itself, but more interesting is the fact that the first peak is always dependent on the second peak. Compare the two forms in (43), the second of which is a colloquial form of the first.

(43)	a.	KE.DA.mo.NO-ga	'wild animal-NOM'
	b.	KE.da.MOn-ga	

Although these two forms consist of the same number of moras, the first high pitch is on /ke.da/ in (43a), but on /ke/ in (43b). This difference directly reflects the difference in the second peak, which is on /no/ in (43a), but on /mo/ in (43b). This indicates that the domain of the first high pitch can be computed on the

21 Kubozono's (2008a) work is based on his fieldwork data obtained in Teuchi, a small village located at the southern edge of the Koshikijima Islands.

22 Myers (1990) reports a similar phenomenon in Shona, a Bantu language spoken in Zimbabwe.

23 Consequently, the penultimate mora is high in /KE.DA.mo.NO-ga/ 'wild animal-NOM', but the antepenultimate mora is high in its colloquial form /KE.da.MOn-ga/.

basis of the position of the second peak. The whole strategy can be summed up as follows.

(44) a. Assign a high pitch (second peak) to the final mora in Tone B words and to the penultimate mora in Tone A words.
b. If the second peak is on a non-syllabic mora in Tone A words, the peak moves one mora to the left, i.e. to the syllabic mora of the same syllable.
c. Keep one syllable low-pitched before the second peak.
d. Assign a high pitch (first peak) to word-initial syllables up to the low-pitched syllable defined in (c).

This right-to-left computation means that the second peak phonologically dominates the first peak. This may not be very surprising in itself, but its real significance lies in the fact that the dominance relationship between the two peaks is reversed in sentence-level phonology. In a sequence of phrases, every phrase except the sentence-final one can have only one peak and it is always the second peak that disappears (Kubozono 2008a). This is illustrated in (45), where some nouns in (41) are followed by the nominative marker /ga/ and put in the carrier sentence /...ga mieta/ '...was seen' (/mi.e.ta/ is a Tone B verb). The forms in parentheses represent the tonal forms when the noun-NOM phrases are pronounced in isolation.

(45) a. *Tone A*

KE.DA.mo.no-ga MI.e.TA	'a wild animal was seen'
(KE.DA.mo.NO-ga)	'wild animal-NOM'
KE.da.mon-ga MI.e.TA	'a wild animal was seen'
(KE.da.MOn-ga)	'wild animal-NOM (coll)'

b. *Tone B*

NI.WA.TO.ri-ga MI.e.TA	'a chicken was seen'
(NI.WA.TO.ri-GA)	'chicken-NOM'
NI.WA.toi-ga MI.e.TA	'a chicken was seen (coll)'
(NI.WA.toi-GA)	'chicken-NOM (coll)'

Words and phrases thus lose one of their two high-pitched portions when they appear in the non-final position of a sentence. It is of special interest that a phonologically dominant peak (the second peak) is defeated by a less dominant peak (the first peak) in sentence-level phonology.

6 Conclusion

In this chapter we have considered some pitch accent systems of Japanese, focusing on the major debates and findings in the past and present. The first half of the chapter discussed the accent system of Tokyo Japanese, the dialect most extensively studied in the literature of Japanese phonetics and phonology. One of the most important findings about this dialect is that accent patterns, including the unaccented one, can be predicted by the linguistic structure of words. Another crucial point is that many seemingly different accent rules can be generalized if we incorporate the foot as a relevant prosodic unit in the language. Specifically, the so-called antepenultimate accent rule and the set of compound accent rules can be generalized with one another and can be expressed by general notions such as non-finality and edgemostness.

The second half of the chapter described the pitch accent systems of Kagoshima and Koshikijima Japanese, to illustrate the diversity of prosodic systems in the language. Unlike Tokyo Japanese, these dialects have a two-pattern system, a system that exhibits only two distinctive pitch patterns irrespective of the length of the word.

There are many important issues that must be addressed in the future. In empirical studies, the most important issue concerns endangered dialects and their prosodic systems. While some dialects such as Tokyo Japanese and Kagoshima Japanese have been studied in considerable depth in the literature, there are many regional dialects that remain largely unstudied or even undocumented. Extensive data must be collected about these dialects before they become extinct. Serious analyses of these dialects may shed a new light on the typology of Japanese pitch accent systems, as well as on the typology of prosodic systems in general. It is equally important to look at the endangered dialects from a sociolinguistic perspective, by examining how their pitch accent systems change under exposure to standard Tokyo Japanese.

There are also interesting questions that remain unanswered about the theoretical aspect of Japanese pitch accent. One of them is how to formalize pitch accent rules of the language in the non-derivational framework of Optimality Theory (Kubozono 1995, 1997; Katayama 1996; Shinohara 2000). How to account for the underspecification of lexical accent in a theoretical framework is also an important issue for future work.

REFERENCES

Akinaga, Kazue. 1981. Akusento syuutoku hoosoku [Laws of accent learning]. Appendix to *Meikai nihongo akusento ziten* [*Japanese accent dictionary*]. Tokyo: Sanseido.

Akinaga, Kazue. 1985. Kyootuugo no akusento [Accent of standard Japanese]. Appendix to NHK (1985).

Beckman, Mary E. 1986. *Stress and non-stress accent*. Dordrecht: Foris.

Haraguchi, Shōsuke. 1977. *The tone pattern of Japanese: An autosegmental theory of tonology*. Tokyo: Kaitakusha.

Haraguchi, Shōsuke. 1991. *A theory of stress and accent*. Dordrecht: Foris.

Hayashi, Ooki (ed.) 1982. *Zusetu nihongo* [*Illustrated Japanese*]. Tokyo: Kadokawa.

Hayata, Teruhiro. 1999. *Ontyoo no taiporozii* [*Typology of prosodic systems*]. Tokyo: Taishukan.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hirayama, Teruo. 1951. *Kyuusyuu hoogen ontyoo no kenkyuu* [*Studies on the tone of the Kyushu dialects*]. Tokyo: Gakkai-no-shishin-sha.

Hirayama, Teruo. 1960. *Zenkoku akusento ziten* [*All Japan accent dictionary*]. Tokyo: Meiji Shoin.

Igarashi, Yosuke. Forthcoming. Typology of intonational phrasing in Japanese dialects. In Sun-Ah Jun (ed.) *Prosodic typology*, vol. 2. Oxford: Oxford University Press.

Itô, Junko, Yoshihisa Kitagawa & Armin Mester. 1996. Prosodic faithfulness and correspondence: Evidence from a Japanese argot. *Journal of East Asian Linguistics* 5. 217–294.

Kamimura, Takaji. 1941. Koshikijima hoogen no akusento [The accent of the Koshikijima dialect]. *Onseigaku Kyokai Kaiho* 65/66. 12–15.
Katayama, Motoko. 1998. Optimality Theory and Japanese loanword phonology. Ph.D. dissertation, University of California, Santa Cruz.
Kibe, Nobuko. 2000. *Seinanbu kyuushuu nikei akusento no kenkyuu* [*Studies on the two-class-accent dialects of southwest Kyushu*]. Tokyo: Benseisha.
Kindaichi, Haruhiko. 1967. *Nihongo on'in no kenkyuu* [*Studies on Japanese phonology*]. Tokyo: Tōkyōdō-shuppan.
Kindaichi, Haruhiko (ed.) 2001. *Shinmeikai nihongo akusento ziten* [*Shinmeikai Japanese accent dictionary*]. Tokyo: Sanseido.
Kubozono, Haruo. 1988. The organization of Japanese prosody. Ph.D. dissertation, University of Edinburgh. Published 1993, Tokyo: Kurosio.
Kubozono, Haruo. 1995. Constraint interaction in Japanese phonology: Evidence from compound accent. In Rachael Walker, Ove Lorentz & Haruo Kubozono (eds.) *Phonology at Santa Cruz* 4. Santa Cruz: Linguistics Research Center, University of Calfornia, Santa Cruz, 21–38.
Kubozono, Haruo. 1996. Syllable and accent in Japanese: Evidence from loanword accentuation. *The Bulletin of the Phonetic Society of Japan* 211. 71–82.
Kubozono, Haruo. 1997. Lexical markedness and variation: A non-derivational account. *Proceedings of the West Coast Conference on Formal Linguistics* 15. 273–287.
Kubozono, Haruo. 1999. Mora and syllable. In Natsuko Tsujimura (ed.) *The handbook of Japanese linguistics*, 31–61. Malden, MA & Oxford: Blackwell.
Kubozono, Haruo. 2001. Epenthetic vowels and accent in Japanese: Facts and paradoxes. In Jeroen van de Weijer & Tetsuo Nishihara (eds.) *Issues in Japanese phonology and morphology*, 113–142. Berlin & New York: Mouton de Gruyter.
Kubozono, Haruo. 2003. Accent of alphabetic acronyms in Tokyo Japanese. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata & Shin-ichi Tanaka (eds.) *A new century of phonology and phonological theory*, 356–370. Tokyo: Kaitakusha.
Kubozono, Haruo. 2006a. Where does loanword prosody come from? A case study of Japanese loanword accent. *Lingua* 116. 1140–1170.
Kubozono, Haruo. 2006b. *Akusento no hoosoku* [*Laws of accent*]. Tokyo: Iwanami.
Kubozono, Haruo. 2007. Tonal change in language contact: Evidence from Kagoshima Japanese. In Riad & Gussenhoven (2007), 323–351.
Kubozono, Haruo. 2008a. The phonetics and phonology of tone in an endangered Japanese dialect: Facts and paradoxes. *Proceedings of the 18th International Congress of Linguists*, vol. 2, 205–206. Seoul: Korea University.
Kubozono, Haruo. 2008b. Japanese accent. In Shigeru Miyagawa & Mamoru Saito (eds.) *Handbook of Japanese linguistics*, 165–191. Oxford: Oxford University Press.
Kubozono, Haruo, Junko Itô & Armin Mester. 1997. On'in koozoo kara mita go to ku no kyookai. [The phonological boundary between the word and the phrase.] In Spoken Language Research Group (ed.) *Bunpoo to onsei* [*Speech and grammar*], 147–166. Tokyo: Kurosio.
Maekawa, Kikuo. 1994. Intonational structure of Kumamoto Japanese: A perceptual validation. *Proceedings of the 3rd International Conference on Spoken Language Processing*, 119–122. Yokohama: Acoustical Society of Japan.
McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague & Paris: Mouton.
McCawley, James D. 1978. What is a tone language? In Victoria A. Fromkin (ed.) *Tone: A linguistic survey*, 113–131. New York: Academic Press.
Myers, Scott. 1990. *Tone and the structure of words in Shona*. New York: Garland.
NHK. 1985. *NHK pronunciation and accent dictionary*, 2nd edn, 1998. Tokyo: Nihon Hoso Shuppan Kyokai.

121 Slavic Palatalization

Jerzy Rubach

1 Introduction

Slavic languages more than any other languages contribute to the understanding of palatalization processes (see also CHAPTER 71: PALATALIZATION). The historical source language – Common Slavic – exhibited several different types of palatalization, and these are reflected in the phonology of its descendants. Palatalization can affect all consonants, producing changes that are either allophonic or phonemic. The determination of whether a change is allophonic or phonemic depends on the system of contrasts in a given language. Looking at /s z t d/ as examples, the identical outcomes have a different status in Russian and Polish. I call the type of palatalization exhibited in (1) Surface Palatalization. The environments are discussed in §2.

(1) *Surface Palatalization*

a. *Russian (phonemic palatalization)*
s z t d → s^j z^j t^j d^j

b. *Polish (allophonic palatalization)*
s z t d → s^j z^j t^j d^j

The difference between (1a) and (1b) is that palatalized consonants are members of the phonemic inventory in Russian but not in Polish. In Polish, a phonemic effect of palatalization for the inputs in (1) yields prepalatal fricatives and affricates. I call this process Coronal Palatalization.[1]

(2) *Coronal Palatalization in Polish*
s z t d → ɕ ʑ tɕ dʑ

How is it possible for Polish to have both types of palatalization: allophonic (1b) and phonemic (2)? The question is intriguing because the environments in which

[1] Coronal Palatalization did not exist in Common Slavic. It is a later development that occurred in some Slavic languages.

(1b) and (2) apply are virtually the same (see §2). Answering this question is a challenge for the researcher, regardless of what theoretical framework is assumed.

Velars can undergo two different kinds of palatalization in addition to Surface Palatalization (3a). I will identify them by the traditional names: First Velar Palatalization and Second Velar Palatalization. The example in (3) is the velar stop /k/.

(3) a. *Surface Palatalization*
k → kʲ
b. *First Velar Palatalization*
k → ʧ
c. *Second Velar Palatalization*
k → ts

The challenge is how to distinguish the three kinds of palatalization in (3) when they occur in overlapping environments in a single language.[2] The problem is compounded by the fact that the outputs of First Velar Palatalization and Second Velar Palatalization do not always bear the feature [−back] that is expected to characterize the outcome of a palatalization process.

Palatalization processes may differ in different Slavic languages in all four relevant ways: with regard to the input, the output, the environment, and the domain. For example, in Polish, the inputs to Coronal Palatalization (2) are dental fricatives, stops, and sonorants, while in Slovak, the class of inputs is limited to /t d n l/. In Czech, on the other hand, the inputs are /t d n r/, so /r/ but not /l/ undergoes palatalization.

Looking at /t d/ as examples of inputs to Coronal Palatalization, the outputs in Slovak and Czech are the prepalatal stops [t̠ʲ d̠ʲ], where underlining denotes that the stops are prepalatal rather than dental. In Eastern Polish /t d/ change into dental affricates [tsʲ dzʲ], whereas in Standard Polish they change into prepalatal affricates [tɕ dʑ].

The environments for palatalization are invariably front vowels and glides (see also CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS), but languages differ from each other with regard to their front vowel inventory. For example, Slovak has three front vowels /i ɛ æ/ and, predictably, they all cause palatalization (see §4). Russian lacks /æ/, so its palatalization triggers are the vowels /i ɛ/. More intriguing is the situation where only a subset of front vowels in a given language induces palatalization. This is exemplified by Ukrainian, which has /i ɛ/, but only /i/ triggers palatalization. Similarly, Polish Surface Palatalization applies only before /i/ (and /j/) but not before /ɛ/ (see §2). Finally, palatalization before front vowels may be different from palatalization before /j/, a process that is called Iotation (Steele 1973). This is illustrated in (4), where I look at dental obstruents in Polish.

(4) a. *Coronal Palatalization in Polish*
s z t d → ɕ ʑ tɕ dʑ / __ i ɛ

b. *Iotation in Polish*
s z t d → ʃ ʒ ts dz / __ j

[2] The example here is Polish. The added difficulty is that [ʧ] and [ts] are hard (velarized) rather than soft (palatalized) consonants. See the discussion in §4.

§4 illustrates how palatalization and related processes can be analyzed in the current phonological theory. The paradigm selected for the analysis is Optimality Theory. For a review of the data and descriptive generalizations referring to particular Slavic languages, the reader is referred to the collection of studies in de Bray (1980) and Comrie and Corbett (1993).

2 Inventories

This section reviews types of consonants in Slavic languages that are products of palatalization as a historical process. The contexts considered here make it impossible to derive these consonants from their original historical sources. The consequence is that the consonants are phonemes, that is, underlying segments.

Descriptive grammars of Slavic languages draw a distinction between "soft" and "hard" consonants. Soft consonants are a product of palatalization, either historically or synchronically in the contemporary Slavic languages. Other consonants are hard. The point is illustrated by Russian. Soft consonants exhibit palatalization whereas hard consonants show velarization. There are no neutral consonants that would be neither palatalized nor velarized. Palatalization and velarization are secondary articulations as the raising of the tongue body that they involve is a gesture that does not affect the primary place of articulation. For example, the soft [d^j] and hard [d] in *d'ad'a* 'uncle' and *da* 'yes', respectively, are dental rather than palatal or velar. Palatalization means that, in the example at hand, the production of the dental stop is accompanied by the simultaneous raising of the front part of the tongue toward the hard palate. Velarization is the raising of the back part of the tongue toward the velum. In terms of distinctive features, palatalized consonants are [−back], while velarized consonants are [+back], a distinction that reflects the part of the tongue that is raised. I will follow the Slavic tradition in leaving velarization unmarked, so [d^j] is palatalized while [d] is velarized.

In the typical case, velarized and palatalized consonants are paired, that is, they constitute minimal or near-minimal pairs that contrast in terms of [±back]. This is illustrated by Russian labials and dentals (Halle 1959; Avanesov 1968).[5]

(5) *Velarized–palatalized contrasts in Russian*

[p]	*pat*	'stalemate'	[p^j]	*p'at'*	'five'
[b]	*burak*	'beetroot'	[b^j]	*b'uro*	'office'
[f]	*forma*	'form'	[f^j]	*šof'or*	'driver'
[v]	*vazon*	'vase'	[v^j]	*v'az*	'elm'
[m]	*maslo*	'butter'	[m^j]	*m'aso*	'meat'
[t]	*tot*	'that'	[t^j]	*t'ot'a*	'aunt'
[d]	*da*	'yes'	[d^j]	*d'ad'a*	'uncle'
[s]	*sud*	'court'	[s^j]	*s'uda*	'here'
[z]	*zad*	'back'	[z^j]	*z'at'*	'son-in-law'
[l]	*luk*	'bow'	[l^j]	*l'uk*	'hatch'
[r]	*rasa*	'race'	[r^j]	*r'asa*	'cassock'

[5] I use the term "dental" to refer to both dentals proper and alveolars. That is, dentals are [+anterior] coronals.

skład	[t]	'storage room'	*składzi+e*	[dʑɛ]	(LOC SG)
			składz+ik	[dʑi]	(DIM)
brud	[t]	'dirt'	*brudzi+e*	[dʑɛ]	(LOC SG)
			brudz+i+ć	[dʑi]	'to make dirty'
kran	[n]	'faucet'	*krani+e*	[ɲɛ]	(LOC SG)
			kran+ik	[ɲi]	(DIM)
dzwon	[n]	'bell'	*dzwoni+e*	[ɲɛ]	(LOC SG)
			dzwon+i+ć	[ɲi]	'to ring'

The palatalization of velars is ascribed to two different rules that apply to the same set of targets but yield different results. Following the traditional parlance, I will refer to these rules as First Velar Palatalization and Second Velar Palatalization. The latter is restricted morphologically, making the former the default palatalization for velars.

First Velar Palatalization turns velars into post-alveolar stridents. This is illustrated by Kashubian, where the outputs are soft segments, as would be expected of a palatalization process (Breza and Treder 1981, Treder 2001, and my fieldwork).

(18) *Kashubian First Velar Palatalization*

k g x → tʃʲ dʒʲ ʃʲ / __ i ɛ

kalek+a	[k]	'invalid'	*kalecz+i*	[tʃʲi]	(NOM PL)
drog+a	[g]	'road'	*drodż+i*	[dʒʲi]	(NOM PL)
much+a	[x]	'fly (N)'	*musz+i*	[ʃʲi]	(NOM PL)
bek	[k]	'bleating'	*becz+e+ć*	[tʃʲɛ]	'to bleat'

Second Velar Palatalization yields dental rather than post-alveolar stridents. In Ukrainian, underlying /k ɣ x/ surface as [tsʲ zʲ sʲ] before the suffix /i/ of the dative/locative singular (Medushevs'kyj and Zyatovs'ka 1963; Bilodid 1969; Rusanovs'kyj *et al.* 1986).

(19) *Ukrainian Second Velar Palatalization*

k ɣ x → tsʲ zʲ sʲ / __ i $]_{\text{DAT/LOC}}$

nom sg		*dat/loc sg*		
ruk+a	[k]	*ruc+i*	[tsʲi]	'hand'
noh+a	[ɦ][12]	*noz+i*	[zʲi]	'leg'
much+a	[x]	*mus+i*	[sʲi]	'fly'

First Velar Palatalization and Second Velar Palatalization are active in all Slavic languages with the exception of Russian, which has the former but not the latter rule. Since the two rules are in competition if velars are inputs, the conflict is resolved by introducing partial morphological conditioning of Second Velar Palatalization. In Ukrainian, the morphological context is that of the dative and the locative declension cases. Other Slavic languages may be subject to other morphological limitations. For example, in Polish, Second Velar Palatalization applies before the /i/ suffix of the nominative plural, the /ɛ/ suffix of the dative and the locative singular (as in Ukrainian) and before the adverb-forming /ɛ/ suffix (see Rubach 1984).

[12] I assume that the surface laryngeal [ɦ] derives from the voiced velar fricative /ɣ/. The arguments parallel those for Slovak /ɣ/ → [ɦ] presented in Rubach (1993).

(30) rux+i+tɕ

ruʃʲitɕ	First Velar Palatalization	x → ʃʲ / __ i
ruʃitɕ	Hardening	ʃʲ → ʃ
ruʃɨtɕ	Retraction	i → ɨ after a hard consonant

The operation of First Velar Palatalization in Polish is not limited to /x/ → [ʃ]. It extends to all velars, as the following examples show.

(31) *Polish velar stops*

a.	*voiceless stop*: k → ʧ					
	krok	[k]	'step (N)'	*krocz+ek*	[ʧ+ɛ]	(DIM)
	skok	[k]	'jump (N)'	*skocz+y+ć*	[ʧ+ɨ]	'to jump'
b.	*voiced stop*: g → ʤ					
	mózg	[k]	'brain'	*móżdż+ek*	[ʤ+ɛ]	(DIM)
	miazg+a	[g]	'pulp'	*miażdż+y+ć*	[ʤ+ɨ]	'to squash'
c.	*voiced stop*: g → ʒ					
	Bóg	[k]	'God'	*Boże*	[ʒ+ɛ]	(VOC SG)
	śnieg	[k]	'snow'	*śnież+ek*	[ʒ+ɛ]	(DIM)
				śnież+y+ć	[ʒ+ɨ]	'to snow'
	sług+a	[g]	'servant'	*służ+eb+n+y*	[ʒ+ɛ]	(ADJ)
				służ+y+ć	[ʒ+ɨ]	'to serve'

The data in (27), (31a), and (31b) show that Polish First Velar Palatalization is the same as Kashubian First Velar Palatalization (18), but the parallel is obscured by Hardening: /ʃʲ ʧʲ ʤʲ/ → [ʃ ʧ ʤ]. The data in (31c) add a new generalization: /ʤ/ is spirantized to [ʒ]. The rule applies if the segment preceding /ʤ/ is a sonorant, hence we see [ʒ] in *śnież+ek* but [ʤ] in *móżdż+ek* in (31c) and (31b), respectively. The [ʒ] derived from /ʤʲ/ is hard, showing that hardening affects all post-alveolars.

(32) *Polish Hardening*

ʃʲ ʒʲ ʧʲ ʤʲ → ʃ ʒ ʧ ʤ

Polish also has a process of Spirantization. Since there is no way to tell whether the rule applies before or after Hardening, the choice between /ʤʲ/ → [ʒʲ] and /ʤ/ → [ʒ] is arbitrary. Assuming that the latter is true, the rule is stated schematically as follows (Rubach 1984).

(33) *Polish Spirantization*

ʤ → ʒ / [+son] __

Spirantization is a derived environment rule in the sense that it applies only to the /ʤ/ that is a result of First Velar Palatalization. The /ʤ/ that comes from the underlying representation resists Spirantization. This is exactly what Kiparsky's (1973, 1982) Derived Environment Constraint (DEC) predicts (see CHAPTER 88: DERIVED ENVIRONMENT EFFECTS). Its instantiation that is relevant here refers to phonologically derived segments. In (34), I look at the derivation of *Boż+e* 'God (VOC SG)' and *brydż* 'bridge (card game)', whose underlying representations are representations because we find velarized rather than palatalized consonants as outputs of palatalization.

The opacity of palatalization is also manifested by the fact that palatalization processes appear to have exceptions in the surface representation. That is, they are not surface true. The problem is illustrated by Polish.

Recall that Polish Coronal Palatalization turns, among other consonants, /s z t d/ into [ɕ ʑ tɕ dʑ] before front vowels, as documented in (17). The process is not observed on the surface in hundreds of words of both native and foreign origin.

(40) *"Exceptions" to Polish Coronal Palatalization*

a.	*ser*	[sɛ]	'cheese'		
	zegar	[zɛ]	'clock'		
	deszcz	[dɛ]	'rain'		
b.	*seps+a*	[sɛps+a]	'sepsis'	*sepsi+e*	[sɛpɕ+ɛ] (LOC SG)
	zez	[zɛs]	'squint'	*zezi+e*	[zɛʑ+ɛ] (LOC SG)
	internet	[intɛrnɛt]	'Internet'	*interneci+e*	[intɛrnɛtɕ+ɛ] (LOC SG)
	dywidend+a	[divʲidɛnd+a]	'dividend'	*dywidendzi+e*	[divʲidɛɲdʑ+ɛ] (LOC SG)

It appears that the words in (40a) are simply exceptions to Coronal Palatalization and should be encoded as such in the lexical representation. This reasoning is faulty, however. The words in (40b) appear to be simultaneously exceptions and regular inputs to Coronal Palatalization. Looking at the word for 'squint', the underlying representation of the locative singular form is /zɛz+ɛ/, with two instances of /z/ followed by /ɛ/. The surface effect is contradictory: one /z/ palatalizes to [ʑ] but the other does not: [zɛʑ+ɛ]. If the morpheme /zɛz/ were an exception, then none of the /z/s should palatalize, yielding *[zɛz+ɛ]. On the other hand, if /zɛz/ were a regular input to Coronal Palatalization, then both of the /z/s should palatalize before /ɛ/, yielding *[ʑɛʑ+ɛ]. Since none of these predictions is true, the generalization concerning Coronal Palatalization must be different from the default assumption that the rule applies whenever its environment is met.

As mentioned above, Kiparsky (1973, 1982) observes that phonological rules may be sensitive to derived environments. The part of the Derived Environment Constraint that is relevant here refers to morphologically derived environments, that is, to concatenations of segments that arise due to the application of a word formation rule. In effect, then, the Derived Environment Constraint limits the application of the rule to the portion of the string that strides a morpheme boundary. This requirement is fulfilled by the second /z/ and /ɛ/ in /zɛz+ɛ/, because the input /z/ and the environment /ɛ/ occur across a morpheme boundary: /z+ɛ/. This structure counts as "derived" because it is not found in the morpheme /zɛz/ listed in the lexicon. Rather, the input to Coronal Palatalization is derived via the application of the word formation rule (WFR) that adds the locative singular suffix /ɛ/. Schematically:

Standard Polish (17). Recall that Slovak coronals surface as prepalatals, as in *advokat* [t] 'lawyer' – *advokat+ik* [t̠ʲ] (DIM). The observation is that palatalization induces a shift in the place of articulation from [+anterior], which characterizes dentals, to [–anterior], which characterizes posteriors. The markedness constraint enforcing this change is Posterior (Rubach 2003a).

(48) Posterior
Palatalized coronals must be [–anterior].

The change from [+anterior] to [–anterior] that Posterior asks for violates the faithfulness constraint Ident[+anterior].

(49) Ident[+anterior]
[+anterior] on an input consonant must be preserved as [+anterior] on a correspondent of that consonant in the output.

The effect of Coronal Palatalization in Eastern Polish is different from that found in Slovak: dental stops are affricated without changing the place of articulation, as shown by the alternation *bat* [bat] 'whip' – *bac+ik* [batsʲik]. Affrication is enforced by Stridency (Rubach 2003a) at the cost of violating the faithfulness constraint Ident[–strident], both of which are stated in (50).

(50) a. Stridency
Palatalized coronals must be [+strident].[24]

b. Ident[–strident]
[–strident] on an input consonant must be preserved as [–strident] on a correspondent of that consonant in the output.

The pattern attested in Standard Polish, /t/ → [tɕ], as in *bat* [bat] 'whip' – *bac+ik* [batɕik] (DIM), is a combined effect of Posterior and Stridency.

The various permutations of these constraints generate the outputs found in Slovak (51), Eastern Polish (52), and Standard Polish (53). The Russian example from (47) is repeated in (54) to clarify how the evaluation works once Posterior and Stridency have been added to the constraint set. To keep the tableaux within reasonable bounds, I ignore the faithfulness constraints in (46) and the candidates that would violate them (see the interaction displayed in (47) above).

(51) *Coronal Palatalization effects in Slovak*: advokat+ik 'lawyer (DIM)'
No affrication, so Ident[–strid] >> Stridency, but a change of the place of articulation, so Posterior>> Ident[+ant].

/advɔkat+ik/	Pal-*i*	Posterior	Ident[+ant]	Ident[–strid]	Strid
a. advɔkatik	*!				
b. advɔkatʲik		*!			*
c. advɔkatsʲik		*!			
☞ d. advɔkat̠ʲik			*		*
e. advɔkatɕik			*	*!	

[24] Since strident sonorants are universally unattested, GEN does not submit for evaluation candidates that contain such segments. That is, the prohibition against strident sonorants is a constraint on GEN.

(52) *Coronal Palatalization effects in Eastern Polish*: bac+ik 'whip (DIM)'

Affrication, so Stridency >> Ident[−strid], but no change of the place of articulation, so Ident[+ant] >> Posterior.

/bat+ik/	Pal-*i*	Ident[+ant]	Posterior	Strid	Ident[−strid]
a. batik	*!				
b. batʲik			*	*!	
☞ c. batsʲik			*		*
d. bat̠ʲik		*!		*	
e. batɕik		*!			*

(53) *Coronal Palatalization effects in Standard Polish*: bac+ik 'whip (DIM)'

Affrication, so Stridency >> Ident[−strid] and a change of the place of articulation, so Posterior >> Ident[+ant].

/bat+ik/	Pal-*i*	Posterior	Ident[+ant]	Strid	Ident[−strid]
a. batik	*!				
b. batʲik		*!		*	
c. batsʲik		*!			*
d. bat̠ʲik			*	*!	
☞ e. batɕik			*		*

(54) *Coronal Palatalization effects in Russian*: brat+ik 'brother (DIM)'

No affrication, so Ident[−strid] >> Stridency and no change of the place of articulation, so Ident[+ant] >> Posterior.

/brat+ik/	Pal-*i*	Ident[+ant]	Posterior	Ident[−strid]	Strid
a. bratik	*!				
☞ b. bratʲik			*		*
c. bratsʲik			*	*!	
d. brat̠ʲik		*!			*
e. bratɕik		*!		*	

Standard Polish challenges OT by admitting two contradictory patterns: Coronal Palatalization and Surface Palatalization. This is exemplified by near-minimal contrastive pairs in (55).

(55) *Polish Coronal Palatalization and Surface Palatalization*

a.	*but*	[t]	'shoe'		*buc+ik*	[tɕi]	(DIM)
				vs.	*but Iwana*	[tʲ#i]	'Ivan's shoe'
b.	*głos*	[s]	'voice'		*głos+ik*	[ɕi]	(DIM)
				vs.	*głos Iwana*	[sʲ#i]	'Ivan's voice'
c.	*kran*	[n]	'faucet'		*kran+ik*	[ɲʲi]	(DIM)
				vs.	*kran Iwana*	[nʲ#i]	'Ivan's faucet'

Classic OT cannot account for the contrast in (55) because of a ranking paradox. In order to generate /t/ → [tɕ] in (55a), POSTERIOR and STRIDENCY must dominate IDENT[+ant] and IDENT[−strid], respectively, as shown in (53). But, in order to generate /t/ → [tʲ], the ranking of the constraints must be exactly the opposite: IDENT[+ant] and IDENT[−strid] must outrank POSTERIOR and STRIDENCY, respectively.

The problem with classic OT is that it assumes strict parallelism as one of its fundamental governing principles. Strict parallelism forbids any form of derivationalism, so all evaluation must be carried out simultaneously (McCarthy and Prince 1995; Prince and Smolensky 2004). This tenet is rejected by Derivational Optimality Theory (DOT), a phonological framework developed in the work of Kiparsky (1997, 2000), Rubach (1997, 2000a, 2000b, 2003b, 2004) and others. DOT admits three derivational levels, which are defined morphologically: the stem level, the word level, and the post-lexical level.[25] (The latter takes constituents produced by syntactic operations as its domain, so these constituents are larger than words.) Each level constitutes a separate system with its own constraint ranking. This means that constraints can be reranked between levels, albeit the reranking must be minimal (Rubach 2000b). The winning candidate from an earlier level becomes the input to a later level, its "new underlying representation." The IDENT constraints take this new input as their point of reference for penalizing the disparity between the input and the output. In sum, each level in DOT constitutes a miniphonology with its own inputs and constraint ranking. These are exactly the assumptions that we need in order to account for the Polish data in (45). From the point of view of these data, the relevant distinction is between the word level and the post-lexical level. The input to the post-lexical level is the winner from the word level, and the constraints can be reranked. The ranking paradox encountered by classic OT is now easily resolved.

At the word level, the constraint ranking is as in (53), with POSTERIOR and STRIDENCY mandating the change of the place of articulation and affrication. The input /but+ik/ 'shoe' (DIM) has [butɕik] as its optimal output, exactly as in (53). At the word level, the phrase *but Iwana* 'Ivan's shoe' does not exist. Rather, we have a separate evaluation for each word: one for *but* and the other for *Iwana*, a tenet that DOT has inherited from Lexical Phonology. Since *but* and *Iwana* are not a constituent at the word level, /t/ is not before /i/, so PAL-*i* is not activated.

The situation is different at the post-lexical level. Now the rules of syntax have put together *but* and *Iwana*, which form a phrase and the phrase is subject to evaluation. The /t/ of *but* in *but Iwana* is before /i/, so PAL-*i* is activated. The faithful candidate [but#ivana] is excluded by PAL-*i* because [t] is hard, that is [+back], and [i] is front, that is [−back], so there is no agreement in the value for [±back]; we witness a violation of PAL-*i*. The candidate [butsʲ ivana] is eliminated by IDENT[−strid] >> STRIDENCY. These constraints have changed places at the post-lexical level, where affrication is not longer optimal. At the word level, they were ranked STRIDENCY >> IDENT[−strid] in order to obtain [tɕ] from /t/ in *buc+ik* 'shoe (DIM)'. The candidate [bu t̠ʲ ivana], with prepalatal [t̠ʲ], loses as well. It violates IDENT[+ant], which dominates POSTERIOR at the post-lexical level. This too is a case of reranking because POSTERIOR >> IDENT[+ant] derives [tɕ] from /t/ in *buc+ik* at the word level. To conclude, the word level ranking of the constraints is the

[25] The similarity between DOT and Lexical Phonology (see footnote 22) is not accidental. For discussion, see Kiparsky (1997, 2000), Rubach (1997, 2000a, 2000b), and Bermúdez-Otero (1999).

Classic OT is unable to solve the ranking paradox in (54) *vs.* (57).[27] In contrast, DOT has no difficulty dealing with these data. At the word level, the Russian response to PAL-*i* is Surface Palatalization, so IDENT-V[−back] is ranked above IDENT-C[+back], as in (54). At the post-lexical level, the response is Retraction, so the constraints are reranked to IDENT-C[+back] >> IDENT-V[−back], as in (57).

6 Conclusion

Slavic languages exhibit complex but regular patterns of palatalization, which uphold the bifurcation of the inventory into hard and soft segments, a bifurcation that is observed in numerous Slavic languages. The distinction between these classes of segments is necessary for morphology (assignment of declension paradigms), so it must exist at the underlying level. This view is supported by phonology since soft segments may occur in environments in which they cannot be derived from hard segments: before back vowels or at the end of the word.

Not all soft segments found in the surface representations come directly from underlying representations. There is solid evidence that in many instances soft segments are derived from hard segments via productive palatalization processes. These processes induce a variety of surface effects, ranging from palatalization as a secondary articulation to palatalization as a change in the place and/or manner of articulation.

Palatalization leads to much opacity in the surface representations, but this does not pose a serious analytical problem. Rather, it is a natural consequence of the interaction between independent generalizations.

OT can make excellent sense of the typological differences regarding both inventories and processes in various Slavic languages. The reason for the success is that OT uses universal constraints and claims that the permutation of these constraints generates different languages (factorial typology). The palatalization material discussed here upholds this claim and adds to it by showing that OT must admit derivational levels, making DOT rather than classic OT the better framework of analysis.

REFERENCES

Avanesov, Ruben I. 1968. *Russkoye literaturnoye proiznoshenie*. Moscow: Prosveshchenie.

Bajec, A., R. Kolarič & M. Rupel. 1956. *Slovenska slovnica*. Ljubljana: Državna Založba Slovenije.

Bermúdez-Otero, Ricardo. 1999. Constraint interaction in language change: Quantity in English and Germanic. Ph.D. dissertation, University of Manchester & University of Santiago de Compostela.

Bethin, Christina Y. 1992. *Polish syllables: The role of prosody in phonology and morphology*. Columbus: Slavica.

Bilodid, I. K. 1969. *Suchasna ukraïns'ka literaturna mova: Fonetyka*. Kiev: Naukova Dumka.

Booij, Geert & Jerzy Rubach. 1987. Postcyclic versus postlexical rules in Lexical Phonology. *Linguistic Inquiry* 18. 1–44.

[27] Classic OT auxiliary theories designed to solve opacity and ranking contradictions are not successful either. See Rubach (2004, 2005) for a discussion of McCarthy's (1999, 2003) Sympathy Theory.

122 Slavic Yers

Tobias Scheer

1 Introduction

Yers is the name given to two vowels of Common Slavic (CS) that represent Indo-European short *i* (CS *vьdova = Lat. *vidua* 'widow') and short *u* (CS *dъva = Lat. *duo* 'two'), which is also the value that they had in CS. Textbook descriptions hold that by the end of the CS period, i.e. around the ninth century, ь and ъ first became centralized and extra short, before being lost altogether in certain positions. They appear in Old Church Slavonic documents, which have been available since the tenth century, but depending on the geographic origin (and invariably in more recent sources) there is evidence that the scribes no longer had access to their original forms (etymologically inconsistent use; confusion between ь and ъ). In modern Slavic languages, yers are represented by different vowels, which typically alternate with zero. This is due to the language-specific vocalization of yers that has occurred after the dialectal differentiation of CS into individual Slavic languages, which may (e.g. Eastern Slavic, Bulgarian) or may not (most of Western Slavic, Bosno-Croato-Serbian) continue to distinguish the original front–back opposition: in Russian for example, *e* represents ь, while *o* derives from ъ (CS dьnь, sъnъ > *d'en'* 'day', *son* 'dream'); and both yers have merged in Czech (> *den*, *sen*) and Bosno-Croato-Serbian (BCS) (> *dan*, *san*).

The diachronic situation is described at length in the traditional literature (e.g. Shevelov 1964: 432ff.; Carlton 1991: 165ff.), and we will not be concerned with it in any further detail in this chapter (except on the occasion of an excursus in §5 and §6.1). Rather, we will concentrate on the consequences of the loss of yers for the phonology of modern Slavic languages, and on their relevance for phonological theory. In the literature on Slavic, yers are typically associated with particular phenomena that are considered to be specifically Slavic. As a consequence, yer-related phenomena have been confined to the phonology of Slavic languages. For example, the literature often talks about yer vowels when describing vowel–zero alternations in Slavic; nobody would talk about yer vowels when it comes to the description of vowel–zero alternations in, say, French (on which see also CHAPTER 26: SCHWA): yers are a phenomenon specific to Slavic phonology.

We will see that the practice of considering yers to be Slavic-specific items is unwarranted, as it impedes phonological insight: Slavic vowel–zero alternations may have come into being on the grounds of vowels that have only existed in Slavic, but they are controlled by the same grammatical principles that are responsible for vowel–zero alternations and other processes in other languages. That is, the Lower rule, which has acquired some notoriety outside Slavic phonology, but is typically thought of as an analysis of a specifically Slavic phenomenon, is no more Slavic than, say, palatalization. The goal of this chapter is to show that Slavic vowel–zero alternations (i.e. yers) can contribute valuable insight into phonological theory, provided that the Slavic bias is abandoned.

2 Vowel–zero alternations: The basic pattern

2.1 *Alternating* vs. *non-alternating vowels, insertion* vs. *deletion*

A general property of Slavic vowel–zero alternations is that whether a vowel alternates with zero or not cannot be predicted from its phonetic or contrastive properties. Some illustration is given under (1) below.[1]

(1) *Alternating and non-alternating vowels of the same quality*

	alternating		*non-alternating*	
	CvC	CøC-V	CvC	CvC-V
Russian	ku'sok	kusø'k-a	ra'bot	ra'bot-a
	'piece (NOM SG/GEN SG)'		'work (GEN PL/NOM SG)'	
Polish	pies	pøs-a	bies	bies-a
	'dog (NOM SG/GEN SG)'		'devil (NOM SG/GEN SG)'	
Czech	lev	løv-a	les	les-a
	'lion (NOM SG/GEN SG)'		'forest (NOM SG/GEN SG)'	
BCS	tajac	tajøc-a	pajac	pajac-a
	'silence (NOM SG/GEN SG)'		'clown (NOM SG/GEN SG)'	

Whether a vowel alternates with zero or not must thus be somehow recorded in the lexicon. That is, analyses must be able to somehow distinguish "true" (i.e. stable) from "false" (i.e. alternating) vowels of the same quality. The occurrence of yers is thus unpredictable (just as their distribution as regular vowels in CS was), and this is also true at a broader level: alternating vowels freely occur across prefixes, roots, and suffixes in Slavic languages.

Another question that has been debated at length in the literature is whether alternating vowels are underlyingly absent and inserted, or present and deleted. Insertion-based analyses have been proposed by, among others, Laskowski (1975), Czaykowska-Higgins (1988), and Piotrowski (1992). They are convincingly

[1] Note that in this chapter the symbol "ø" indicates the absence of a vowel that alternates with zero. Data are presented in orthography (or transliteration for Russian) throughout. In most cases symbols are self-explanatory. Specifics are as follows: for Russian an apostrophe after a consonant indicates its palatality (as in *d'en'* 'day'). Polish *cz* and Czech *č* are [ʧ]; in (3), they are the palatalized version of the underlying suffixal /k/. Polish *ł* is [w], *y* is [ɨ] in Polish but [i] in Czech, and the diacritic on Czech *ě* indicates the palatality of the preceding consonant. In Czech (and Slovak), vowel length is noted by an acute accent or by a circle on *u* (*ů* is a long [uu]). In Polish, *ó* is pronounced [u] (just like *u*). Finally, Polish *ę*, *ą* are nasal vowels.

(3)

	a. *open syllable*		b. *closed syllable*	
	zero	vowel	vowel	vowel
	C __ C-V	C __ C-yer Cø	C __ C-ø	C __ C-CV
Russian	d'n'-a	d'e'n'-ok	'd'en'	d'en'-ø'k-a
Czech	dom-øk-u	dom-eč-ek	dom-ek	dom-eč-øk-u
Slovak	kríd-øl-o	kríd-el-iec	kríd-el	kríd-el-øc-e
Polish	buł-øk-a	buł-ecz-ek	buł-ek	buł-ecz-øk-a

The paradigms shown are fully regular in the languages in question, and the relevant distributional regularity is thus as in (4).[6]

(4) Alternation sites are vocalized in open syllables iff the following vowel alternates with zero.

Indeed, in all cases where an alternation site is vocalized in an open syllable (Pol. *bu.łe.czek*), the vowel of the following syllable alternates with zero itself (Pol. *bu.łe.czø.ka*). In other words, the existence of a vowel in *-ecz-* is a consequence of the fact that the vowel in *-ek* alternates with zero. Alternation sites are never vocalized in open syllables when the following vowel is stable.

For historical reasons that were discussed in the introduction, vowels which alternate with zero in Slavic languages are called yers. Following this tradition, the distributional generalization that covers all facts discussed identifies as the disjunction under (5) below that is rendered in *SPE*-type notation.

(5) *The yer context*

alternation sites show

V / __	{ C.CV	} in closed syllables	buł-ecz-k-a
	C#		buł-ek
	C ь,ъ }	before yers	buł-ecz-ek
ø / __	CV	iff V ≠ ь,ъ	buł-øk-a

The challenge raised by this distribution is its disjunctivity: vocalization occurs in closed syllables and in open syllables iff the following vowel is a yer. The question thus arises in which way closed syllables and yers constitute a natural class: i.e. what they have in common. We will see below that the syllable-based generalization can be maintained if certain assumptions are made regarding underlying representations and the cyclic (or phase-based) nature of the derivation.

3 Lower

3.1 *Original linear implementation*

If there is any chance to capture the distribution of vocalized and unvocalized alternation sites in terms of a non-disjunctive statement at all, the formulation must

[6] The yer literature typically talks about vocalized and unvocalized yers: yers are lexically present and thus may or may not appear on the surface. In the former case they are said to be vocalized, while in the latter they are unvocalized. The same goes for the more neutral (and less Slavo-centristic) term "alternation site." These vocabulary items are commonly used in the chapter.

The derivations under (7) show Lower and yer deletion at work, also in case several alternating vowels occur in a row. The example used is the Polish word for bread roll *bułka* (see (3); the rule that palatalizes *k* to *cz* is not represented).

(7) *Sample derivations showing the operation of Lower*

underlying	a.	buł-ĭcz-ĭk-a	b.	buł-ɨ̆k-ɨ̆
Lower		buł-ecz-ĭk-a		buł-ek-ɨ̆
yer-deletion		buł-ecz-k-a		buł-ek
surface		bułeczk-a		bułek
relevant yer occurs in		__C yer CV		__C yer #
underlying	c.	buł-ĭcz-ĭk-ɨ̆	d.	buł-ĭk-a
Lower		buł-ecz-ek-ɨ̆		buł-ĭk-a
yer-deletion		buł-ecz-ek		buł-k-a
surface		bułecz-ek		bułk-a
relevant yer occurs in		__C yer C yer #		__C V

Note that, under (7c), Lower must apply twice and from left to right, i.e. cyclically following the morphological structure [[[[buł] ɨ̆k] ĭk] ɨ̆], in order to transform /buł-ĭk-ɨ̆k-ɨ̆/ into [buł-ecz-ek-ɨ̆]. Were [buł ĭk ɨ̆k ɨ̆] interpreted in one go, it would not be clear to which yer Lower should apply first (more on this in §6.2). The traditional assumption is therefore that Lower is applied cyclically (e.g. Lightner 1965: 111f.; Pesetsky 1979; Rubach 1984: 184ff.). Based on Anderson (1974), however, Gussmann (1980, 2007) advocates a non-cyclic version of Lower whereby "the string is first scanned for the [alternating] segments; once these are identified, the change is implemented simultaneously" (Gussmann 1980: 30). That is, all yers are vocalized in one go, according to whether or not the following vowel is a yer in the underlying form.

4 Autosegmental analyses

4.1 *Autosegmentalized Lower*

In the 1980s, when the autosegmental idea was applied to all areas of phonological theory, Lower also evolved. The autosegmentalization of Lower was adopted by Hyman (1985: 58f.) and Rubach (1986) (see also Kenstowicz and Rubach 1987; Bethin 1998: 2[illegible]5 provides an informed overview). Rather than the rule itself, the autosegmentalization concerns the lexical identity of yers: recall that their distribution is unpredictable, and that they must be distinguished from non-alternating vowels of the same quality at the underlying level. In a non-autosegmental environment, the only way to express that two vowels are different is to make them contrast in quality. Hence, a six-vowel system such as the one encountered in Polish ([i u ɨ ɛ ɔ a]) will have to be augmented by two yers, whose melodic identity must not coincide with any of the existing vowels. The traditional solution since Lightner (1965) has been to assume that yers are high vowels, but ones which are assigned a [–tense] feature, which isolates them from the other three high vowels.

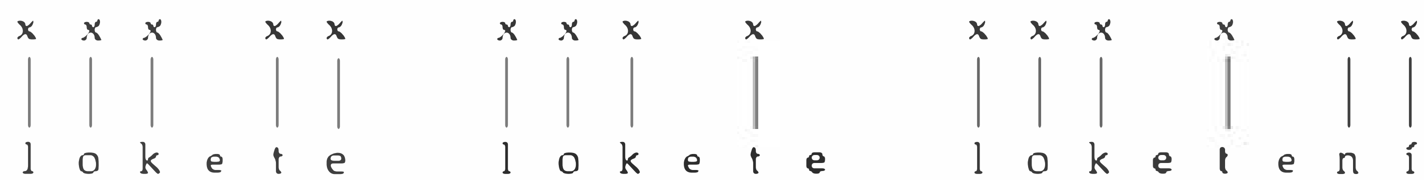
The result is a system where Polish possesses no fewer than five high vowels: /i u ɨ ĭ ɨ̆ ɛ ɔ a/.[8]

In autosegmental representations, a vowel that enjoys phonetic expression is defined as the association of a melodic unit with an x-slot, which in turn is dominated by a syllabic constituent. If there is an x-slot but no melody, nothing is heard (empty onset or nucleus); if there is a melody available but no x-slot, no phonetic trace will appear (e.g. floating consonants like in French liaison); finally, if both melody and x-slot are present but remain unassociated, nothing is heard either (see also CHAPTER 54: THE SKELETON).

Autosegmental representations thus offer an alternative way of making yers different from other vowels: their peculiar properties may be encoded structurally, rather than melodically. The alternative proposed by Rubach (1986) and Kenstowicz and Rubach (1987) therefore grants melodic, but no skeletal, properties to yers (see also Hyman 1985: 58f. along the same lines in a mora-based environment): yers are floating pieces of melody that do not possess any skeletal anchor in the lexicon, while stable vowels (which may be melodically identical) are lexically associated to an x-slot. The corresponding underlying representations are shown under (8) for the three relevant distributional situations, which are illustrated by Czech data from (2).

(8) *Yers are floating pieces of melody* (Rubach 1986)
Czech 'elbow'

a. loke̲t-e (GEN SG) b. loket (NOM SG) c. loket-ní (ADJ)



This option offers several advantages.[9] First, there is no longer any need to make yers high vowels, a solution whose only motivation was their historical CS properties. Also, extra features no longer need to be invoked: the choice of [–tense]

[8] The situation is actually more complicated than this; see Gussmann (1980: 63ff.) and Rubach (1984: 27ff., 139ff.).

[9] The autosegmentalization of Lower introduces representations as in (8), where syllabification is contrastive: two lexical items may be distinct only by the fact that a vowel either floats or is lexically attached. Two reviewers point out that contrastive syllabification may be viewed as a suspicious device from a general perspective. Note that since Hyman (1985) and Rubach (1986) it has been part-and-parcel of all (autosegmental) accounts of Slavic vowel–zero alternations across different theories (including OT). Also, apprehension regarding contrastive syllabification appears to be a bias that is introduced by theoretical preferences, rather than by fact: the general perspective inherited from the 1970s is that lexical entries are a linear sequence of pieces of melody, on which a syllabification algorithm superimposes syllable structure during phonological computation. In this perspective, therefore, syllable structure is absent from the lexicon. From this point of view, contrastive syllabification is indeed suspicious. This conception, however, is not without alternatives: there are theories like Government Phonology where lexical entries are fully syllabified, and the autosegmental analysis of Slavic yers is a step in this direction (although here only x-slots, not full syllable structure, are present in the lexicon). Also, the common practice of marking extrasyllabic consonants as <C> in lexical representations is nothing else than prespecifying the behavior of the item at hand during syllabification. Finally, the lexical association of melody with moras has been common practice in mora-based approaches to syllable structure since Hayes (1989). This option is represented by Yearley (1995) in the analysis of yers.

was entirely arbitrary. Any diacritic underlying representation for the two yers, such as /@/ and /%/, or any position in the vocalic triangle defined by whatever feature, would have done the job in the linear environment, provided that the two items are different from all other underlying vowels.

The alternative in (8) thus does away with absolute neutralization (i.e. a vowel that never appears on the surface in its underlying form) and its associated arbitrariness. Rather, alternating (yers) and non-alternating vowels are now identical as far as their melodic representation is concerned: the contrast is expressed in terms of association (compare the two *e*'s under (8a). There is thus a trade-off between an inventory that is expressed only by melodic contrast and an inventory where contrast is achieved both by melodic features and autosegmental structure. Of course, the number of objects that need to be contrasted remains the same – but the advantage of the autosegmental option is that no arbitrary contrast-achieving machinery ([–tense] vowels) needs to be used: the new way of expressing contrast only uses the tools that are provided by the autosegmental environment anyway.

Another advantage is that the rule of yer deletion can be dispensed with: the phonetic absence of unassociated melodic material is automatic in the autosegmental environment.

Finally, the autosegmentalization of Lower emancipates synchronic vowel–zero alternations from their diachronic affiliation, which also makes them a phenomenon that is not necessarily specific to Slavic languages (more on this in §8.2). That is, vowels which alternate with zero are no longer subject to any melodic restriction: in the linear system, two distinct vowels at most (i.e. the Eastern Slavic situation) could alternate with zero. Languages like Slovak, where more than two vowels alternate with zero, therefore begged the question. Rubach (1993: 139ff.) reports relevant evidence: *o*–zero *bahor* – *baho̲r-a* 'belly (NOM SG/GEN SG)', *e*–zero *šev* – *šo̲v-u* 'seam (NOM SG/GEN SG)', *a*–zero *jedo̲l-o* – *jedál* 'food (NOM SG/GEN PL)'.

In autosegmental terms, the relevant representations are straightforward: whatever vowel alternates with zero and whatever the number of different alternating vowels, their underlying representation is simply their floating melodic identity without a skeletal slot. Any alternating vowel will thereby be different from its melodically identical stable peer because it is associated to a skeletal slot. While every additional vowel that alternates with zero augments the inventory of underlying melodies in the linear system, the number of melodic identities is stable no matter whether a language possesses vowels that alternate with zero or not in the autosegmental environment.[10]

The autosegmentalized version of Lower is shown in (9).

(9) *Autosegmentalized Lower* (Rubach 1986; Kenstowicz and Rubach 1987)

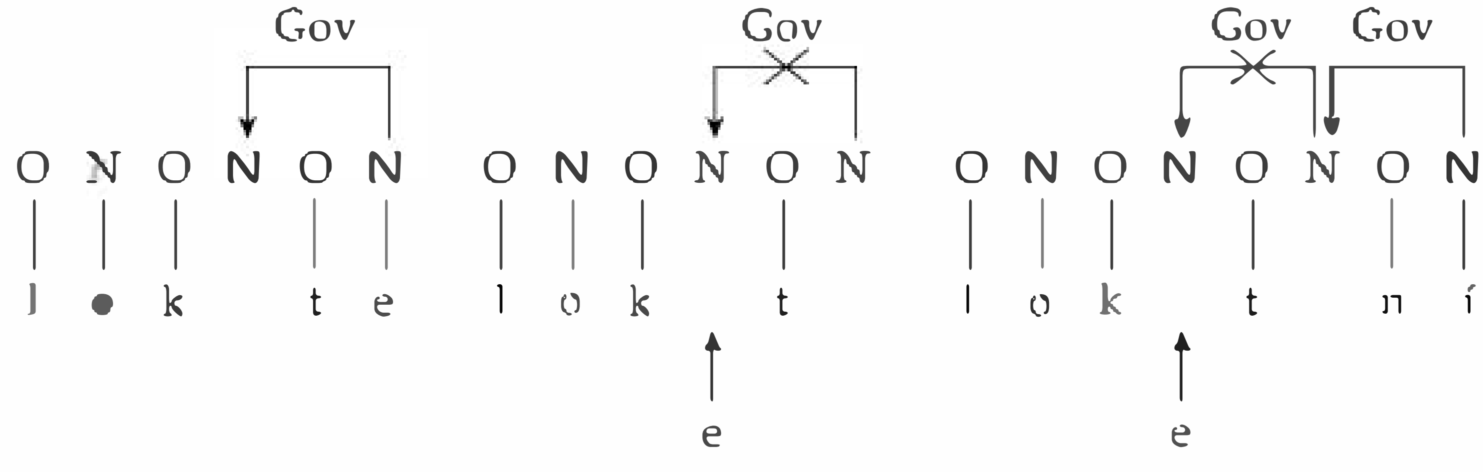
$$\text{ⓥ} \rightarrow \begin{array}{c} \text{x} \\ | \\ \text{V} \end{array} \;/\; \_\_\, C_0 \; \text{ⓥ}$$

[10] The number of alternating, i.e. lexically unassociated, vowels that a language possesses is thus a lexical and unpredictable property of the vocalic inventory: the reason why *e* and *o*, but not other vowels, are lexically floating in Russian, or why only *e* enjoys this status in Czech, is the same as the reason why Russian does not have, say, front rounded vowels.

In Government Phonology, Gussmann and Kaye (1993) apply the general distribution of empty nuclei that has no specific bond with Slavic to Polish vowel–zero alternations. In this perspective, alternating vowels (i.e. yers) have the opposite characteristics from those proposed by Rubach (1986): rather than melodies that lack lexical association (to an x-slot), they are represented as empty nuclei, i.e. a constituent that lacks melody. This option is shown in (12), where the Cezch data from (8) are used.

(12) *Yers are empty nuclei* (Gussmann and Kaye 1993)
Czech 'elbow'

a. lok$\underline{ø}$t-e (GEN SG) b. loket (NOM SG) c. loket-ní (ADJ)



In this perspective, *e* is epenthesized into empty nuclei that fail to be governed. Government is defined as a lateral relation whose head (origin) must be phonetically expressed: it relates the final and preceding nuclei in (12a) because the former is filled with the genitive case marker, but breaks down in (12b) when the final nucleus is empty (in classical terms: when the internal yer is followed by a final yer). When applying, government has thus the effect of silencing its target: the empty nucleus in (12a) remains unpronounced, but must appear on the surface when it fails to be governed as in (12b). Since government is a regressive lateral relation, strings that are subject to phonological computation are parsed from right to left. Thus the final filled nucleus first governs its left-hand neighbor in (12c), which therefore remains empty. Being phonetically unexpressed, this nucleus is then unable to act as a governor, a fact that causes the preceding empty nucleus to escape government and hence to vocalize.

Like the traditional analysis, the government-based version of Lower needs to recur to cyclic derivation in order to account for sequences of alternating vowels. In Polish *buł-ecz-ek* 'bread roll (DOUBLE DIM, NOM SG)' (see (7c)), the application of government to the entire string /buł$\underline{ø}_3$k$\underline{ø}_2$k$\underline{ø}_1$/ in one go would produce **buł$\underline{ø}$czek*: $\underline{ø}_1$ would be unable to govern $\underline{ø}_2$, which would therefore vocalize and govern $\underline{ø}_3$.

In order to derive the vocalization of all alternating vowels in a row, Gussmann and Kaye (1993) therefore apply government cyclically: *buł-ecz-ek* is divided into three cycles that are computed each in its own right: given [[[buł$\underline{ø}_3$] k$\underline{ø}_2$] k$\underline{ø}_1$],[15] nothing happens on the first cycle, [buł$\underline{ø}_3$], since there is only one domain-final empty nucleus. The second cycle involves [buł$\underline{ø}_3$k$\underline{ø}_2$], with two successive

[15] This underlying form is simplified for the sake of exposition: the suffix *-ek* is identified as /-$\underline{ø}$k$\underline{ø}$/ in Gussmann and Kaye (1993), but due to an operation called reduction loses its leftmost empty nucleus during the derivation.

empty nuclei, a situation that triggers the vocalization of $\underline{\emptyset}_3$ through the regular (non-)application of government ($\underline{\emptyset}_2$ is unable to govern $\underline{\emptyset}_3$, which therefore vocalizes). The input to the third cycle is thus [bułek$\underline{\emptyset}_2$k$\underline{\emptyset}_1$], and the presence of two successive empty nuclei again causes the leftmost to vocalize, producing the surface result *bułeczek*.

Gussmann and Kaye's (1993) analysis is an interesting blend of deletion- and insertion-based strategies: on the one hand the locus of alternating vowels is lexically specified by the presence of empty nuclei; these are pronounced by default but may be silenced when they are struck by government. In this sense we are talking about deletion. On the other hand, empty nuclei are (melodically) empty and therefore need to be filled in with melody that is lexically absent. In this sense the analysis is insertion-based.

Recall from §2.1 that insertion-based analyses were refuted mainly on the grounds of two arguments: (i) the locus of insertion cannot be predicted and (ii) the quality of the vowel that alternates with zero may be an idiosyncratic property of morphemes (like in Eastern Slavic, for example). While Gussmann and Kaye's (1993) system escapes the first criticism (empty nuclei determine the locus of alternation sites lexically), the second argument hits the target: in languages like Russian where more than one vowel alternates with zero, it cannot be predicted which vowel will be inserted into which morpheme.

The quality of alternating vowels must thus be recorded in the lexicon. Scheer (2004: §81f.) has therefore proposed to unify Rubach's original take, where yers are floating melodies that are recorded in the lexical make-up of morphemes with the government-based strand. That is, alternating vowels (yers) are nuclei that possess a melody, which however is not associated lexically (see also Gussmann 2007). By contrast, the melody of non-alternating vowels is lexically associated. The three options discussed are contrasted in (13), using the Czech word *pes* – *p*$\underline{\emptyset}$*s-a* 'dog (NOM SG/GEN SG)'.

(13) *Underlying representations of vowels that alternate with zero (yers)*

a. Rubach (1986) b. Gussmann and Kaye (1993) c. Scheer (2004,2005)

On the assumption of (13c), then, government acts as an association inhibitor: floating melodies associate by default except when their nucleus is governed. Or, in other words, melodies can only associate to ungoverned nuclei.[16] Another interesting property of (13c) is that it distinguishes between two kinds of "empty" nuclei: one that is really empty, and one that possesses an unassociated

[16] This analysis is along the lines of the behavior that is classically assumed for floating pieces of melody: for example, French liaison consonants (e.g. the final -*t* of *petit*, cf. *pet*[it] *enfant* 'small child' *vs. pet*[i] *café* 'small coffee') are traditionally taken to be lexically floating; they associate whenever an accessible, i.e. empty onset is available (e.g. Encrevé 1988). Hence association works in the absence of any explicit lexical relationship between the floater and the receiving constituent. The same is true for (13c): the floater associates whenever it can, and to the only nucleus that is available. The accessibility of this nucleus is defined by government (governed = inaccessible, ungoverned = accessible).

from all versions of the word in all possible donor languages (see also CHAPTER 95: LOANWORD PHONOLOGY).

A decision to break up a cluster by an epenthetic vowel or not thus needs to be made upon the lexicalization of new vocabulary items, and this produces random distribution of alternating vowels. Also, children when acquiring their native tongue have no way of knowing whether the cluster of something that they hear as *swetr-a* 'jumper (GEN SG)' in Polish or *metr-o* 'metro' in Czech will or will not break up until they have a chance to hear the word without the final vowel. Relevant mislexicalizations are indeed typical "mistakes" that are reported from children.

The arbitrary lexical distribution of alternating vowels in the modern language is also confirmed by the reverse evolution: there are numerous cases of vowels that were CS yers and thus alternated, but today are stable. Examples include *blech-a* < CS blъch-a 'flea (nom sg)', *bez* – *bez-u* 'elder (bot.) (NOM SG/GEN SG)' < CS bъzъ (e.g. Trávníček 1935: 48).

Let us now turn to alternating vowels without etymological basis that occur in native vocabulary. (14) provides some illustrations from Czech.

(14) *Non-etymological yers in native vocabulary*

a. *epenthesis in Old Czech*

CS	*Old Cz*	
ogn-ь	oheň	'fire (NOM SG)'
od-	od(e)-	'from'
orz-	roz(e)-	'separating, inchoative'
bez-	bez(e)-	'without'

b. *epenthesis in Modern Czech*

CS	*Old Cz*	*ModCz*	
vydr-ъ	vydr	vyder	'otter (GEN PL)'
sestr-ъ	sestr	sester	'sister (GEN PL)'
stьbl-ъ	stébl	stébel	'blade (GEN PL)'
kridl-ъ	křídl	křídel	'wing (GEN PL)'

(14) confirms that epenthesis was also active in native vocabulary, and throughout different stages of the evolution of the language: (14a) illustrates epenthetic alternating vowels that appeared between CS and Old Czech,[18] while (14b) shows epenthesis in Old Czech final clusters that occurred in Modern Czech.[19]

In sum, it may be said that while the initial spark for vowel–zero alternations in Slavic was the loss of CS yers, modern alternating vowels are entirely independent: it is true neither that they all go back to a CS yer, nor that all CS yers have produced modern alternating vowels.

[18] The three latter items under (14a) are prefixes/prepositions (whose behavior in Modern Czech is further discussed in §7.1). They show an alternating vowel in Old Czech despite the fact that there was no yer in CS: e.g. OCz *ote dne* 'from the day', *beze všeho* 'without all', *roze-hnal* 'dispel, scatter' (Trávníček 1935: 50). Isačenko (1979) describes the state of affairs in Russian prefixes.

[19] Unlike for the recent loans, epenthesis in (14a) and (14b) is predictable. Its conditions are discussed in Scheer *et al.* (2009).

[dom-Ek-Ek] without internal cyclic structure (in syntactic terms, it represents one single phase). The suffix *-ek* then became cyclic in diachronic evolution, which means that a cycle is created upon concatenation or, in syntactic terms, interpretation is triggered upon concatenation and the string has three phases: [[[dom] Ek] Ek]. By contrast, case markers such as GEN SG *-u* were non-cyclic in Old Czech and still are today: OCz = MdCz [dom-Ek-u] *domček* (Ziková 2008).

In other words, whether interpretation is triggered or not depends on a lexical property of affixes. This scenario contrasts with current syntactic phase theory, where a given class of arboreal nodes that is defined independently of lexical material (phase heads: CP, vP, and perhaps DP in Chomsky's original conception) determines when interpretation occurs. Affix classes in phonology (of which the Lower–Havlík contrast is an instance) thus call for what could be named piece-driven phases where lexical material is specified for triggering or not triggering interpretation, while current syntactic theory employs node-driven phases (see Scheer 2008 for further discussion).

7 Prepositions and prefixes

7.1 *The general picture*

Yers also occur after the final consonant of consonant-final prepositions and prefixes: e.g. Czech *v/ve* 'in', *nad/nade* 'over, above'. Since the vocalization of yers is decided by the right-hand context, the morpheme or word boundary that intervenes between the yer and the conditioning context plays a role: its greater or lesser transparency will impact on yer vocalization. This is the basic reason for the typically irregular vocalization patterns that are found in prefixes and prepositions. But there are also other factors that further complicate the picture: sensitivity to identical (or similar) preposition-/prefix-final and stem-initial consonants, sensitivity to stem-initial clusters and, in the case of prefixes, a good deal of analogical activity. Finally, lexicalized preposition–noun sequences also appear to occur.

As may be expected in such a case, there is a fair amount of variation among Slavic languages regarding the treatment of alternation sites in prepositions and prefixes. Relevant literature includes Zubritskaya (1995), Rubach (2000), Matushansky (2002), Gribanova (2008) on Russian, Rubach (1993) on Slovak, and Scheer (1996), Ziková (2008) on Czech (literature on Polish is mentioned below).

Let us first look at the stock of Slavic consonant-final prepositions and prefixes, which is more or less stable across Slavic languages. (18) below shows the Czech inventory (only basic, i.e. morphologically non-complex, items are displayed). It may be seen that prepositions and prefixes are by and large the same objects and have comparable meaning.

(18) *Czech prepositions and prefixes*

	prepositions	*prefixes*	
a.	v(e)	v(e)-	'in'
	s(e)	s(e)-	'with'
	z(e)	z(e)-	'from inside'
	k(e)	—	'towards'

b. *before unexpressed alternating vowels*

od-wszyć	wesz, wszy	'to de-louse, louse (NOM SG/GEN SG)'
od-pchlić	pchla, pcheł	'to de-flea, flea (NOM SG/GEN PL)'
bez-cłowy	cło, ceł	'duty-free, duty (NOM SG/GEN PL)'
nad-dniówka	dzień, dnia	'extra day's work, day (NOM SG/GEN SG)'
w-śnić się	sen, snu	'to start dreaming, dream (NOM SG/GEN SG)'
roz-łzawić	łza, łez	'to draw tears, tears (NOM SG/GEN PL)'

In these cases, the prefix boundary impacts on the vocalization of prefixes, which escape the influence of the root: the prefix does not "see" the root vowel.[23] Various implementations of this insight can be found in, among others, Rubach (1984: 18ff.), Szpyra (1989: 215ff., 1992b, 1995: 132f.), Cyran and Gussmann (1999), and Rowicka (1999a: 267ff.). Rubach (1984: 186ff.) develops a solution along the lines of Lexical Phonology: he captures the phonological autonomy of prefixes by feeding them into the derivation on the last cycle, which makes them immune to the action of Lower.

Whatever analysis is set up, though, it will face conflicting evidence since the same prefixes sometimes do – but at other times do not – vocalize in identical phonological environments.

8 The relevance of yers for phenomena other than vowel–zero alternations

8.1 *Within Slavic*

Yers usually draw attention because of their own behavior: we have seen how their alternation with zero is analyzed. The central empirical fact that led to the Lower rule and the view that yers entertain a lateral relation with the following vowel is the yer context (5): a yer appears on the surface in closed syllables and in open syllables iff the following vowel alternates with zero. Yers thus react to other yers or, looked at from the other end, a stable vowel triggers the phonetic absence of a preceding yer, while a yer triggers its presence. As conditioners, yers thus behave as if they were not there, although they are phonetically present.

Interestingly, this effect is not only observable on yers, but also on other vowels: the yer context also controls alternations beyond vowel–zero alternations, a fact that typically goes unnoticed in the literature. Scheer (2004: §428) has collected the following examples in Western Slavic.[24]

[23] Laskowski (1975: 34ff.) suggests that the strength or weakness of the prefixal boundary depends on the type of derivation: prefixes behave regularly in verbal forms, but produce the irregular pattern in nouns. This falls foul of fact in a number of cases in (20). Nykiel-Herbert (1985) contends that the vocalization of prefixes is a function of the number of morphological brackets that separate the prefix and the stem. This proposal also faces conflicting evidence; cf. Szpyra (1989: 211f.).

[24] Glosses: (a) 'frog' (NOM SG/DIM, GEN PL/GEN PL/DIM, NOM SG); 'name' (NOM SG/ADJ, GEN PL); (b) 'knife' (GEN SG)/'scissors' (i.e. DIM of 'knife'), GEN PL/'knife' (NOM SG)/'scissors' (NOM PL); (c) 'cow' (NOM SG/DIM, GEN PL)/GEN PL/DIM, NOM SG); (d) 'tooth' (GEN SG/DIM, NOM SG/NOM SG/DIM, GEN SG).

(21) *Alternations conditioned by the yer context "in closed syllables and in open syllables if the following vowel is a yer"*

		open syllable		closed syllable	
		C __ C-V	C __ C-yer	C __ C#	C __ C-CV
a.	Czech VV-V	žáb-a	žab-ek	žab	žab-øk-a
		jmén-o		jmen	jmen-ný
b.	Czech o-ů	nož-e	nůž-ek	nůž	nůž-øk-y
c.	Polish o-ó	krov-a	krów-ek	krów	krów-øk-a
d.	Polish ę-ą	zęb-a	ząb-ek	ząb	ząb-øk-a

Space restrictions hardly allow going into any more detail. For the sake of illustration, let us however consider the distribution of long and short vowels in Czech. These are restricted to a well-defined paradigm, feminine a-stems and neuter o-stems (a typical situation for Czech vowel length). At first glance the alternation looks like an instance of regular closed syllable shortening: short vowels occur in closed syllables (*žab* 'frog (GEN PL)', *žab-ka* 'frog (DIM, NOM SG)'), while long vowels appear in open syllables (*žáb-a* 'frog (NOM SG)'). Just as with vowel–zero alternations, however, this syllabic generalization is refuted by cases such as *žab-ek* 'frog (DIM, GEN PL)', where a short, rather than the expected long vowel appears in an open syllable. This only happens, though, if the following vowel alternates with zero, i.e. is a yer (its alternating character is established by *žab-ka*). In other words, the disjunctive yer context can be reduced to a simple non-disjunctive statement as before: short vowels (just as zeros) occur before yers, while long vowels are found before non-yers.

Before drawing conclusions from the fact that a number of other alternations are controlled by the same yer-based contextual conditions as vowel–zero alternations, a word is in order regarding the alternations in (21), whose non-productivity is notorious. There are many lexical items that do not undergo the alternations shown (which however are regular in case an item reacts). Diachronically, all four alternations amount to the same original Western Slavic process that manipulated vowel length.[25] The fact that the alternations are either not synchronically active (in Polish) or restricted to specific paradigms (in Czech) does not mean that they are less indicative, or do not witness a phonological process that once was synchronically active.

Therefore, if the yer context (5) is responsible for vowel–zero alternations as much as for other alternations, the Lower rule turns out to be but a specific subregularity of a much broader process whereby yers play the central role. That is, regarding the alternations in (21) as much as regarding vowel–zero alternations, the striking property of yers is that they behave as if they were not there even when they are phonetically expressed.

8.2 *Beyond Slavic*

We have already seen that the analysis of yers progressively turned away from a scenario where they are a specific property of Slavic languages. While linear

[25] Vowel length was lost in Modern Polish: alternations in this language are only witnessed by cases where original long and short versions of a vowel at some point also diverged in vowel quality. The diachronic and philological detail of the processes in (21) is discussed in Scheer (2004: §426).

approaches represented them as two idiosyncratic melodic items in the underlying vocalic inventory, they became regular and non-Slavic-specific phonological objects in the autosegmental era: only Slavic languages have [−tense] yers, but all languages can have floating pieces of melody. Government-based analyses have gone one step further: the lateral relation embodied by Lower identifies as government, and word-final consonants are followed by an empty nucleus, rather than by a yer with morphological value. Finally, it was shown in §5 that vowels which alternate with zero in modern Slavic languages may be completely independent from the Common Slavic vowels that are known as yers.

The insight that the phenomena at hand are not specifically Slavic, but phonological in nature, is also supported by the fact that the yer context in (5), which is rather specific, is found to control alternations beyond Slavic. Scheer (2004: §426, 2006) reports a number of relevant cases in point, among which two well-known alternations in French that concern schwa and [ɛ] on the one hand, and the ATR-ness of mid vowels on the other (e.g. Tranel 1987 for an overview). In French, the only vowel that alternates with zero is schwa [ə]. It has already been mentioned that, unlike in Slavic, the alternation is optional.

The schwa–[ɛ] alternation may be illustrated by the word *appeler* 'to call': [ɛ] appears in closed syllables (*j'appelle* [apɛl] 'I call', *il appellera* [apɛlʁa] 'he will call') and before a vowel that alternates with zero (*il appellera* [apɛləʁa] 'he will call'), while schwa is found in open syllables (*appeler* [apəle] 'to call'). Regarding the other alternation, French possesses six mid vowels that subdivide into two sets: [+ATR] [e o ø] and [−ATR] [ɛ ɔ œ]. In a number of Southern varieties, ATR-ness is distributed according to the yer context: [+ATR] versions occur in open syllables if the following vowel does not alternate with zero (*fêter* [fete] 'to party'), while mid vowels are [−ATR] in closed syllables (*je fête* [fɛt] 'I party', *perdu* [pɛʁdy] 'lost'), and in open syllables if the following vowel alternates with zero (*céleri* [sɛləʁi]/[sɛlʁi] 'celery').

The yer context also conditions consonants. The well-known alternation of the German velar nasal is a case in point (Scheer 2004: §480; relevant literature includes Dressler 1981 and Hall 1992: 199ff.). Like the English velar nasal, the underlying /Ng/ reduces to [ŋ] in closed syllables (*Ding* [dɪŋ] 'thing', *Angst* [ʔaŋst] 'fear'), but appears as [ŋg] before full vowels (*Ingo* [ʔɪŋgoo] (name), *evangelisch* [ʔɛfaŋgeelɪʃ] 'protestant'). Parting company with English (cf. Eng *finger* [fɪŋgə]), though, German also produces the reduced form before schwa (*Inge* [ʔɪŋə] (name), *Bengel* [bɛŋəl] 'rascal'). The same pattern, i.e. where consonants behave alike in coda position and before schwa, also occurs in Dutch (Kager and Zonneveld 1986).

Of course there is no point in trying to attribute these alternations to yer vowels or other Slavic-specific items. What we are facing is a truly phonological pattern that occurs in Slavic as much as in other languages, and whose key feature is the behavior of vowels that alternate with zero. Unlike other vowels, they behave as if they were not there even when they are phonetically expressed. Note that the quality of these vowels is entirely irrelevant: they may be peripheral, as in Slavic, or central ("true" schwas), as in French and German.

Analyses of so-called Slavic yers will therefore have to make sure that their instruments are not bound to Slavic, but can express more general phonological processes. Yers need to be extracted from their narrow Slavic context, where they have lived in a waterproof environment in much of the structuralist and the generative tradition. They can offer rich insight into phonological theory if they are placed in a broader context: "yers" condition processes that are different from vowel–zero alternations, and they are active beyond Slavic.

REFERENCES

Anderson, Stephen R. 1974. *The organization of phonology.* New York: Academic Press.

Anderson, Stephen R. 1982. The analysis of French shwa: Or, how to get something for nothing. *Language* 58. 534–573.

Bethin, Christina Y. 1992. *Polish syllables: The role of prosody in phonology and morphology.* Columbus: Slavica.

Bethin, Christina Y. 1998. *Slavic prosody: Language change and phonological theory.* Cambridge: Cambridge University Press.

Burzio, Luigi. 1994. *Principles of English stress.* Cambridge: Cambridge University Press.

Carlton, Terence. 1991. *Introduction to the phonological history of the Slavic languages.* Columbus: Slavica.

Carr, Philip. 1993. *Phonology.* London: Macmillan.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.) *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.) *Ken Hale: A life in language*, 1–52. Cambridge, MA: MIT Press.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable.* Cambridge, MA: MIT Press.

Cyran, Eugeniusz. 2003. *Complexity scales and licensing strength in phonology.* Lublin: Wydawnictwo Katolickiego Uniwersytetu Lubelskiego.

Cyran, Eugeniusz & Edmund Gussmann. 1999. Consonant clusters and governing relations: Polish initial consonant sequences. In Harry van der Hulst & Nancy Ritter (eds.) *The syllable: Views and facts*, 219–247. Berlin & New York: Mouton de Gruyter.

Czaykowska-Higgins, Ewa. 1988. *Investigations into Polish morphology and phonology.* Cambridge, MA: MIT Press.

Dell, François. 1995. Consonant clusters and phonological syllables in French. *Lingua* 95. 5–26.

Dressler, Wolfgang U. 1981. External evidence for an abstract analysis of the German velar nasal. In Didier Goyvaerts (ed.) *Phonology in the 1980's*, 445–467. Ghent: Story-Scientia.

Encrevé, Pierre. 1988. *La liaison avec et sans enchaînement: Phonologie tridimensionnelle et usages du français.* Paris: Seuil.

Farina, Donna. 1991. Palatalization and jers in modern Russian phonology: An underspecification approach. Ph.D. dissertation, University of Illinois.

Fisiak, Jacek & Stanislaw Puppel (eds.) 1992. *Phonological investigations.* Amsterdam: John Benjamins.

Gebauer, Jan. 1894–8. *Historická mluvnice jazyka českého.* 3 vols. Prague: Unie.

Grammont, Maurice. 1914. *Traité pratique de prononciation française.* Paris: Delgrave.

Gribanova, Vera. 2008. Russian prefixes, prepositions and palatalization in Stratal OT. *Proceedings of the West Coast Conference of Formal Linguistics* 26. 217–225.

Gussmann, Edmund. 1980. *Studies in abstract phonology.* Cambridge, MA: MIT Press.

Gussmann, Edmund. 1992. Back to front: Non-linear palatalization and vowels in Polish. In Fisiak & Puppel (1992), 5–66.

Gussmann, Edmund. 2004. Polish front vowels or Baudouin de Courtenay redivivus. *Studies in Polish Linguistics* 1. 103–130.

Gussmann, Edmund. 2007. *The phonology of Polish.* Oxford: Oxford University Press.

Gussmann, Edmund & Jonathan Kaye. 1993. Polish notes from a Dubrovnik Café: I. The yers. *SOAS Working Papers in Linguistics and Phonetics* 3. 427–462.

Hall, T. A. 1992. *Syllable structure and syllable-related processes in German.* Tübingen: Niemeyer.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress.* Cambridge, MA: MIT Press.

Havlík, Antonín. 1889. K otázce jerové v staré češtině. *Listy Filologické* 16. 45–51, 106–116, 248–258, 342–353, 436–445.

Hayes, Bruce. 1989. The prosodic hierarchy in meter. In Paul Kiparsky & Gilbert Youmans (eds.) *Rhythm and meter*, 201–260. San Diego: Academic Press.

Hristova, Vanya. 1994. Nominal vowel/zero alternations in Bulgarian and Russian. Ph.D. dissertation, University of Delaware.

Hyman, Larry M. 1985. *A theory of phonological weight*. Dordrecht: Foris.

Isačenko, Alexander V. 1970. East Slavic Morphophonemics and the treatment of the jers in Russian: A revision of Havlík's Law. *International Journal of Slavic Linguistics and Poetics* 13. 73–124.

Isačenko, Alexander V. 1979. Secondary vocalization of the jers. *Folia Slavica* 3. 154–160.

Kager, René & Wim Zonneveld. 1986. Schwa, syllables, and extrametricality in Dutch. *The Linguistic Review* 5. 197–221.

Kaye, Jonathan. 1990a. "Coda" licensing. *Phonology* 7. 301–330.

Kaye, Jonathan. 1990b. Government in phonology: The case of Moroccan Arabic. *The Linguistic Review* 6. 131–159.

Kaye, Jonathan. 1992. Do you believe in magic? The story of s+C sequences. *SOAS Working Papers in Linguistics and Phonetics* 2. 293–313. Reprinted 1996 in Henryk Kardela & Bogdan Szymanek (eds.) *A Festschrift for Edmund Gussmann*, 155–176. Lublin: University Press of the Catholic University of Lublin.

Kaye, Jonathan. 1995. Derivations and interfaces. In Jacques Durand & Francis Katamba (eds.) *Frontiers of phonology: Atoms, structures, derivations*, 289–332. London & New York: Longman.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology* 7. 193–231.

Kenstowicz, Michael & Jerzy Rubach. 1987. The phonology of syllabic nuclei in Slovak. *Language* 63. 463–497.

Kiparsky, Paul. 1991. Catalexis. Unpublished ms., Stanford University.

Komárek, Miroslav. 1969. *Historická mluvnice česká*, I: *Hláskosloví*. Prague: SPN.

Lamprecht, Arnošt. 1987. *Praslovanština*. Brno: Univerzita Purkyně.

Laskowski, Roman. 1975. *Studia nad morfonologią współczesnego języka polskiego*. Wrocław: Wydawnictwo Polskiej Akademii Nauk.

Liewehr, Ferdinand. 1933. *Einführung in die historische Grammatik der tschechischen Sprache*, I: *Lautlehre*, *Erste Lieferung*. Brünn: Rohrer.

Lightner, Theodore M. 1965. Segmental phonology of Contemporary Standard Russian. Ph.D. dissertation, MIT.

Lowenstamm, Jean. 1996. CV as the only syllable type. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, 419–441. Salford: ESRI.

Matushansky, Ora. 2002. On formal identity of Russian prefixes and prepositions. *MIT Working Papers in Linguistics* 42. 217–253.

Melvold, Janis. 1989. Structure and stress in the phonology of Russian. Ph.D. dissertation, MIT.

Noske, Roland. 1993. *A theory of syllabification and segmental alternation: With studies on the phonology of French, German, Tonkawa and Yawelmani*. Tübingen: Niemeyer.

Nykiel-Herbert, Barbara. 1985. The vowel–zero alternation in Polish prefixes. In Edmund Gussmann (ed.) *Phono-morphology: Studies in the interaction of phonology and morphology*, 113–130. Lublin: Wydawnictwo Katolicki Universytet Lubelski.

Oostendorp, Marc van. 2005. The first person singular in Dutch dialects. *Papers from the Annual Meeting of the North East Linguistic Society* 35. 1–12.

Szpyra, Jolanta. 1992b. The phonology of Polish prefixation. In Fisiak & Puppel (1992), 185–218.

Szpyra, Jolanta. 1995. *Three tiers in Polish and English phonology*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Tranel, Bernard. 1987. Floating schwas and closed syllable adjustment in French. In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.) *Phonologica 1984*, 311–317. London: Cambridge University Press.

Trávníček, František. 1935. *Historická mluvnice československá*. Prague: Melantrich.

Trávníček, František. 1948–9. *Mluvnice spisovné češtiny*. 2 vols. Prague: Melantrich.

Wiese, Richard. 1996. *The phonology of German*. Oxford: Clarendon Press.

Yearley, Jennifer. 1995. Jer vowels in Russian. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 533–571. Amherst: GSLA.

Ziková, Markéta. 2008. Alternace e-nula v současné češtině: Autosegmentální laterální analýza. Ph.D. dissertation, Masarykova Univerzita, Brně.

Ziková, Markéta. 2009. Vowel–zero alternations in Czech double diminutives: From the Havlík to the Lower pattern. In Mojmír Dočekal & Markéta Ziková (eds.) *Czech in formal grammar*, 237–250. Munich: Lincom Europa.

Zubritskaya, Ekatarina. 1995. The categorial and variable phonology of Russian. Ph.D. dissertation, University of Pennsylvania.

9 Handshape in Sign Language Phonology

Diane Brentari

1 Introduction

This chapter addresses the particular importance of the handshape parameter in sign language phonology, using evidence from a variety of sources: psycholinguistic and cross-linguistic, as well as system-internal evidence from phonological operations. The chapter is organized as follows. In §1, by way of introduction to this topic, some of the ways that handshape behaves differently in gesture and sign languages will be described. In §2, portions of basic handshape structure will be presented, the terms that will be used throughout the chapter will be defined, and some descriptive generalizations concerning handshape will be outlined. Further elaboration of the features of handshape is discussed in §3. A description on how handshape interacts with the other parameters of signs will be discussed in §4. In §5, several examples of the role of handshape at the grammatical interfaces will be described, and §6 will consist of some concluding remarks on the topic of handshape.

1.1 Handshape in gesture and sign language

Handshape includes the form that the hand(s) assume(s) during the articulation of a sign or gesture; however, handshape has been shown to behave differently in gesture and in sign languages in several important ways.

One notable difference concerning handshape in gesture and sign is the greater range of forms used by signers in comparison with gesturers. This has been demonstrated in a number of psycholinguistic studies using tasks in which signing and non-signing (gesturing) participants were asked to describe events using only signs or gestures (Singleton *et al.* 1993; Goldin-Meadow *et al.* 1996; Emmorey and Herzig 2003; Schembri *et al.* 2005). I will describe just one of these studies, that of Schembri *et al.* (2005). In this study the gesturers did not use their voices. A test that consisted of animated video stimuli of objects moving in space was used to elicit productions from three groups (Supalla *et al.* 1995): non-signing hearing gesturers from Australia, signers of Australian Sign Language, and signers of Taiwanese Sign Language. Properties of handshape, place of articulation (where the sign is articulated), and movement (displacement of the hands in space) were

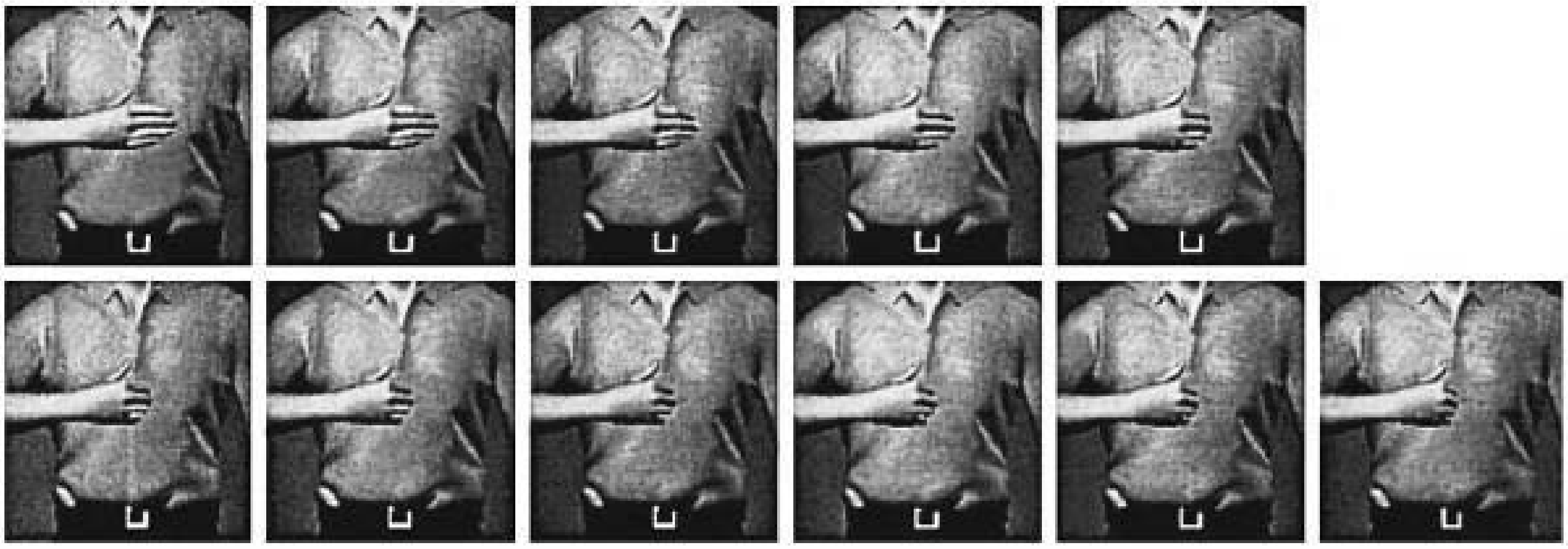


Figure 9.1 The eleven equal intervals between the two contrastive handshapes that were used in the categorical perception experiment presented in Emmorey *et al.* (2003): in the sign PLEASE (top left) and in the sign SORRY (bottom right)

counted for each subject and each group. I will describe only the results comparing the Australian gesturers and signers. The largest difference occurred in the handshape parameter. There was a 76 percent match on movement and place of articulation *vs.* a 44 percent match on handshape. The difference consisted in the greater number and complexity of the handshapes used by signers.[1]

Categorical perception is a second way that handshape in gesture and sign languages behaves differently; that is, native signers of American Sign Language (ASL) exhibit categorical perception for handshape (Emmorey *et al.* 2003; Baker *et al.* 2005), but gesturers do not. In Figure 9.1, 11 equal intervals between two handshapes are shown; static images were used in this study. These two handshapes are contrastive in ASL, and are seen here in the minimal pair PLEASE *vs.* SORRY . Categorical perception is described in the psycholinguistic literature as being present when performance on an identification task and a discrimination task produces a boundary between the categories that is in the same place along the continuum (CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). In this case the identification task consisted of putting the images shown in Figure 9.1 into either the or group. The discrimination task consisted of taking images that are two intervals apart and asking participants to judge whether the two forms were the same or different.

In native signers the boundary between the and groups in the identification task and the strongest point of discrimination were in the same place, namely between intervals 4 and 6. Non-signers did not display categorical perception for these handshapes. Moreover, the native signers showed categorical perception for only the pairs of handshapes known to be contrastive; other pairs of stimulus handshapes that are not contrastive in ASL did not yield results indicating categorical perception. Handshapes that are considered allophonic and contrastive are described in detail in §2. The phenomenon of categorical perception does not demonstrate conclusively that a phenomenon is phonological in nature (Harnad 1987),

[1] A greater range of handshape forms is also seen in signers when compared with those of homesigners (Singleton *et al.* 1993; Goldin-Meadow 2001). Homesigners are deaf individuals who have not been exposed to a language model for a variety of reasons, and who have invented their own system of gestures to communicate with their family and acquaintances.

Figure 9.5 Two signs that are a minimal pair in ASL for joints. All properties of CANDY (a) and APPLE (b) are identical, except that CANDY has a straight index finger and APPLE has a bent index finger

is different because of the joint configuration of that finger: in CANDY it is straight; in APPLE it is bent (Figure 9.5). As mentioned previously, unselected fingers allow only extended and flexed specifications, but selected fingers allow for several more. The seven different positions that are contrastive in ASL are defined in (3) and shown in (4).

(3) *Definitions of joint configurations in handshape* (cf. Brentari 1998: 106–109)

a. Fully open: no joints of the hand are flexed.
b. Bent (closed): non-base joints flexed.
c. Flat-open: base joints flexed less than 90°.
d. Flat-closed: base joints flexed more than 90°.
e. Curved open: base and non-base joints flexed without contact.
f. Curved closed: base and non-base joints flexed with contact.
g. Fully closed: base and non-base joints fully flexed.

(4) *Contrastive positions of the joints in ASL*

(3a) (3b) (3c) (3d) (3e) (3f) (3g)

In addition to its ability to create minimal pairs, the joint position can change during the articulation of a single sign, as seen earlier in the HKSL sign SPECIAL

(Figure 9.3). These changes are called aperture changes, and they occur in a predictable fashion. Each joint configuration in (3) is assigned membership to either the open or the closed class, determined by its degree of closure and distribution in the system. If there are two handshapes in a sign, one must be from the open class and the other from the closed class. This was also observed by Mandel (1981) for ASL, first formalized in Sandler (1987) as the Handshape Sequencing Constraint, and further refined by Brentari (1990, 1998) as the Maximize Aperture Constraint. A recent formulation of this constraint is given in (5). This phonotactic constraint was also used in the word-segmentation task described in §1.1.

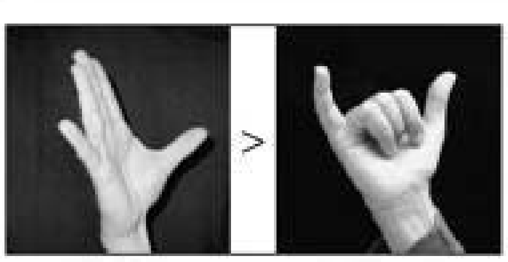
(5) *Handshape Sequencing Constraint* (Sander and Lillo-Martin 2006: 154)

If there are two joints positions in a sign, one must be [open] and the other [closed].

Examples of handshape sequences that are and are not permitted in a single sign are shown in (6) and (7). In (6a) the index, middle, and ring fingers change from open to closed, and in (6b) and (6c) all of the fingers change from closed to open in a way that conforms to the constraints in (2) and (5). Consequently, all things being equal, these sequences of handshapes should be perceived as single signs by signers of ASL. In contrast, the handshape sequences in (7) all violate the constraints in (2) or (5) in some way, and, all things being equal, such sequences ought to be perceived as two signs by ASL signers. The sequence in (7a) has a change in selected finger groups, violating the constraint in (2), and the sequence in (7b) and (7c) violates the constraint in (5) because the two handshapes in the sequence are either both closed (7b) or both open (7c). The Prosodic Model Notation System for handshape (Eccarius and Brentari 2008) is employed from now on in this chapter to underscore the differences and similarities among handshapes. This notation is explained in Appendix 9.1.

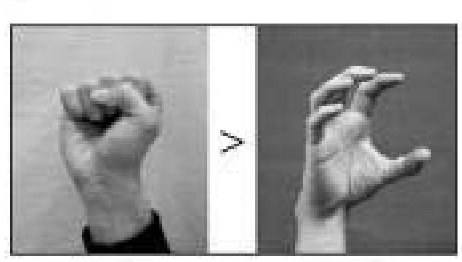
(6) *Permissible handshape changes within a sign*

a.



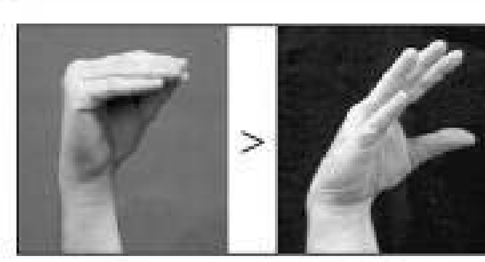
MT;/ > M@;/

b.



BT@ > BT^c

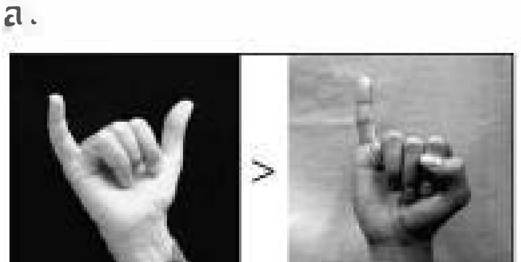
c.



BT> > BT^<

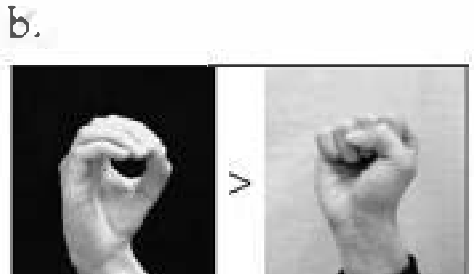
(7) *Impermissible handshape changes within a sign*

a.



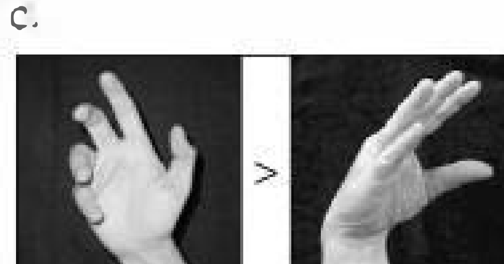
*M@;/ > BT@

b.



*BTo > BT@

c.



*BT-ĸc > BT^<

The different roles of joint features were formalized by van der Hulst (1995), who divided them into two different classes: those used for aperture change with their limited set of possibilities (i.e. "open" or "closed" variants of the underlying handshape) and joint selection for the seven distinctive joint contrasts.

In addition to selected fingers, unselected fingers, and joints, handshape can include other structures that will not be further discussed here due to space limitations. The thumb has the possibility of being a selected finger, and sometimes behaves differently than the other selected fingers in a sign. The arm can sometimes form a part of the larger articulatory structure. And two hands are sometimes used to articulate a sign; the dominant hand (H1) is the hand used to sign one-handed signs in a neutral linguistic context. The non-dominant hand (H2) is the other hand used in two-handed signs. This topic is covered in CHAPTER 10: THE OTHER HAND IN SIGN LANGUAGE PHONOLOGY. In addition, orientation of the hand in space is considered part of the handshape structure; this will be discussed in §3.

Based on the facts just presented, the areas of consensus among current models of sign language phonology about handshape are given in Table 9.1 (see Sandler 1989; Brentari 1990, 1998; van der Hulst 1993, 1995; Uyechi 1995; Channon 2002a, 2002b; van der Kooij 2002; Sandler and Lillo-Martin 2006). The generalizations in Table 9.1 extend to all sign languages studied thus far.

The hierarchical feature structure that includes the characteristics in Table 9.1 is given in (8). Notice the articulatory distinction between selected and unselected fingers, discussed in §2.1, as well as the feature distinction between the joints and fingers, discussed in §2.2.

Table 9.1 General characteristics of handshape

Structural property	*Description*
Feature geometry	The internal structure of handshape is best accounted for using a specific hierarchy of features that captures the relationships among the substructures of handshape.
H1/H2	An articulatory distinction of handshape that captures the roles of the two hands in a two-handed sign: the dominant hand (H1) and the other the non-dominant hand (H2) (see CHAPTER 10: THE OTHER HAND IN SIGN LANGUAGE PHONOLOGY).
Selected/ unselected fingers	An articulatory distinction of handshape that captures which fingers are foregrounded (selected) or backgrounded (unselected). Selected fingers can move during the articulation of a sign, and also can assume a larger range and more elaborate configurations of joints. Unselected fingers can assume only two joint configurations: fully open or fully closed.
Selected fingers/joints	A feature distinction within selected fingers capturing which fingers are selected and the disposition of their finger joints.

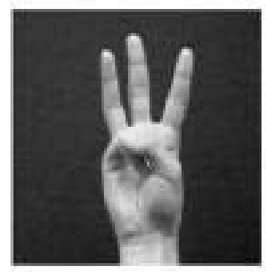
– but the index, middle, and ring fingers are selected in (10b): ASL WEIRD, and HKSL SUCCEED. In (11a) – HKSL NOT-HAVE and ASL INTERPRET – the index and thumb are selected, but in (11b) – HKSL NAME – the middle, ring, and pinkie fingers are selected.

(10) *Phonetically similar handshapes:*
Thumb + pinkie vs. *index, middle, and ring selected*

a. thumb + pinkie selected

HKSL: LITTLE
ASL: SIX
JTo;/

b. index, middle, and ring selected



HKSL: SUCCEED
ASL: WEIRD
M^;#

(11) *Phonetically similar handshapes:*
Thumb + index selected vs. *middle, ring, and pinkie selected*

a. thumb + index selected

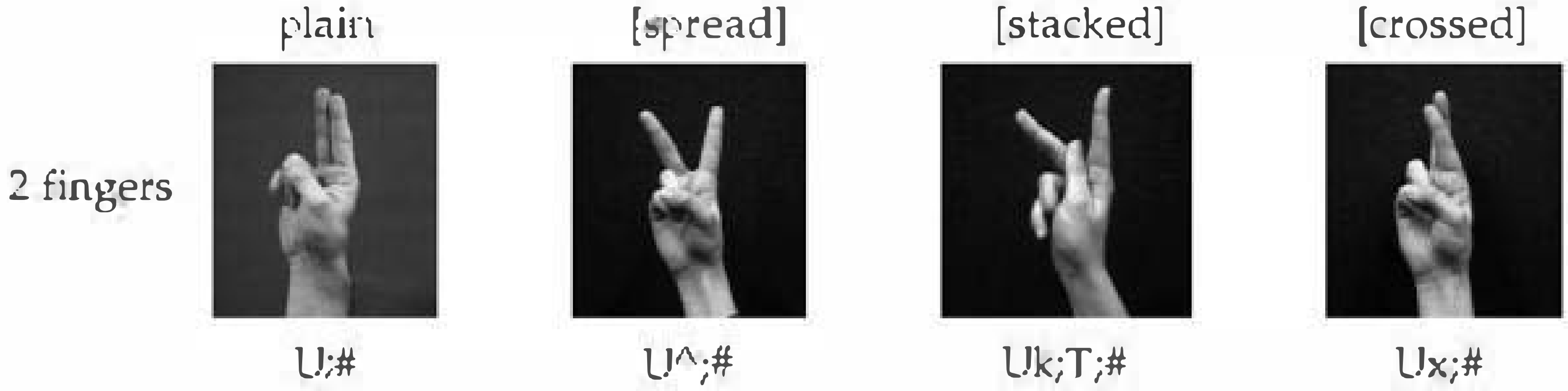
HKSL: NOT-HAVE
ASL: INTERPRET
1T@;/

b. middle, ring, and pinkie selected

HKSL: NAME
D^;#

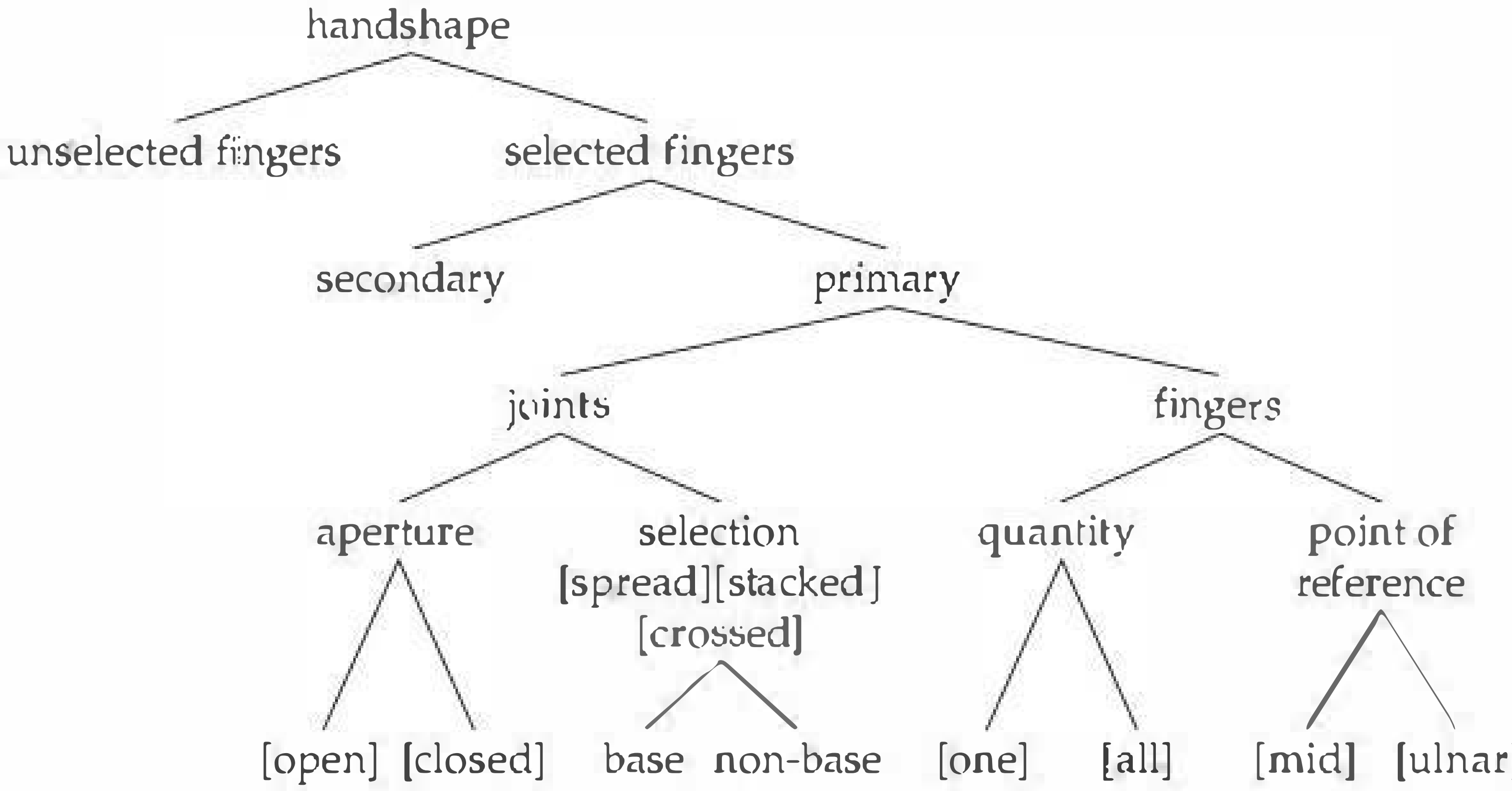
3.2 *Secondary selected fingers*

The distinction between selected and unselected fingers is sufficient to capture the majority of handshape contrasts in most sign languages, but there are a number of handshapes that require a third level of fingers, which has been called "secondary selected fingers" and can be distinguished from the "primary selected fingers" when necessary (Eccarius 2002). In such cases, the unselected fingers assume an unmarked open or closed joint position as described earlier, but the selected fingers have two different sets: one that is more marked, requiring the more elaborate specification (i.e. the primary selected fingers), and one that is less marked, but which still requires more specification than the unselected fingers (i.e. the secondary selected fingers). Cross-linguistic work – particularly from the sign languages of Asia – has demonstrated the need for this structure (Eccarius 2002, 2008; Fischer and Dong 2010). Some examples are given in (12)–(13). In (12), all three levels are represented: primary selected fingers, secondary selected fingers, and unselected fingers. The handshape in (12a) is from HKSL; (12b)–(12d) are from ASL.



If the elaborations of handshape structure described in §3 are added to the basic structure in (8), the full handshape structure is as in (16).

(16) *The structure of handshape with elaborations*



4 The role of handshape with respect to other parameters

The handshape structure in (16) is part of a larger feature hierarchy that includes Movement (M; see CHAPTER 24: THE PHONOLOGY OF MOVEMENT IN SIGN LANGUAGE), Place of Articulation (POA), and Orientation (OR). Models differ on how handshape should be represented with respect to the other parameters. Because of space, only three representative models are mentioned: the Hand Tier model (Sandler 1989; Sandler and Lillo-Martin 2006), the Prosodic Model (Brentari 1998), and the Dependency Model (van der Hulst 1995; van der Kooij 2002). Also because of space, only two points that bear on the issue of handshape (HS) and other phonological structures are raised: aperture change (ApChange; already discussed above) and orientation. A full discussion of the arguments for these structural differences is beyond the scope of this chapter, so the purpose of the representations below is simply to indicate where aperture change and orientation are represented in these different models. The differences arise from trying to best determine the simplest and most elegant analysis that can account for specific sets of forms. In addition, the Hand Tier Model adheres to the view that articulatory relatedness should

Table 9.2 Places of articulation internal to the hand (i.e. handparts). Each sign in the left column orientates the indicated handpart of the dominant hand toward the sign's place of articulation. Each sign in the right column uses the indicated handpart as a place of contact on the non-dominant hand during the production of the sign

	Role in orientation in ASL	Role in two-handed signs in ASL
1. finger fronts	LABEL	DISMISS
2. palm of hand	MY	LEARN
3. back of palm	LOVE (a thing)	NATIVE-AMERICAN
4. back of fingers	CHERISH	EASY
5. radial side of fingers	OLD	WOOD
6. ulnar side of fingers	BROKE ('no money')	TICKET
7. tip of fingers/thumb	COMPLAIN	TOP
8. heel of hand (wrist)	SLIP	CHEESE

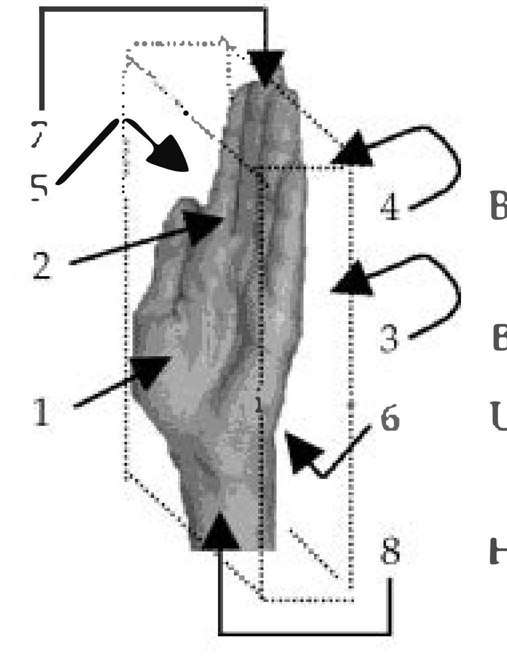
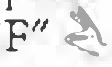


Figure 9.6 The eight surfaces of the hand used to indicate orientation

5 Handshape at the grammatical interfaces

Handshape has received a great deal of attention on purely phonological grounds, as we have seen in the previous sections. It has also been used to describe phenomena at the grammatical interfaces between phonology and other components of the grammar. Here we address handshape phenomena at the phonology–phonetics, phonology–morphology, and phonology–syntax interfaces. Each of the phenomena addressed in §5.2–§5.4 involves one or more of the phonological structures discussed earlier in this chapter. In order to understand how these phenomena are manifested in a sign language grammar, some background on the architecture of a sign language lexicon is needed.

5.1 Handshape in the core–periphery framework

Many sign languages have multiple origins, resulting in a lexicon divided into subcomponents according to the historical origins of signs, as well as their morphological and phonological behavior. For a model addressing similar facts for spoken languages, see Itô and Mester (1995a, 1995b); the sign language model is presented of these signs corresponds to the first letter of the English word (e.g. WATER, CAFETERIA, BACHELOR, and DIVORCE are all initialized signs). The spatial lexicon consists primarily of the classifier constructions, which are used to express events of motion and location. Classifier constructions are polymorphemic verbal complexes with a root – the movement – and affixes that involve place of articulation and handshape. The classifier morphemes that refer to the arguments of the predicate are the handshapes (Janis 1992; Benedicto and Brentari 2004). In classifier predicates, parts of the handshape structure can carry information about the size and shape of the object that may be phonological, morphological, and iconic at the same time. In Figure 9.7b examples of signs in each of the components using the same handshapes are shown: "F" , "O" , and "C"  handshapes.

5.2 *The phonology–phonetics interface and joint features*

A recent study was conducted to address whether the phonetic targets for joints were the same across the three lexical components described above (Eccarius 2008; Eccarius and Brentari, forthcoming). The stimuli consisted of video clips of core vocabulary, initialized signs from the foreign component, and classifier forms from the spatial component. Three handshape groups were tested ("O," "F," and "C"). Each item included three variants of the handshape (round, intermediate, and flat) at the same place of articulation using the same movement. The variants were designed to address the stability of joint specifications throughout the lexicon. A foreign item using the "C" handshape is given in Figure 9.8, with the round, intermediate, and flat forms. The round form is the sign for CAFETERIA and the flat form is the sign for BACHELOR (a minimal pair); the intermediate form could be interpreted as either sign or neither of them. Twelve Deaf signers who were native or early learners of ASL participated in this experiment.[10]

A total of 154 randomized stimulus video clips, balanced across lexical component (core, foreign, spatial), handshape group (O, F, C), and variant (round, intermediate, flat) were shown to each participant. Participants were asked to watch each clip and to respond to two questions presented immediately afterwards on the computer screen: "What is the best English meaning for this sign?" and "Please rate the handshape of this sign," from very bad to very good, on a five-point scale.

Figure 9.8 The round, intermediate, and flat variants of "C" used with the target signs CAFETERIA (left), BACHELOR (right), and an intermediate form (middle)

[10] A "native signer" is someone with two Deaf signing parents; an "early learner" is defined here as someone who learned ASL before age five.

Figure 9.9 (a) A whole entity and (b) a handling classifier handshape in ASL

of selected finger complexity.[12] Low-complexity handshapes have the simplest phonological representation (Brentari 1998), are the most frequent handshapes cross-linguistically (Hara 2003; Eccarius and Brentari 2007) and are the earliest handshapes acquired by native signers (Boyes Braem 1981). These three criteria converge on the set of handshapes represented in (18a): all fingers, index finger, and thumb. These three handshapes have also been found to be frequent in the spontaneous gestures that accompany speech (Singleton *et al.* 1993) and in child homesign (Goldin-Meadow *et al.* 1995). Medium-complexity handshapes include one additional elaboration of the finger representation of a one-finger handshape. The elaboration indicates that the single selected digit is not on the "radial" (thumb) side of the hand (the default specification) – e.g. in the finger is on the [ulnar] "pinkie" side of the hand and is [mid] "middle" – or that there is an additional finger selected, as in where two fingers are selected rather than one. High-complexity handshapes include all other handshapes, e.g. and.

(18) *Low- and medium-complexity handshapes* (Brentari 1998; Eccarius and Brentari 2007)

a. *low complexity*

- (B): fingers – quantity [all]
- (1): fingers – quantity [one]
- (T): selected fingers – thumb

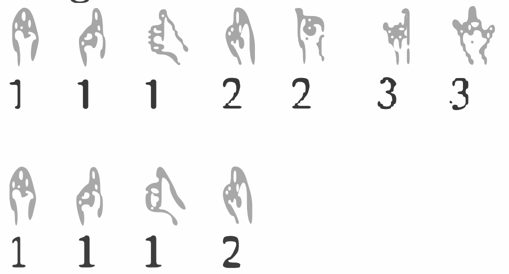
b. *medium complexity*

- (I): fingers – quantity – (quantity [one], POR [ulnar])
- (8): fingers – quantity – (quantity [one], POR [mid])
- (U): fingers – quantity [one] – [all]

[12] Joint complexity has been set aside for future analysis and is not considered here.

Using both elicitation methods and grammaticality judgments, Eccarius (2008) and Brentari and Eccarius (2010) found that whole entity and handling classifiers differ systematically in their distribution of finger complexity. Whole entity classifiers have a larger set and a more complex set of finger contrasts than handling classifiers (see (19); the level of complexity is shown beneath each handshape).[13] Both types of classifier handshapes have the potential to use selected fingers to show the size and shape of an object, yet only whole entity classifiers do this extensively.

(19) *Number of finger contrasts for classifier handshapes in ASL, HKSL, and DSGS, with the finger complexity given for each*



	Finger contrasts						
Object CL-HSs	1	1	1	2	2	3	3
Handling CL-HSs	1	1	1	2			

This distribution is a part of the phonology–morphology interface for several reasons. First, the morphological categories for whole entity and handling classifiers are associated with a corresponding phonological pattern, but this state of affairs is not obligatory. For example, if a hypothetical morphological classifier type in a sign language comprised the following handshapes – – the set would also form a phonological class, because the handshapes in the set share a phonological property (their joints are all fully extended). New handshapes that enter the set would be predicted to be fully extended as well. If, instead, the hypothetical classifier type comprised this set of handshapes – – the set would not form a phonological class as there is no common property that the handshapes share. In this event, the handshapes would constitute a morphological class, but not a phonological one. In the case discussed here, finger complexity provides a unifying phonological property associated with whole entity and handling classifiers; high complexity is associated with whole entity classifiers and low complexity with handling classifiers. New forms entering the system are therefore expected to conform to this pattern. Second, the phenomenon is morphophonological because it is tied to specific morphological classes, similar to the way that Trisyllabic Laxing is tied to English words having specific Class 1 affixes (CHAPTER 94: LEXICAL PHONOLOGY AND THE LEXICAL SYNDROME). And third, this phenomenon is phonological because selected fingers are phonological throughout the system, as has been discussed earlier in this chapter, and whatever iconicity might exist in classifiers is subject to phonological pressures on the system as a whole, such as ease of articulation and ease of perception (Eccarius 2008).

This work expands the range of generalizations possible in sign language phonology, specifically with regard to handshape, which is important in its own right. In addition, it provides a new way to consider the emergence of phonology as it has evolved from gesture to homesign and from homesign to a sign language. Sign languages utilize iconicity, but it is not used in the same way by gesturers,

[13] Handshapes in (19) are abstract classes, so, for example, the handshape , with or without the thumb, represents any handshape with one selected finger, i.e. one finger active, e.g. , . The thumb is ignored in our analysis, except when it is the only finger selected.

The second diagnostic syntactic test is sensitive to objects/themes. The sign languages we have investigated have a form for the "distributive," with the meaning '. . . to each'. As above, sentences with verbs from the core lexicon are tested first (22). The sentence in (22a) with MELT, an unaccusative verb, is grammatical with the 'to-each' form. The sentence in (22b) with LAUGH, an unergative verb, obtains an ungrammatical result for the intended meaning 'Each woman laughed.' The only interpretation of this sentence that is acceptable is 'The woman laughed at each [of them].' As above, the sentences in (23) express a minimal difference in meaning and form (i.e. in both cases each book is moving onto its side). The form in (23a) contains the whole entity classifier and the one in (23b) contains the handling classifier, the same two forms discussed above. The results of this test determine that the sentences in both (23a) and (23b) are grammatical because both the whole entity (23a) and handling (23b) classifiers contain an object/theme. The whole entity classifiers are therefore intransitive since only an object is present, and the handling classifiers transitive because both an object and subject are present.

(22) *The Distributive Test applied to core lexical verbs*

a. BUTTER MELT+ 'to-each'
butter melt DISTRIBUTIVE
'Each butter melted.'

b. WOMAN LAUGH + 'to-each'
woman laugh DISTRIBUTIVE
\# 'Each woman laughed.'
'The woman laughed at each of them.'

(23) *The Distributive Test applied to classifiers*

a. BOOK B/+ MOVE + 'to-each'
book OBJECT CLASSIFIER + move + DISTRIBUTIVE
'Each of the books fell down (on its side).'

b. BOOK C/+ MOVE + 'to-each'
book HANDLING CLASSIFIER + move + DISTRIBUTIVE
'Each book was put on its side.'

This transitive–intransitive alternation concerns the phonology because it can be captured by a phonological analysis specifying which nodes of the feature tree are associated with this syntactic distinction. The phonological element involved in the transitive–intransitive alternation concerns the differential use of orientation of the hand.[15] In whole entity classifiers orientation is purely phonological and does not interact with the syntax. In handling classifiers, the parts of the hand shown in Figure 9.6 can be used morphologically, and these alternations also interact with the syntax to express an agent.

The studies presented here provide a few examples of how phonology interacts with the phonetic, morphological, and syntactic components. In the phonology–phonetics example, it was the joint features that were used. In the phonology–morphology example, it was the selected fingers that allowed for the generalization

[15] As stated in §4, orientation is represented differently in the phonological models available; however, these handpart features are part of the handshape structure in all models.

be searched for. This notation also captures the case of two handshapes that are very similar phonetically, but which belong to different handshape groups because of differences in their selected fingers, such as examples (10) and (11) in the text. A complex handshape meaning SPACE-SHUTTLE is transcribed below:

1. finger base symbol
2. thumb base symbol
3. thumb joint symbol
4. spread "stacked" and/or "crossed" joint symbol(s)
5. remaining joint symbol
6. secondary selected fingers symbols (4 are possible)
7. unselected fingers symbol

1	2	3	4	5		6					7
U			x		;	J	T			;	#

primary selected fingers (1–5); secondary selected fingers (6); unselected fingers (7)

See Eccarius and Brentari (2008) for full details of the notation system.

ACKNOWLEDGMENT

This work was supported in part by NSF Grant BCS 0547554 to the author.

REFERENCES

Allen, J. Sean & Joanne L. Miller. 2001. Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception and Psychophysics* 63. 798–810.

Baker, Stephanie, William J. Idsardi, Roberta Golinkoff & Laura Petitto. 2005. The perception of handshapes in American Sign Language. *Memory and Cognition* 33. 887–904.

Benedicto, Elena & Diane Brentari. 2004. Where did all the arguments go? Argument changing properties of Classifiers in ASL. *Natural Language and Linguistic Theory* 22. 743–810.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Boyes Braem, Penny. 1981. Distinctive features of the handshapes of American Sign Language. Ph.D. dissertation, University of California, Berkeley.

Brentari, Diane. 1990. Theoretical foundations of American Sign Language phonology. Ph.D. dissertation, University of Chicago.

Brentari, Diane. 1998. *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

Brentari, Diane (ed.) 2010. *Sign languages: A Cambridge language survey*. Cambridge: Cambridge University Press.

Brentari, Diane & Petra Eccarius. 2010. Handshape contrasts in sign language. In Brentari (2010), 284–311.

Brentari, Diane & Carol Padden. 2001. Native and foreign vocabulary in American Sign Language: A lexicon with multiple origins. In Diane Brentari (ed.) *Foreign vocabulary in sign languages*, 87–119. Mahwah, NJ: Lawrence Erlbaum.

Brentari, Diane, Marie Coppola, Laura Mazzoni & Susan Goldin-Meadow. 2010. When does a system become phonological? Handshape production in gesturers, signers, and homesigners. Unpublished ms., Purdue University.

Brentari, Diane, Carolina González, Amanda Seidl & Ronnie Wilbur. Forthcoming. Sensitivity to visual prosodic cues in signers and nonsigners. *Language and Speech*.

Brentari, Diane, Harry van der Hulst, Els van der Kooij & Wendy Sandler. 1996. [One] Over [All]; [All] Over [One]: A Dependency Phonology analysis of handshape in sign languages. Unpublished ms., Purdue University, University of Connecticut & Haifa University.

Channon, Rachel. 2002a. Signs are single segments: Phonological representations and temporal sequencing in ASL and other sign languages. Ph.D. dissertation, University of Maryland at College Park.

Channon, Rachel. 2002b. Beads on a string? Representations of repetition in spoken and signed languages. In Richard Meier, David Quinto-Pozos & Kearsy Cormier (eds.) *Modality and structure in signed and spoken languages*, 65-87. Cambridge: Cambridge University Press.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. 2001. Representational economy in constraint-based phonology. In T. A. Hall (ed.) *Distinctive feature theory*, 71–146. Berlin & New York: Mouton de Gruyter.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

Corina, David. 1990. Reassessing the role of sonority in syllable structure: Evidence from a visual-gestural language. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26(2). 33–44.

Crasborn, Onno, Harry van der Hulst & Els van der Kooij. 2001. SignPhon: A phonological database for sign language. *Sign Language and Linguistics* 4. 215–228.

Crasborn, Onno & Els van der Kooij. 1997. Relative orientation in sign language phonology. In Jane Coerts & Helen de Hoop (ed.) *Linguistics in the Netherlands* 1997, 37–48. Amsterdam & Philadelphia: John Benjamins.

Eccarius, Petra. 2002. Finding common ground: A comparison of handshape across multiple sign languages. M.A. thesis, Purdue University.

Eccarius, Petra. 2008. A constraint-based account of handshape contrast in sign languages. Ph.D. dissertation, Purdue University.

Eccarius, Petra & Diane Brentari. 2007. Symmetry and dominance: A cross-linguistic study of signs and classifier constructions. *Lingua* 117. 1169–1201.

Eccarius, Petra & Diane Brentari. 2008. Handshape coding made easier: A theoretically based notation for phonological transcription. *Sign Language and Linguistics* 11. 69–101.

Eccarius, Petra & Diane Brentari. Forthcoming. Contrast differences across lexical substrata: Evidence from ASL handshapes. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 43.

Emmorey, Karen & Melissa Herzig. 2003. Categorical *vs.* gradient properties in classifier constructions in ASL. In Karen Emmorey (ed.) *Perspectives on classifier constructions in sign languages*, 221–246. Mahwah, NJ: Lawrence Erlbaum.

Emmorey, Karen, Stephen McCullough & Diane Brentari. 2003. Categorical perception in American Sign Language. *Language and Cognitive Processes* 18. 21–45.

Engberg-Pederson, Elisabeth. 1993. *Space in Danish Sign Language*. Hamburg: Signum Verlag.

Fischer, Susan D. & Qunhu Gong. 2010. Variation in East Asian sign language structures. In Brentari (2010), 499–518.

Frishberg, Nancy. 1975. Arbitrariness and iconicity: Historical change in American Sign Language. *Language* 51. 696–719.

Goldin-Meadow, Susan, David McNeill & Jenny Singleton. 1996. Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality. *Psychological Review* 103. 34–55.

Goldin-Meadow, Susan, Carolyn Mylander & Cynthia Butcher. 1995. The resilience of combinatorial structure at the word level: Morphology in self-styled gesture systems. *Cognition* 56. 195–262.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.

Hara, Daisuke. 2003. A complexity-based approach to the syllable formation in sign language. Ph.D. dissertation, University of Chicago.

Harnad, Steven. 1987. *Categorical perception: The groundwork of cognition*. Cambridge: Cambridge University Press.

Hulst, Harry van der. 1993. Units in the analysis of signs. *Phonology* 10. 209–241.

Hulst, Harry van der. 1995. The composition of handshapes. *University of Trondheim Working Papers in Linguistics*. 1–18.

Hulst, Harry van der & Rachel Channon. 2010. Notation systems. In Brentari (2010), 152–172.

Itô, Junko & Armin Mester. 1995a. Japanese phonology. In Goldsmith (1995), 817–838.

Itô, Junko & Armin Mester. 1995b. The core–periphery structure of the lexicon and constraints on reranking. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 181–209. Amherst: GLSA.

Janis, Wynn. 1992. Morphosyntax of ASL verb phrase. Ph.D. dissertation, State University of New York, Buffalo.

Johnson, Robert E. 1990. Distinctive features for handshapes in American Sign Language. Paper presented at the International Conference on Theoretical Issues in Sign Language Research, Boston.

Jusczyk, Peter W., Anne Cutler & Nancy J. Redanz. 1993. Preference for the predominant stress patterns of English words. *Child Development* 64. 675–687.

Jusczyk, Peter W., Elizabeth Hohne & Angela Bauman. 1999. Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics* 61. 1465–1473.

Kooij, Els van der. 2002. Phonological categories in Sign Language of the Netherlands: Phonetic implementation and iconic motivation. Ph.D. dissertation, University of Leiden.

Ladefoged, Peter. 1993. *A course in phonetics*. 3rd edn. New York: Harcourt Brace Jovanovich.

Liddell, Scott K. & Robert E. Johnson. 1986. American Sign Language compound formation processes, lexicalization and phonological remnants. *Natural Language and Linguistic Theory* 4. 445–513.

Liddell, Scott K. & Robert E. Johnson. 1989. American Sign Language: The phonological base. *Sign Language Studies* 64. 197–277.

Mandel, Mark A. 1981. Phonotactics and morphophonology in American Sign Language. Ph.D. dissertation, University of California, Berkeley.

Mazzoni, Laura. 2008. *Classificatori e impersonamento nella Lingua dei Segni Italiana*. Pisa: Plus.

Padden, Carol, Irit Meir, Mark Aronoff & Wendy Sandler. 2010. The grammar of space in two new sign languages. In Brentari (2010), 570–592.

Poizner, Howard. 1983. Perception of movement in American Sign Language: Effects of linguistic structure and linguistic experience. *Perception and Psychophysics* 33. 215–231.

Prillwitz, Siegmund, Regina Leven, Heiko Zienert, Thomas Hamke & Jan Henning. 1989. *Hamburg notation system for sign language: An introductory guide*. Hamburg: Signum.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT. Published 1990, New York: Garland.

Sandler, Wendy. 1987. Assimilation and feature hierarchy in American Sign Language. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 266–278.

Sandler, Wendy. 1989. *Phonological representation of the sign: Linearity and non-linearity in American Sign Language*. Dordrecht: Foris.

123 Hungarian Vowel Harmony

Miklós Törkenczy

1 Introduction

Vowel harmony is undoubtedly the best known and most studied aspect of Hungarian phonology. This is especially true if one has the international literature in mind: other Hungarian phonological phenomena have received relatively little attention. There is a host of analyses of Hungarian vowel harmony, pre-generative, "mainstream" generative (especially post-*SPE*), including optimality-theoretic approaches, and "non-mainstream" generative.[1] The reason for this attention is that Hungarian vowel harmony has been used as a testing ground for various theories and theoretical tools/devices of generative phonology. This is primarily due to (the coexistence of) a number of features displayed by Hungarian vowel harmony: long-distance effects, the transparency of neutral vowels and the variation in harmony associated with them, the intricate patterning of alternating and invariant suffixes, the existence of antiharmonic roots, and the combination and very different character of backness and roundness harmony. Attention increased especially with the advent of autosegmental phonology and underspecification theory (e.g. Clements 1976; Vago 1976, 1978, 1980a; Jensen 1978; Ringen 1978, 1980, 1988a; Booij 1984; van der Hulst 1985), and again with the recent shift of research towards phonetic motivation (e.g. Beňuš *et al.* 2003; Beňuš and Gafos 2005, 2007) and variation in phonology (e.g. Hayes and Londe 2006; Hayes *et al.* 2009). Because of this theoretical bias in the research into Hungarian vowel harmony, the linguist reading the literature may learn much, but also too little: the available analyses tend to focus only on some aspects of Hungarian vowel harmony[2] – typically those crucial to the theoretical issue at hand – and tend to disregard/ignore others, thus sometimes failing to do justice to the complexities and intricacies of Hungarian vowel harmony. Therefore, instead of surveying the analyses (which would not be useful or possible because of the sheer number of – often very different – competing analyses and the transitory nature of some

[1] For a bibliography of Hungarian vowel harmony, which is intended to be fairly exhaustive, see the further reading section below.

[2] Some comprehensive treatments are notable exceptions, e.g. Vago (1980c), Abondolo (1988), Olsson (1992), and Siptár and Törkenczy (2000).

of the theoretical problems some of them focus on), I will concentrate on the empirical aspect and the major analytic issues of Hungarian vowel harmony. In what follows, after a brief review of the Hungarian vowel system (§2), I will discuss in detail the properties of Hungarian vowel harmony in §3, attempting to give comprehensive descriptive coverage. I will pay special attention to areas where vowel harmony interacts with other, lesser-known (or rarely discussed) areas of Hungarian phonology. Finally, in §4 I will briefly survey the major analytic issues that a phonological treatment of Hungarian vowel harmony has to address. Throughout, I will try to be as theoretically neutral as possible and will take pains to highlight exceptional behavior or data/patterns not (sufficiently) understood.

2 Background: The Hungarian vowels

The phonological classification of the Standard Hungarian[3] vowels is given in (1) below, where the italicized characters that appear next to the IPA symbols are the vowel letters of Hungarian orthography. The vowel spelled <a> is transcribed as [ɔ], [ɒ], or [ɑ] in the literature. Phonetically, this vowel is more open than [ɔ] and less open than [aː ɑ]. It has considerably less lip rounding than [ɔ ɒ], but more than [ɑ]. As nothing hinges on this matter as far as vowel harmony is concerned, I will (arbitrarily) use the symbol [ɔ] for <a>.

(1)

	front		back	
	unrounded	rounded	unrounded	rounded
high	iː *í* i *i*	yː *ű* y *ü*		uː *ú* u *u*
mid	eː *é*	øː *ő* ø *ö*		oː *ó* o *o*
low	ɛ *e*		aː *á*	ɔ *a*

Although there are subtle phonetic differences between the heights of non-high long and short vowels, and [aː] is phonetically central, it is usually assumed that phonologically it is sufficient to distinguish only three vowel heights and two degrees of backness (see Siptár and Törkenczy 2000). The long–short pairs are established on the basis of length alternations like those in (2):[4, 5]

[3] It is also referred to as Educated Colloquial Hungarian (ECH), spoken in Budapest; see Nádasdy (1985). There is very little on Hungarian vowel harmony in dialects; see Rebrus (2000a), Ringen and Szentgyörgyi (2000), and Ringen and Vago (1998).

[4] There are sporadic cases of other pairings, e.g. [oː ~ ɔ]: *hó* ~ *hav-at* [hoː ~ hɔvɔt] 'snow (NOM ~ ACC)'.

[5] Hungarian consonant letters typically have transparent phonetic values, except for *sz* [s], *s* [ʃ], *zs* [ʒ], *cs* [ʧ], *c* [ts], *ny* [ɲ], *ty* [c], *gy* [ɟ], *j*, *ly* [j]. The doubling of a consonant letter or the first letter of a consonant digraph indicates a geminate consonant: *tt* [tː], *ggy* [ɟː]. As can be seen in (1), a single or a double acute accent above a vowel letter denotes length, not stress (which is always on the first syllable of the word). I will use capital letter abbreviations for harmonizing vowels in suffixes, as is customary in analyses of Hungarian vowel harmony (see §3.3.1). Wherever necessary, I will indicate roots by underlining and affixes in bold. The symbol "≈" will be used to indicate vacillation.

(2)

	nominative		*accusative*		
iː ~ i	*víz*	[viːz]	*viz-et*	[vizɛt]	'water'
yː ~ y	*tűz*	[tyːz]	*tüz-et*	[tyzɛt]	'fire'
uː ~ u	*kút*	[kuːt]	*kut-at*	[kutɔt]	'well'
øː ~ ø	*kő*	[køː]	*köv-et*	[køvɛt]	'stone'
oː ~ o	*ló*	[loː]	*lov-at*	[lovɔt]	'horse'
eː ~ ɛ	*kéz*	[keːz]	*kez-et*	[kɛzɛt]	'hand'
aː ~ ɔ	*nyár*	[ɲaːr]	*nyar-at*	[ɲɔrɔt]	'summer'

Given the classification in (1) the paired vowels differ in length only phonologically, except for *é* ~ *e* [eː ~ ɛ], which differ in height too, and *á* ~ *a* [aː ~ ɔ], which also differ in rounding. These predictable differences are often abstracted away from, and seen as the result of, late rules or phonetic interpretation.

3 Vowel harmony

Hungarian has both backness (palatal) and roundness (labial) harmony (see CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS and CHAPTER 118: TURKISH VOWEL HARMONY for more discussion of vowel harmony). Both are root-controlled or, more precisely, stem-controlled, i.e. the harmonic properties of the root/stem determine the harmonic properties of the affix(es). Directionality is left-to-right, but this does not have to be stipulated since all prefixes are outside the harmonic domain, which consists of root+ suffixes. Phonologically, prefix+ root combinations behave like compounds, in that each prefix forms its own harmonic domain, and thus does not harmonize with the root. Compare compound *fa* + *gyöngy* [fɔɟøɲɟ] 'mistletoe' (literally 'tree-pearl') and pre-verb + verb *át* + *köt* [aːtkøt] 'tie up' (literally 'across-tie'), where "+" indicates the morpheme boundary (see the discussion of the domain issue later in §4.1). Intervening consonants do not interact with Hungarian vowel harmony; they neither block nor initiate the harmonizing features.

Backness harmony, which in general requires that vowels should agree in backness within the harmonic domain, is based on the following groupings:

(3)

Back	B	u uː o oː ɔ aː
Front rounded	Fr	y yː ø øː
Neutral: front unrounded	N	i iː ɛ eː

Neutral vowels form a separate group, since they may in general interrupt a back harmonic span and may behave as transparent and/or opaque (see §3.4.1 for the details).

While backness harmony is pervasive in the phonology, roundness harmony is restricted. In general it requires that the *front* alternants of some suffixes should agree in roundness with the stem – in this sense it is parasitic on backness harmony (see §3.5 for details).

3.1 Roots

With respect to backness harmony, roots can be classified according to (a) the distribution of back, front rounded, and neutral vowels in them, and (b) the

3.2 The distribution of vowels in roots

(4) shows that Fr, B, and N vowels freely combine in (bisyllabic) monomorphemic roots, as all possible combinations exist. However, this picture is quite misleading, since the type frequency of some types of roots in the lexicon is strikingly lower than that of others. Notably, monomorphemic roots (bisyllabic or longer) in which Fr and B vowels combine in any order (types a.ii and b.i) are very rare (these are called "disharmonic stems"). For example, in a dictionary of about 80,000 lexical items I have found only 11 monomorphemic roots in which [ø] and [ɔ/aː] combine (e.g. *manöken* [mɔnøkɛn] 'model', *pönálé* [pønaːleː] 'penalty'), 12 in which [øː] and [ɔ/aː] combine (e.g. *amőba* [ɔmøːmɔ] 'amoeba', *amatőr* [ɔmɔtøːr] 'amateur'), and only one in which [øː] and [u] (*krőzus* [krøːzuʃ] 'very rich person') combine. That this is not due to the incompatibility of back and front vowels in monomorphemic roots can be seen if we examine the type frequency of "mixed" roots, i.e. roots in which B and N vowels combine in any order, namely types (b.iii) and (c.ii) in (4). There are many more mixed roots (of any length) than disharmonic roots. In the same search I found 1,084 monomorphemic roots in which [ɛ] and [ɔ/aː] combine (e.g. *ceruza* [tsɛruzɔ] 'pencil', *kráter* [kraːtɛr] 'crater'), and 149 in which [iː] and [ɔ/aː] combine (e.g. *papír* [pɔpiːr] 'paper', *hínár* [hiːnaːr] 'seaweed') in any order. Even though the frequency of [ɛ] or [eː] is known to be much higher than that of [ø] or [øː], the difference is telling. Thus, disharmonic roots, which are typically recent loans,[6] can be seen as special or irregular, while mixed roots – together with "simple roots" (roots not disharmonic or mixed) – are considered normal or regular (this is assumed in most analyses). Although (4) shows monosyllabic and bisyllabic roots only, as we have implied in the discussion above, the simple/mixed/disharmonic distinction carries over to longer roots as well: a root of any length is (a) simple if it contains Fr vowels only, B vowels only, N vowels only, or only Fr and N vowels, (b) mixed if *only B* and N vowels combine in it, and (c) disharmonic if it contains B and Fr vowels. A root that contains B, Fr, and N vowels is disharmonic (e.g. *manöken*). The frequency of the mixed type as compared to the disharmonic type is (part of) the lexical motivation for considering front unrounded vowels neutral. These are lexical classes independent of the harmonic behavior a root induces in a harmonizing suffix.

Roots are stable in their classes: there is typically no vacillation or allomorphy that involves harmony. If the root is subject to some alternation, the alternants are faithful in that they both belong to the same harmonic lexical class. This is true of vowel alternations, e.g. length alternations (see (2)), and vowel–zero alternations (*bokrot* [bokrot] 'bush (ACC)', *bokorban* [bokorbɔn] 'bush (INESS)', *epret* [ɛprɛt] 'strawberry (ACC)', *eperben* [ɛpɛrbɛn] 'strawberry (INESS)'), and trivially of all alternations that only involve consonants.

There are two kinds of exceptions to this generalization. "Vacillating disharmonic roots" (Vago 1978; van der Hulst 1985) are a marginal group of stems that are optionally either mixed or disharmonic. These roots are a special case of a phenomenon in which *some* syllables with [ɛ] in *some* roots have a free (or sometimes

[6] Intuitively, disharmonic roots "feel" foreign while mixed roots do not (by virtue of their mixed nature, other phonological properties may make them feel foreign too). To the best of my knowledge, this native speaker intuition about well-formedness has never been tested experimentally.

3.3 Suffixes

3.3.1 Suffix classes

Based on the quality of the alternating vowels involved and the number of suffix alternants, Hungarian suffixes fall into three alternating classes plus one in which the suffix vowel does not alternate (see Rebrus 2000a for an exhaustive list of suffixes). In the examples below I will use the same four roots whenever possible to facilitate comparison: *bot* [bot] 'stick', *fal* [fɔl] 'wall', *év* [e:v] 'year', *tök* [tøk] 'marrow'.[8]

(7) *Quaternary suffixes*
 [o~ɔ~ɛ~ø] *o~a~e~ö* (abbreviation: @)
 example: plural -@*k*
 botok [botok], ***falak*** [fɔlɔk], ***évek*** [e:vɛk], ***tökök*** [tøkøk]

There are 17 productive suffixes of this kind. Suffix-initial vowels that show vowel–zero alternation (traditionally called "linking" vowels) typically alternate in this way.[9]

(8) *Ternary suffixes*
 [o~ɛ~ø] *o~e~ö* (abbreviation: O)
 example: allative -*hOz* 'to'
 bothoz [bothoz], ***falhoz*** [fɔlhoz], ***évhez*** [e:fhɛz], ***tökhöz*** [tøkhøz]

There are nine productive suffixes of this kind.

Quaternary and ternary suffixes show the effect of backness and roundness harmony. The difference between quaternary and ternary suffix alternations (i.e. the occurrence of [ɔ] in the former) is due to Lowering, which is independent of harmony but interacts with it; see §3.5.1.

(9) *Binary suffixes*

a. [u~y] *u~ü* (abbreviation: U)
 example: -*Unk* 'our'
 botunk [botuŋk], ***falunk*** [fɔluŋk], ***évünk*** [e:vyŋk], ***tökünk*** [tøkyŋk]

b. [u:~y:] *ú~ű* (abbreviation: Ú)
 example: -*Ú* 'having'
 botú [botu:], ***falú*** [fɔlu:], ***évű*** [e:vy:], ***tökű*** [tøky:]

c. [ɔ~ɛ] *a~e* (abbreviation: A)
 example: inessive -*bAn* 'in'
 botban [bodbɔn], ***falban*** [fɔlbɔn], ***évben*** [e:vbɛn], ***tökben*** [tøgbɛn]

d. [a:~e:] *á~é* (abbreviation: Á)
 example: adessive -*nÁl* 'at'
 botnál [botna:l], ***falnál*** [fɔlna:l], ***évnél*** [e:vne:l], ***töknél*** [tøkne:l]

e. [o~ø] *o~ö* (no abbreviation)
 example: -*nok/nök* [nok/nøk] 'person acting in X capacity'
 látnok [la:tnok] 'visionary', ***mérnök*** [me:rnøk] 'engineer'

[8] The suffix counts are based on Rebrus (2000a) and Siptár and Törkenczy (2000); non-productive suffixes and semi/non-productive Latinate/Greek suffixes have been excluded.

[9] There are four exceptions: superessive -*n/on/en/ön* [n/on/ɛn/øn] 'on' (ternary); past -*t(t)/ott/ett/ött* [t(:)/ot:/ɛt:/øt:] (ternary); 1st plural possessive 'our' -*nk/unk/ünk* [ŋk/uŋk/yŋk] (binary); 1st personal plural (verb) -*nk/unk/ünk* [ŋk/uŋk/yŋk] (binary).

(13a) is not exceptionless: *-kor* is the only exception; it is productive, invariant and has a non-neutral vowel.

(14) *Non-neutral vowels*

a. Non-neutral vowels only exceptionally occur in productive invariant suffixes.
b. It is accidental which non-neutral vowels occur in a (productive or non-productive) invariant suffix (although some generalizations can be made: e.g. Fr vowels do not occur).

This asymmetrical distribution of vowels in alternating and invariant suffixes is (part of) the lexical motivation for considering front unrounded vowels neutral. Note that the neutrality of [ɛ] is questionable in this respect, since it does not occur in invariant suffixes (see §3.4.1).

The suffix analogue of disharmonic stems (in which Fr and B vowels would combine) does not exist.

3.4 Backness harmony

(4) above shows that the final vowel of the root (i.e. the one closest to the suffix) determines backness harmony if it is Fr or B (4a, 4b):

(15) If the closest vowel is Fr, the harmonizing suffix vowel will be front; if it is B the vowel will be back.

(15) also holds if the suffix is preceded by another suffix: *bot-Unk-bAn* = *botunkban* [botuŋgbɔn] 'in our stick', *év-Unk-bAn* = *évünkben* [e:vyŋgbɛn] 'in our year'.

If the final vowel is N, the preceding vowels determine backness harmony: *papír-bÓl* = *papírból* [pɔpi:rbo:l] 'of paper', *üveg-bÓl* = *üvegből* [yvɛgbø:l] 'of glass' (see (4c)). In fact, it is not exactly the final vowel that matters in this case but the final *uninterrupted sequence* of N-voweled syllables (one or more than one (N*)):

(16) If the closest vowel, not counting the final sequence of N-voweled syllables, is Fr (FrN*), the harmonizing suffix vowel will be front (4c.i).

Again, it does not matter if the final sequence of N-voweled syllables is contained in a monomorphemic root, or a polymorphemic stem: *kümité-tÓl* = *kümitétől* [kymite:tø:l] 'from kumite [sparring in karate]', *üd-ít-Ó* = *üdítő* [ydi:tø:] 'refreshing', *kövér-ít-i-tOk* = *kövérítitek* [køve:ri:titɛk] 'you (PL) make it/him/her fat'.

3.4.1 Variation

The remaining types of stems (4c.ii–iv) are in "zones of variation" (Hayes *et al.* 2009). Here vowel harmony is not (fully) predictable. Exactly what kind of harmonic behavior a given stem induces in harmonizing suffixes is the lexically arbitrary property of each stem. There are two types of behavior: (i) lexical variation and vacillation, and (ii) lexical variation, no vacillation.

(19) *Backness harmony after BNN stems*

	suffix behavior		
NN	back	back ≈ front	front
[iː i]		*analizisnak* ≈ *analizisnek* 'analysis' [ɔnɔliːziʃnɔk ≈ ɔnɔliːziʃnɛk]	
[iː i eː]		*klarinétnak* ≈*klarinétnek* 'clarinet' [klɔrineːtnɔk ≈ klɔrineːtnɛk]	*oxigénnek* 'oxigen' [oksigeːnːɛk]
[ɛ]			*novembernek* 'November' [novɛmbɛrnɛk]

Third, there are minor statistical patterns where backness harmony seems to be influenced by "unnatural" constraints independent of the height effect and the count effect. The following tendencies have been found by Hayes *et al.* (2009) to be statistically significant in the lexicon:[11]

(20) *Front harmonizing suffixes are preferred if the stem ends in a:*

a. bilabial non-continuant [p b m]
b. sibilant [s z ʃ ʒ ʧ ʤ]
c. coronal sonorant [n ɲ l r]
d. two-member consonant cluster

Ringen and Kontra (1989) found a preference for front harmonizing suffixes after BN stems which were longer than two syllables (but had just one N vowel) compared to two-syllable BN stems (e.g. *majonéz-Os* = *majonézes* [mɔjoneːzɛʃ] **majonézos* *[mɔjoneːzoʃ] 'with mayonnaise').

There seems to be a (potentially contrary) tendency for harmonizing suffixes to be back after BN stems in which the N vowel is stem-final: e.g. *parádé-bAn* = *parádéban* [pɔraːdeːbɔn] *?*parádében* [pɔraːdeːbɛn] 'in the parade'.

The height effect, the count effect, and these minor statistical effects combine to give the probability of the three types of suffix behavior after BN* stems. Consider, for instance, BN stems with stem-final [ɛ] (e.g. *dózse* 'doge'). In these stems the final [ɛ] lengthens to [eː] before most suffixes by Low Vowel Lengthening (LVL; see Siptár and Törkenczy 2000), a process that "upgrades" the vowel in the hierarchy of transparency, as [eː] is more transparent than [ɛ], thus *dózsé-nAk* = *dózsénak* ≈*dózsének* [doːʒeːnɔk ≈ doːʒeːnɛk] 'doge (DAT)', where the back-voweled suffix alternant is preferred, in accordance with Ringen and Kontra's findings. However, when we add the harmonizing suffix *-sÁg*, which does not trigger LVL, the vowel is not upgraded, and a back suffix alternant is less likely or impossible: *dózse-sÁg* = *dózseség* [doːʒɛʃeːg], *?*dózseság* *?[doːʒɛʃaːg] 'dogehood'.

Probability manifests itself in type frequency and native-speaker intuition. These were tested for some effects by Hayes and Londe (2006) and Hayes *et al.* (2009), using Google searches and a wug test, and by Ringen and Kontra (1989), using a sentence completion test with real BN stems.

Vacillation seems also to be sensitive to larger (phrasal) context (Kontra *et al.* 1989, 1991; Londe 2005). The choice between vacillating suffix alternants is

[11] These tendencies were found in all the zones of variation, i.e. in all-neutral stems as well.

of *Vígszínház* (Comic Theatre)), *Víg-bAn* = *Vígben* [vi:gbɛn], **Vígban* *[vi:gbɔn]. This also suggests that the antiharmonic pattern is not the regular/productive one. It is only roots that can be antiharmonic: there is no suffix that contains the relevant N vowels and, when added to an all-neutral non-antiharmonic (or Fr) root, would make the resulting stem antiharmonic.

(21) If in an all-neutral stem (N*) the closest vowel is [ɛ], the harmonizing suffix vowel will be *front*; if the closest vowel is [i: i e:], the harmonizing suffix vowel will be *front* after some stems and *back* after others.

Harmonic behavior is sometimes not only stem-specific, but also suffix-specific in the case of stems that show variation. The best-known example is *férfi* 'man' (see note 14), but there are others: e.g. *mágnes* [ma:gnɛʃ] 'magnet' vacillates in the dative *mágnesnek* ≈ *mágnesnak* [ma:gnɛʃnɛk ≈ ma:gnɛʃnɔk], but requires the front alternant of the adjective-forming suffix -(@)s: *mágneses* [ma:gnɛʃɛʃ], **mágnesos* *[ma:gnɛʃoʃ] 'magnetic', and *matek* [mɔtɛk] 'math' vacillates in the dative *mateknek* ≈ *mateknak*, [mɔtɛknɛk ≈ mɔtɛknɔk], but requires the back alternant of -(●)s *?*matekes* *?[mɔtɛkɛs], *matekos* [mɔtɛkos] if the meaning is 'person good at/interested/specializing in math'. The preference for one or another type of behavior may not be absolute. Very little is known about the suffix-specific aspect of Hungarian vowel harmony.

3.4.2 *Antiharmony*

Antiharmony occurs if an all-neutral stem takes the back alternant of a harmonizing suffix, a stem whose final vowel is Fr takes the back alternant of a harmonizing suffix, or a stem whose final vowel is B takes the front alternant of a harmonizing suffix. All three occur in Hungarian:

(i) *Antiharmonic roots*: this is the type of antiharmony we discussed in §3.4.1, which applies after a closed set of all-neutral roots.

(ii) *"Foreign" antiharmonic roots*: Some recent loans (from German and English) that are all neutral or whose final vowel is Fr take the back allomorph of the verb-forming suffix -(●)*l*, which otherwise harmonizes in the expected quaternary way (see (7), §3.5.1 and §4.3); cf. Kertész (2003). Compare (22a), where the suffix appears after "regular" Hungarian stems, with (22b), where it occurs after recent loans to form verbs:[15]

(22)	a.	*lábal*	[la:bɔl]	'wade'
		gombol	[gombol]	'button up'
		énekel	[e:nɛkɛl]	'sing'
		gőzöl	[gø:zøl]	'steam'
	b.	*(be)csekkol*	[ʧɛk:ol]	'check in'
		gründol	[gryndol]	'found (business)'
		risztol	[ristol]	'share (money)'
		dekkol	[dɛk:ol]	'wait (idly)'
		flémol	[fle:mol]	'flame' (in computing)

[15] Hungarian can borrow nouns, but not verbs, as bare roots. Verbs must be suffixed by verb-forming -(@)z or -(@)l: e.g. *sztrájk* [stra:jk] 'strike (N)', but *sztrájkol* [stra:jkol] 'go on strike'.

c.	*hekkel*	[hɛkːɛl]	'hack'
	printel	[printɛl]	'print'
	csetel	[ʧɛtɛl]	'chat (on the Internet)'
	netel	[nɛtɛl]	'use the Internet'

(22c) shows that recent verbal loans do not necessarily behave in an antiharmonic way. Kertész (2003) suggests that while it is not predictable which recent loan will behave antiharmonically with -(@)*l*, some regularities seem to obtain: foreign antiharmonic roots all end in a heavy syllable (in the stem containing the root and the suffix): ***dek**.kol*, ***flé**.mol* vs. ***cse**.tel*. Note that a heavy root-final syllable does not imply antiharmonic behavior: compare ***dek**.kol* and ***hek**.kel*. This can be seen as an "unnatural" constraint comparable to those in (20). This phenomenon is restricted to the verb-forming suffix -(@)*l*. The antiharmonic examples can be analyzed as containing a different, invariant -*ol* suffix (Siptár and Törkenczy 2000), in which case this type of antiharmony disappears. Then the problem becomes one of selection between the harmonizing verb-forming suffix -@*l* (22c) and the invariant -*ol* (22b) in recent loans.

(iii) *Systematic suffix-inconsistent antiharmony*: In Standard Hungarian the harmonizing conditional suffix -*nA*, which as a result of LVL becomes -*nÁ* before suffixes, appears in its *front* allomorph after stems with a final B vowel in the 1st singular indefinite present (Rebrus 2000b; Rebrus and Törkenczy 2005). Everywhere else the suffix harmonizes in a normal way. In the examples below I use the stems *vár* [vaːr] 'wait' and *tör* [tør] 'break' to show this.

(23) *Indefinite present conditional*

	vár		*tör*	
1SG	*vár-né-k*	[vaːrneːk]	*tör-né-k*	[tørneːk]
2SG	*vár-ná-l*	[vaːrnaːl]	*tör-né-l*	[tørneːl]
3SG	*vár-na*	[vaːrnɔ]	*tör-ne*	[tørnɛ]
1PL	*vár-ná-nk*	[vaːrnaːŋk]	*tör-né-nk*	[tørneːŋk]
2PL	*vár-ná-tok*	[vaːrnaːtɔk]	*tör-né-tek*	[tørneːtɛk]
3PL	*vár-ná-nak*	[vaːrnaːnɔk]	*tör-né-nek*	[tørneːnɛk]

The conditional suffix is not invariant, since it harmonizes in all other persons, and it is not the 1st singular morphosyntactic value that is associated with disharmony, since the same conditional suffix harmonizes in e.g. 1st singular definite present conditional: *várnám* [vaːrnaːm] vs. *törném* [tørneːm]. It is only before the 1st singular indefinite suffix -(@)*k* that the conditional suffix is invariably -*né*-. This inconsistent behavior with respect to backness harmony is exceptional: no other suffix behaves in this way. Otherwise, suffixes are stable in their classes: they are either always alternating (in the same way) or always invariant. In non-standard Hungarian and some regional dialects, this type of antiharmony does not occur and the conditional harmonizes in the usual way in the 1st singular indefinite present too: *várnák* [vaːrnaːk], *törnék* [tørneːk]. See §4.5 for further discussion of systematic suffix-inconsistent antiharmony.

3.4.3 -*ja/i* [jɔ/i]

The definiteness marker in the 3rd singular and 2nd and 3rd plural present indicative is peculiar because – although it alternates according to the harmonic type of the stem – the alternation is suppletive: the "front" alternant is *-i* [i] and the "back" alternant is *-ja/já*[16] [jɔ/jaː]: *töri* [tøri] *vs.* *várja* [vaːrjɔ] (3SG DEF PRES INDIC); *töritek* [tøritɛk] *vs.* *várjátok* [vaːrjaːtok] (2PL DEF PRES INDIC). Even though this alternation is suppletive, it does not violate backness harmony, since the vowel of *-ja/já* is back and *-i* is front. Interestingly, the *-i* alternant behaves in an opaque way: only a front suffix can follow it. This must be related to the fact that *-i* is the front alternant of a harmonizing suffix rather than an invariant one (compare invariant verb-forming *-ít* [iːt], which is transparent). Consider the following examples: *martini-z-i* [mɔrtinizi] and *martini-z-za*[17] [mɔrtinizːɔ] 'drink Martini (3SG DEF PRES INDIC)' are both possible due to the count effect. But when a harmonizing suffix e.g. *-tOk* (2PL) is attached to the first variant, only the front suffix alternant is possible *martini-z-i-tek* [mɔrtinizitɛk], **martinini-z-i-tok* *[mɔrtinizitok] 'drink Martini (3SG DEF PRES INDIC)' in spite of the fact that the count effect suggests both should be possible after a BN* stem.

However, vacillation *does* occur in similar examples with invariant *-ít*: consider *martini-s-ít-Om* ('treat as/turn something into Martini (1SG DEF PRES INDIC)'): *martini-s-ít-em* ≈ *martini-s-ít-om*, [mɔrtinisiːtɛm ≈ mɔrtinisiːtom] where both variants are possible. Furthermore, neutral-only invariant suffixes never change the harmonizing effect of the root and thus do not contribute to the count effect. Adding an invariant neutral-only suffix to a BN root or more than one to a B root never results in the count effect; the following harmonizing suffix will always be back without vacillation (as the root itself would require): *Martin-ék-hoz* [mɔrtineːkhoz], **Martin-ék-hez* *[mɔrtineːkhɛz] 'to the Martins'; *Márt-i-ék-hoz* [maːrtieːkhoz], *Márt-i-ék-hez *[maːrtieːkhɛz] 'to Martha (hypocoristic) and her group'. The vacillation in *martini-s-ít-em* ≈ *martini-s-ít-om* occurs because the root itself induces vacillation (as BNN) even without the invariant suffix: *martinitől* ≈ *martinitól* [mɔrtinitøːl ≈ mɔrtinitoːl]. This is a somewhat marginal phenomenon, because relevant examples are hard to come by, but it suggests that transparency is not only due to vowel quality or the rank of a vowel in the transparency hierarchy. It also shows that the association of transparency with invariance in suffixes manifests itself in very complex ways, which are unlikely to have a simple representational explanation.

3.5 Roundness harmony

Roundness harmony is relatively rarely discussed in the literature on Hungarian vowel harmony (see Polgárdi and Rebrus 1998; Ringen and Vago 1998; Rebrus 2000a; Siptár and Törkenczy 2000). Setting aside quaternary suffixes for the time being, roundness harmony can be stated as follows:

(24) If a ternary suffix is attached to a stem which is not antiharmonic and whose final vowel is front (N or Fr), the harmonizing suffix vowel will be front rounded if the final stem vowel is rounded and front unrounded if the final stem vowel is unrounded (e.g. *tökhöz, évhez*).

[16] The occurrence of [aː] in the alternant *-já* is due to LVL.

[17] The [j] of *-ja* assimilates to a preceding [z]; see Siptár and Törkenczy (2000).

The parasitic character of roundness harmony ((i) above) and its height sensitivity (ii) is typologically the most frequent state of affairs (Kaun 2004), although the restriction of *target* vowels to non-high while there are no height restrictions on the trigger seems atypical.

3.5.1 Quaternary and ternary suffixes: Lowering

Ternary suffixes, [o ~ ɛ ~ ø] *o* ~ *e* ~ *ö*, show the effect of backness harmony and roundness harmony in the way described above. Quaternary suffixes are different in that they show the usual ternary alternation after some stems and the *binary* alternation [ɔ ~ ɛ] *a* ~ *e* after others:

(25)	root			quaternary plural -@k suffix		ternary superessive -On suffix	
a.	*gáz*	[gaːz]	'gas'	*gázok*	[gaːzok]	*gázon*	[gaːzon]
	bot	[bot]	'stick'	*botok*	[botok]	*boton*	[boton]
	sün	[ʃyn]	'hedgehog'	*sünök*	[ʃynøk]	*sünön*	[ʃynøn]
	rög	[røg]	'clod'	*rögök*	[røgøk]	*rögön*	[røgøn]
b.	*gyep*	[ɟɛp]	'lawn'	*gyepek*	[ɟɛpɛk]	*gyepen*	[ɟɛpɛn]
	kép	[keːp]	'picture'	*képek*	[keːpɛk]	*képen*	[keːpɛn]
c.	*ház*	[haːz]	'house'	*házak*	[haːzɔk]	*házon*	[haːzon]
	fog	[fog]	'tooth'	*fogak*	[fogɔk]	*fogon*	[fogon]
	fül	[fyl]	'ear'	*fülek*	[fylɛk]	*fülön*	[fyløn]
	szög	[søg]	'nail'	*szögek*	[søgɛk]	*szögön*	[søgøn]

As can be seen in (25c), the vowel of the quaternary suffix shows up as [ɔ] instead of [o] after a final stem vowel and as [ɛ] instead of [ø] after a final front rounded stem vowel (i.e. roundness harmony does not apply). This is traditionally called Lowering, and the stems after which it happens are called lowering stems (25c). After non-lowering stems (25a) and stems whose final stem vowel is front unrounded, quaternary suffixes behave in the usual ternary way. In the standard dialect the lowering effect cannot be detected when the final stem vowel is front unrounded (25b), but in dialects that have a distinction between mid [e] and low [ɛ] it can, and there are other consequences of lowering in the standard dialect too (see Siptár and Törkenczy 2000).

All quaternary suffixes begin with a vowel that alternates with zero (a linking vowel), but not all suffixes with linking vowels are quaternary (see note 9).

It is a lexical property of a root whether it is lowering or not, but the frequency of lowering roots is different in different part-of-speech categories (verb roots are never lowering, most adjectives are lowering, and a closed set of about 200 nominal roots are lowering). In general, derivational suffixes do not turn a root or a stem into a lowering stem, but inflectional ones invariably do. Observe the behavior of the 1st singular definite suffix -(@)m: *üt-@m* = ***ütöm*** [ytøm] 'hit (1sg pres def)', *bánt-@m* = ***bántom*** [baːntom] 'hurt (1sg pres def)'; but *üt-@tt-@m* = ***ütöttem*** [ytøtːɛm] 'hit (1sg past def)', *bánt-@tt-@m* = ***bántottam*** [baːntotːɔm] 'hurt (1sg past def)'. The quaternary suffix lowers after the intervening past inflectional suffix -@tt: the roots are non-lowering, but the stems formed by the past suffix are (see Siptár and Törkenczy 2000).

4 Issues

Hungarian vowel harmony has been analyzed with binary features (both feature values active, or only one feature value active), unary features, or a mixture of binary and unary features. It has been given segmental and autosegmental treatments in derivational frameworks and in constraint-based ones, including Government Phonology and Optimality Theory (see further reading below). In this final section I briefly survey the most important or more interesting issues that an analysis of Hungarian vowel harmony must handle.

4.1 *Domains*

Hungarian vowel harmony applies within a domain that consists of a root plus suffixes, a domain which is often identified as the phonological/prosodic word. However, there are invariant suffixes that do not alternate. Most of these contain neutral (transparent) vowels, but there are a few with non-neutral vowels (see (10)). The former are usually analyzed as falling *within* the harmonic domain, and their invariance is seen as the consequence of their neutral character. The back invariant suffixes are either analyzed as falling outside the harmonic domain {*öt*}***kor*** [øtkor] 'at five' (where I use "{ }" to indicate the boundaries of the harmonic domain) or as being inside the harmonic domain {*öt**kor***}, but then the suffix vowel must be specified differently from the similar harmonizing ones (e.g. allative: -*hOz* = ***öthöz*** [øthøz] 'to five'). The first approach has been criticized on the grounds that if a harmonizing suffix follows the invariant back suffix, it shows up in its back alternant (Siptár and Törkenczy 2000). However, this hardly ever happens unambiguously, since the back invariant suffixes that *can* be followed by other suffixes are not (fully) productive, and as such they can be analyzed as part of the root (e.g. *cicus-hOz* = *cicushoz* [tsitsuʃhoz] 'to (a) kitten'). Also, the back invariant suffix could be analyzed as initiating its own harmonic domain in which the invariant suffix itself determines the harmonic properties of the following alternating suffix: {*cic*} {*ushoz*}. It can be seen as a drawback of the domain approach that it increases the proliferation of domains required by various processes whose domain boundaries do not line up; see Rebrus *et al.* (1996). Table 123.1 shows five suffixes (3rd plural possessive -*Unk*, -*Ul* 'in (a language)', -*ig* 'up to', -*sÁg* 'ness', -*ként* 'in the capacity of') and whether or not they obey backness harmony (VH), Low Vowel Lengthening (LVL), and the ban on adjacent identical vowels (${}^{*}V_aV_a$). For comparison I have included a pre-verb *le-* [lɛ] 'down' (since pre-verbs are known to be outside the harmonic domain).

As can be seen, various (types of) affixes delimit their own domains differently (there would be more types if more processes were included) and that invariant suffixes do not necessarily behave in the same way as prefixes (which are affixes that are not in the same harmonic domain as the stem).

4.2 *Harmony in roots*

Harmony obviously applies to suffixes as can be seen in the suffix alternations. There is sometimes disagreement about whether it applies in roots since there is no alternation evidence but only restrictions on co-occurrence (Kiparsky 1968;

Table 123.1 Domain mismatch

	VH	*LVL*	$*V_aV_a$
-Unk	+ *törökünk* 'our Turk' [tørøkyŋk] *dánunk* 'our Dane' [daːnuŋk]	+ *bembánk* 'our Bemba' [bɛmbaːŋk]	+ *bantunk* 'our Bantu' [bɔntuŋk]
-Ul	+ *törökül* 'in Turkish' [tørøkyl] *dánul* 'in Danish' [daːnul]	+ *bembául* 'in Bemba' [bɛmbaːul]	– *bantuul* 'in Bantu' [bɔntuul]
-ig	– *törökig* 'up to (a) Turk' [tørøkig] *dánig* 'up to (a) Dane' [daːnig]	+ *bembáig* 'up to (a) Bemba' [bɛmbaːig]	– *kocsiig* 'up to (a) car' [kotʃiig]
-sÁg	+ *törökség* 'Turkishness' [tørøkʃeːg] *dánság* 'Danishness' [daːnʃaːg]	– *bembaság* 'Bemba-ness' [bɛmbɔʃaːg]	X
-ként	– *törökként* 'as a Turk' [tørøkːeːnt] *dánként* 'as a Dane' [daːŋkeːnt]	– *bembaként* 'as a Bemba' [bɛmbɔkeːnt]	X
le-	– *ledob* 'throw down' [lɛdob]	– *ledob* 'throw down' [lɛdob]	– *leesik* 'fall down' [lɛɛsik]

Vago 1974; van der Hulst 1985). Disharmonic roots do exist, but they are clearly exceptional (see §3.2). While the answer is ultimately theory dependent and the motivation underlying the distinction between handling root-internal static regularities with Morpheme Structure Constraints (MSCs; see CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS) and suffix alternations with a rule of harmony has disappeared with the introduction of autosegmental representations and/or underspecification, it is notable that in OT treatments there is always a highly ranked IDENT-IO constraint relativized to the root that expresses the non-application of Hungarian vowel harmony in roots (see Ringen and Vago 1998; Hayes and Londe 2006). There are two points of interest:

(i) Vowel harmony seems to actively apply within a class of roots, called "epenthetic,"[20] which display vowel–zero alternations in their final syllable (see

[20] "Epenthetic" is only a convenient traditional label here; arguably, the vowel–zero alternation is neither epenthesis nor deletion; see Siptár and Törkenczy (2000). This has no bearing on the issue at hand.

§3.2 above). These roots all end in a VCVC string before consonant-initial suffixes and word-finally, but a VCC string before vowel-initial suffixes. Crucially, the vowel that alternates with zero behaves like the vowel of a ternary suffix [o ~ ɛ ~ ø]: its backness and rounding is determined by the preceding vowel:[21] *bokor* [bokor] 'bush', *bokorban* [bokorban] 'in (a) bush', *bokrok* [bokrok] 'bushes'; *eper* [ɛpɛr] 'strawberry', *eperben* [ɛpɛrbɛn] 'in (a) strawberry', *eprek* [ɛprɛk] 'strawberries'; *tükör* [tykør] 'mirror', *tükörben* [tykørbɛn] 'in (a) mirror', *tükrök* [tykrøk] 'mirrors'.

(ii) Root-internal and suffix harmony seem to be asymmetrical in that (a) there are no front rounded invariable suffixes, i.e. exceptions to vowel harmony that would be the equivalent of the *sofőr* type of disharmonic roots (4a.ii); and (b) unlike root + suffix combinations (e.g. *dzsungel-nÁl* = *dzsungelnál* ≈ *dzsungelnél* [ʤuŋgɛlnaːl ≈ ʤuŋgɛlneːl] 'near the jungle'), roots never vacillate between BN*[back] and BN*[front]: *mohamedán* ≈ **mohamedén* [mohɔmɛdaːn ≈ *mohɔmɛdeːn] 'Muslim'; *planéta* ≈ **planéte* [plɔneːtɔ ≈ *plɔneːtɛ] 'planet', *vanília* ≈ **vaníliе* [vɔniliɔ ≈ *vɔniliɛ] 'vanilla'. If vowel harmony is permitted to apply in roots, (i) is naturally explained, but (ii) will have to be seen as accidental unless some general principle (like Strict Cyclicity[22]; see also CHAPTER 85: CYCLICITY) can be used to distinguish the two cases.

4.3 *Harmony in suffixes*

All analyses have to address the question of how to distinguish suffixes that alternate in different ways (§3.3.1). This can be done in a fairly straightforward way with reference to vowel height. Some analyses assume that the alternating suffix vowels are fully specified and that the underlying identity of a suffix vowel is revealed when the suffix is used as a stem in pronominal forms (e.g. Vago 1980c; Ringen and Vago 1998): e.g. delative *-rÓl* in *botról/tökről* [botroːl/tøkrøːl] 'about (a) stick/marrow' has underlying /oː/ because of *rólam* [roːlɔm] 'about me', but ablative *-tÓl* in *bottól/töktől* [botːoːl/tøktøːl] 'from (a) stick/marrow' has underlying /øː/ because of *tőlem* [tøːlɛm] 'about me'. It is not clear why a phonological connection should hold between the suffixes and the corresponding pronominal forms: consider superessive *-On* in *boton/tökön* [boton/tøkøn] 'on (a) stick/marrow' and *rajtam* [rɔjtɔm] 'on me', where the same connection obviously does not obtain.

Perhaps the most vexing problem (other than the representation of neutral vowels; see §4.4) is how to distinguish ternary alternations from quaternary ones (§3.3.1). Recall that the difference is that quaternary suffixes show the effect of Lowering, but ternary ones do not. The analytic problem is that mid and low vowels (specifically the low vowel [ɛ]) are involved in both so it is not possible to distinguish them on the basis of surface height. [ɛ] is two-faced: it patterns with the *low vowels* in length alternations (e.g. LVL) and in binary suffixes like *-bAn* 'in', but with the *mid vowels* in ternary alternations. This is usually analyzed by

[21] The details are more complicated: there are a handful of exceptional "epenthetic" stems whose alternating vowel is not [o ~ ɛ ~ ø], and there are some vowel-initial suffixes before which the alternating vowel is retained; for details see Siptár and Törkenczy (2000).

[22] Presumably, Strict Cyclicity can only block vowel harmony within the root if it is a structure-changing operation (cf. Kiparsky 1982), thus only if root vowels are fully specified for the harmonizing feature.

assuming that the vowel of ternary alternations (and the alternating vowel of "epenthetic" roots) including *e* is underlyingly mid, and that it becomes low as a result of phonetic interpretation. By contrast, the vowel involved in quaternary alternations is specified in such a way that it becomes low by Lowering after lowering stems. Thus the reason why roundness harmony does not apply after lowering stems is that the resulting vowel would be a *low* front rounded vowel, which does not exist in Hungarian (see Siptár and Törkenczy 2000). This kind of analysis is problematic for surface-oriented frameworks unless some special provisions are made to handle opacity.

4.4 Transparency, neutral vowels, and variation

The behavior of neutral vowels is perhaps the main reason why Hungarian vowel harmony has received so much attention in phonology. Virtually all available theoretical approaches and analytic tools have been applied to Hungarian neutral vowels. In these analyses the fact that (some) neutral vowels are transparent is either representationally or procedurally expressed, i.e. it is either coded in the way these vowels are represented (e.g. Ringen 1988a, 1988b) or the way these vowels are "calculated" by the rules/constraints (e.g. Ringen and Vago 1998; Siptár and Törkenczy 2000). If the representational analysis is chosen, the vowel harmony rule(s)/constraint(s) can be formulated in a general way; if the procedural analysis is adopted, no special representations have to be assumed for transparent vowels. Note that the behavior of BN* stems, specifically the count effect, is a serious problem for the representational view if the transparency of vowels is analyzed as a consequence of the fact that they are not specified for the harmonizing feature. Given the usual autosegmental assumptions, the B trigger and the target of harmony are adjacent no matter how many transparent vowels intervene: there is no reason why behavior should change depending on the number of vowels, but it does.

Until very recently, variation, vacillation, and gradience in Hungarian vowel harmony (§3.4.1) – although they were noted and sometimes dealt with in the analyses – were seen as less important than the "overall" pattern of Hungarian vowel harmony, which was considered to be essentially categorical. Accordingly, a great deal of attention was paid to the exact identity of neutral/transparent vowels, specifically to whether [ɛ] is transparent (e.g. Siptár and Törkenczy 2000) or opaque (e.g. Ringen and Kontra 1989). Even if gradience in transparency was sometimes recognized, it was not built into the models (derivational, OT, etc.), which were categorical by nature. Accordingly, variation/vacillation was handled representationally, by setting up more than one underlying form for the vacillators or procedurally, by optional rules or variable constraint ranking (see Siptár and Törkenczy 2000). Hayes and Londe (2006) propose an analysis in which gradience is directly built into the model, in the form of stochastic OT, where the constraints referring to the various neutral vowels can be independently and stochastically ordered in the hierarchy of constraints. In such an analysis the categorization problem concerning [ɛ] disappears.

Some recent experimental studies (Benus *et al.* 2003; Benus and Gafos 2007) suggest that transparent vowels also participate in backness harmony *phonetically*, contrary to what all other analyses assumed. They claim that transparent vowels in the front-selecting stems (e.g. *bili-vAl* = *bilivel* [bilivɛl] 'with the chamber pot') are produced with more advanced tongue body than the corresponding vowels that occur in back-selecting stems (e.g. *buli-vAl* = *bulival* [bulivɔl] 'with the party'). Crucially, it is also claimed that the relative retraction of the [i] in *bulival* is *not* a co-articulation effect due to the neighboring back vowels, since it also shows up in *isolated* forms of antiharmonic roots like *híd* [hi:d] 'bridge' and *cél* [ʧe:l] 'aim' (as opposed to *víz* [vi:z] 'water' and *szél* [se:l] 'wind'). Beňuš and Gafos (2007) also connect the differential phonological behavior of [ɛ] compared to [i: i e:], with their different articulatory and acoustic properties, a difference which is completely arbitrary or a lexical effect in most other analyses. (For a discussion of transparent vowels see CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS.)

4.5 *Antiharmony*

The problem with antiharmonic roots (§3.4.1) is how to distinguish them from non-antiharmonic ones. They are typically assumed to have some kind of arbitrary lexical marking (in the form of abstract vowels, floating features expressing "backness" or explicitly arbitrary diacritics/full listing; see e.g. Vago 1980c; Siptár and Törkenczy 2000; Hayes and Londe 2006). Beňuš and Gafos (2007) suggest that antiharmonic roots are phonetically distinguishable from harmonic ones, thus their phonological patterning is phonetically motivated (see the discussion above). Hayes *et al.* argue that there are non-categorical biases for antiharmony based on the statistics of the lexicon, which are phonologically "unnatural" but nevertheless are accessed by native speakers.

The systematic suffix-inconsistent antiharmony of conditional *-nÁ* (§3.4.2.iii) is usually either not discussed in the literature or it is treated as entirely arbitrary (Vago 1976, 1980c). Rebrus and Törkenczy (2005) argue that the antiharmony of the conditional in the 1st singular present indefinite after back roots (e.g. *tudnék* [tudne:k] 'I would know (INDEF)') is a paradigm contrast effect. The "regular" form with the back allomorph of the conditional suffix **tudnák* *[tudna:k] 'I would know (INDEF)' would result in homophony with the present 3rd plural definite conditional form *tudnák* [tudna:k] 'they would know (DEF)'. What is interesting is that Hungarian vowel harmony can be shown to be active in spite of antiharmony. There *is* homophony between the present 1st singular indefinite conditional forms and the present 3rd plural definite conditional forms of front stems (*ölnék* [ølne:k] 'I would kill (INDEF)'), but this cannot be avoided through antiharmony, since the hypothetical antiharmonic form **ölnák* *[ølna:k] would violate Hungarian vowel harmony more than *tudnék* [tudne:k], where the suffix has a neutral vowel.

4.6 *Backness harmony* vs. *roundness harmony*

The central problem about roundness harmony is its restricted character (§3.5). This is rarely addressed in the literature, but some authors argue that it is not harmony at all, but a form of (local) licensing of the feature expressing roundness in suffixes and that accounts for some of its properties, specifically, its strictly local nature and its dependence on backness harmony (§3.5); see Polgárdi and Rebrus (1998) and Ringen and Vago (1998). The restriction of targets to non-high vowels makes Hungarian roundness harmony an interesting problem for phonetically based theories that make the prediction on perceptual grounds that the preferred targets of rounding harmony are high vowels (Kaun 2004).

5 Conclusion

The complexities of Hungarian vowel harmony – especially the long-distance effects involving neutral vowels, the numerous classes of harmonic and invariant suffixes, and the class of antiharmonic roots – made Hungarian vowel harmony an especially interesting phenomenon, a "hot" topic when the nature of representations was the central issue in phonological theory. Now that the focus of phonological research has shifted to other areas, such as variation, gradience, phonetic motivation, the role of the statistics of the lexicon, the role of paradigms in phonological phenomena, and the analysis of opacity in surface-oriented constraint-based theories (such as OT), there is every reason for Hungarian vowel harmony to receive just as much attention, since some aspects of it are crucially relevant to these new issues, specifically the gradient transparency of neutral vowels as manifested in the statistics of variation (§4.4), the phonetics of neutral vowels and the connection between surface vowel quality and transparency (§4.4), the role of the paradigm in antiharmony (§4.5), the statistics of the lexicon and the antiharmonic behavior of some all-neutral roots (§4.5), the restrictions on roundness harmony (§4.6), and the opacity of the distinction between ternary and quaternary alternations, both involving [ɛ] (§4.3). Some of these issues have recently been addressed in the literature (Rebrus 2000b; Beňuš *et al.* 2003; Rebrus and Törkenczy 2005; Hayes and Londe 2006; Beňuš and Gafos 2007; Wanlass 2008; Hayes *et al.* 2009), and it is clear that the intricacies of Hungarian vowel harmony will continue to contribute to the development of phonological theory.

ACKNOWLEDGMENTS

I would like to thank Péter Rebrus and Péter Siptár for their comments on various issues discussed here. I am grateful to Keren Rice, Paul de Lacy, Marc van Oostendorp, and an anonymous reviewer for their comments on an earlier version of this chapter.

REFERENCES

Abondolo, Daniel. 1988. *Hungarian inflectional morphology.* Budapest: Akadémiai Kiadó.

Beňuš, Stefan & Adamantios I. Gafos. 2005. Qualitative and quantitative aspects of vowel harmony: A dynamics model. In Bruno G. Bara, Lawrence Barsalou & Monica Bucciarelli (eds.) *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, 226–231. Mahwah, NJ: Lawrence Erlbaum.

Beňuš, Stefan & Adamantios I. Gafos. 2007. Articulatory characteristics of Hungarian "transparent" vowels. *Journal of Phonetics* 35. 271–300.

Beňuš, Stefan, Adamantios I. Gafos & Louis Goldstein. 2003. Phonetics and phonology of transparent vowels in Hungarian. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 29. 485–497.

Booij, Geert. 1984. Neutral vowels and the autosegmental analysis of Hungarian vowel harmony. *Linguistics* 22. 629–641.

Clements, G. N. 1976. Neutral vowels in Hungarian vowel harmony: An autosegmental interpretation. *Papers from the Annual Meeting of the North East Linguistic Society* 7. 49–64.

Hayes, Bruce & Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104.

Ringen, Catherine & Miklós Kontra. 1989. Hungarian neutral vowels. *Lingua* 78. 181–191.
Ringen, Catherine & Szilárd Szentgyörgyi. 2000. Constraint reranking in the Szeged dialect of Hungarian. In Gábor Álberti & István Kenesei (eds.) *Approaches to Hungarian: Papers from the Pécs Conference*, vol. 7, 5–21. Szeged: Attila József University.
Ringen, Catherine & Robert M. Vago. 1998. Hungarian vowel harmony in Optimality Theory. *Phonology* 15. 393–416.
Siptár, Péter & Miklós Törkenczy. 2000. *The phonology of Hungarian.* Oxford: Oxford University Press.
Vago, Robert M. 1974. Hungarian generative phonology. Ph.D. dissertation, Harvard University.
Vago, Robert M. 1976. Theoretical implications of Hungarian vowel harmony. *Linguistic Inquiry* 7. 243–263.
Vago, Robert M. 1978. Some controversial questions concerning the description of Hungarian vowel harmony. *Linguistic Inquiry* 9. 116–126.
Vago, Robert M. 1980a. A critique of suprasegmental theories of vowel harmony. In Vago (1980b), 155–183.
Vago, Robert M. (ed.) 1980b. *Issues in vowel harmony.* Amsterdam: John Benjamins.
Vago, Robert M. 1980c. *The sound pattern of Hungarian.* Washington: Georgetown University Press.
Wanlass, Zara. 2008. Evidence from online corpora for variation in Hungarian vowel harmony. In Christopher Piñón & Szilárd Szentgyörgyi (eds.) *Approaches to Hungarian: Papers from the Veszprém Conference*, vol. 10, 233–245. Budapest: Akadémiai Kiadó.

FURTHER READING

Austerlitz, Robert Paul. 1950. Phonemic analysis of Hungarian. M.A. thesis, Columbia University, New York.
Battistella, Edwin. 1982. More on Hungarian vowel harmony. *Linguistic Analysis* 9. 95–118.
Becker Makkai, Valerie. 1970. Vowel harmony in Hungarian reexamined in the light of recent developments in phonological theory. In Valerie Becker Makkai (ed.) *Phonological theory: Evolution and current practice*, 634–648. New York: Holt, Rinehart & Winston.
Dienes, Péter. 1997. Hungarian neutral vowels. *The Odd Yearbook 1997*. 151–180.
Esztergár, Marianne. 1971. A generative phonology of nouns and vowel harmony in Hungarian. Ph.D. dissertation, University of California, San Diego.
Ewen, Colin J. & Harry van der Hulst. 1985. Single-valued features and the non-linear analysis of vowel harmony. In Hans Bennis & Frits Beukema (eds.) *Linguistics in the Netherlands 1985*, 39–48. Dordrecht: Foris.
Farkas, Donca & Patrice Speeter Beddor. 1987. Privative and equipollent backness in Hungarian. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 91–105.
Goldsmith, John A. 1985. Vowel harmony in Khalkha Mongolian, Yaka, Finnish and Hungarian. *Phonology Yearbook* 2. 253–275.
Gósy, Mária. 1989. Vowel harmony: Interrelations of speech production, speech perception, and the phonological rules. *Acta Linguistica Hungarica* 39. 93–118.
Hare, Mary. 1990. The role of similarity in Hungarian vowel harmony: A connectionist account. *Connection Science* 2. 123–150.
Hetzron, Robert. 1972. Studies in Hungarian morphophonology. *Ural-Altaische Jahrbücher* 44. 79–106.
Jensen, John T. 1971. Theoretical implications of Hungarian vowel harmony. Paper presented at the 2nd Annual Meeting of the North East Linguistic Society.
Jensen, John T. 1984. A Lexical Phonology treatment of Hungarian vowel harmony. *Linguistic Analysis* 14. 231–253.

124 Word Stress in Arabic

JANET C. E. WATSON

1 Introduction

Within phonological theory, Arabic word stress has received arguably more attention than the word stress of any language other than English; and within Arabic linguistics, more work has been devoted to stress-related phenomena than any other single topic. This chapter addresses some of the reasons behind this phenomenon.

The chapter is structured as follows: §2 considers the characteristics of Arabic word stress, discussing general features common to different varieties and basic ways in which modern Arabic dialects differ both from Classical Arabic and from each other. §3 provides a historical overview of the treatment of Arabic word stress within generative paradigms, focusing on major contributions in the analysis of Arabic stress, and cases where data from Arabic has contributed to the development of stress theory. §4 considers stress in Classical Arabic, and then examines in more detail word stress in three modern dialects – Cairene, San'ani (Yemen), and Levantine – focusing particularly on phenomena that pose a challenge for metrical phonology.

2 Characteristics of Arabic word stress

With over 250 million speakers, Arabic is the official language of 18 sovereign states from Mauritania in the west to Iraq in the east. It is also spoken in parts of southern Turkey, by Maronite Christians in northern Cyprus, and in parts of sub-Saharan Africa. Arabic language enclaves are found in the Balkh region of Afghanistan, parts of Iran, and Uzbekistan. All Arabic dialects exhibit word stress; however, the socially and geographically diverse area over which Arabic is spoken leads to differences in the mechanics of word stress assignment. In all cases stress location is a function of both syllable weight and syllable position, but dialects differ in the distribution of syllable types, the leftmost extent of stress (third or fourth syllable from the right), the rhythmic grouping of syllables, the interaction of stress, syncope and epenthesis, and the degree to which lexical information may affect stress.

Arabic recognizes three weights of syllable: light, heavy, and superheavy. Light syllables are always open, heavy syllables are open or closed, and superheavy syllables are closed or doubly closed. Examples of these syllable types from Classical Arabic are given below:

(1)

	open		*closed*		*doubly closed*	
light	CV	wa 'and'				
heavy	CVV	sā.fara 'he traveled'	CVC	min 'from' ka.tab.tu 'I wrote'		
super-heavy			CVVC	bāb# 'door'	CVCC	bint# 'girl'
					CVVGG[1]	mādd# 'stretching'
			CVVG	mād.dun 'stretching (NOM)'		

CV and CVC are unrestricted, although unstressed short vowels in open syllables are often deleted in modern dialects. CVCC and CVVGG are restricted to word- or utterance-final position.[2] In the distribution of other syllable types, however, dialects vary. Levantine, Sudanese, some Peninsular, and North African dialects allow CVGG and/or CVVC in derived environments word internally, as in: /māsik-īn/ > [māskīn] 'holding (MASC PL)', /ʃāf-ha/ > [ʃāfha] 'he saw her', /muʕallim-īn/ > [mʕallmīn] 'teachers'. Cairene allows CVVC syllables word finally only, as in: [kitāb kibīr] 'a big book', but /kitāb-na/ > [kitabna] 'our book'; Cairene restricts CVCC to utterance-final position, breaking up word-final non-utterance-final CVCC syllables through epenthesis, as in: /bint ṭawīla/ > [bint[i] ṭawīla] 'a tall girl'; CVV occurs only when stressed in Cairene: initial CVV in ['ʃāfit] 'she saw' and ['ʕālam] 'world' contrast with initial CV in [ʃa'fitu] 'she saw him' and [ʕa'lamu] 'his world'.

Stress falls on one of the last three syllables, in some dialects one of the last four syllables, with assignment dependent on the weight and position of the stressed syllable. Modern dialects follow the assumed rules of Classical Arabic (§3.1) whereby stress is assigned to a final superheavy (CVVC, CVCC, or CVVGG) syllable, as in Cairene: [fi'luus] 'money', [ma-xa'bazʃ] 'he didn't cook', and Palestinian: [ja'waab] 'answer', [bi-'ḥuṭṭ] 'he puts'. In the absence of a final superheavy, stress is assigned to a heavy penult (CVV or CVC), as in Cairene: [ka'tabtu] 'you (PL) wrote', [fih'muuha] 'they understood her', and Palestinian: [mus'taʃfa] 'hospital', [mu'naafis] 'competitor'. In the absence of either a final superheavy or a heavy penult, the dialects differ. In words with a heavy antepenult, Cairene stresses the light penult, while most other dialects stress the antepenult: Cairene [mad'rasa] 'school' contrasts with Beirut/Damascene ['madrase].

Modern Arabic dialects differ in their rhythmic grouping of light syllables. Western and Bedouin-type dialects tend to group light syllables into weak–strong

[1] GG denotes *geminate*.

[2] In Classical Arabic, superheavy syllables occur pre-pausally only, resulting from pre-pausal deletion of short final vowels or case endings.

pairs (iambs): Cyrenaican Bedouin stresses the penult in forms such as: [kitab-at] > [ik'tibat] 'she wrote', [ingital-aw] > [inig'tilaw] 'they were killed', and the final syllable in [ki'tab] 'he wrote', [nu'xal] 'palm-trees' (Mitchell 1960); eastern urban dialects group light syllables into strong–weak pairs (trochees), stressing the antepenult in forms such as Cairene: ['katabit] 'she wrote', [in'kasarit] 'it (FEM) broke', and the penult in ['katab] 'he wrote', ['walad] 'boy' (SEE CHAPTER 44: THE IAMBIC–TROCHAIC LAW).

All modern dialects differ from Classical Arabic in at least optionally deleting short vowels in unstressed open syllables (Birkeland 1954). Some dialects delete short high vowels only, as in Damascene: /fihim-u/ > ['fihmu] 'they understood', /fihim-na/ > [f'himna] 'we understood' but /katab-u/ > ['katabu] 'they wrote' (Cowell 1964). Other dialects delete short vowels irrespective of their quality, as in Lebanese Kfar-Ṣġāb: /ḍarab-ak/ > ['ḍarbak] 'he hit you' and /samak-i/ > ['samki] 'one fish' (Fleisch 1974).

Several dialects differ from the assumed predictable quantity-based system of Classical Arabic in that certain morphemes affect stress placement. In Cairene and Tunisian the 3rd feminine singular perfect inflectional suffix *-it* attracts stress on suffixation: Cairene ['ramit] 'she threw' becomes [ra'mitu] 'she threw it (MASC)', contrasting with other CVCVCV forms, such as 'katabu 'they wrote', where the antepenult is stressed. In Iraqi, the dual suffix *-een* retracts stress, although all other cases of word-final CVVC attract stress: ['ʧalbeen] 'two dogs' contrasts with [taʕ'baan] 'tired' (Erwin 1963: 43). In Muslim Mosul Iraqi, stress always falls on the final syllable of a verbal or nominal stem when it takes a suffix, as in: [nəx'ləṭ-u] 'we mix it (MASC)' (Jastrow 2007). In some western dialects and dialects of Oman, word stress is phonemic in disyllabic noun–verb pairs: initial stress in ['fihim] 'understanding' contrasts with final stress in [fi'him] 'he understood' (Janssens 1972).

Finally, modern dialects differ as to whether or not epenthetic vowels count for stress purposes. In Cairene, a penultimate post-CVC syllable with an epenthetic vowel is stressed like any other penultimate post-CVC syllable: compare [bin'tina] 'our daughter' with [mad'rasa] 'school' and [fih'mitu] 'she understood him'. In Iraqi and Levantine, by contrast, stress is assigned as if the epenthetic vowel were not there: penultimate stress in Muslim Mosul [ka'tabit] 'I wrote' contrasts with initial stress in ['katabət] 'she wrote' (Jastrow 2007).

3 Theoretical accounts of Arabic word stress

This section provides a historical overview of theoretical accounts of Arabic word stress, focusing on ways in which research on Arabic has contributed both to the development of metrical theory and to a deeper understanding of Arabic prosodic structure and cross-dialectal differences.

3.1 *Pre-generative approaches*

Concepts upon which generative studies of Arabic word stress draw have their roots in early pre-generative approaches. The older traditional studies of Erpenius (1656), Brockelmann (1908), and Wright (1971) recognized the role of the syllable and syllable weight in stress assignment, distinguishing between light (CV) and heavy (CVV and CVC) syllables (see CHAPTER 57: QUANTITY-SENSITIVITY). The analyses of stress in Cairene by Harrell (1957) and Cairene and Cyrenaican Bedouin by Mitchell (1956, 1960) are based on the position and relative weight of syllables.

The Prague School (Jakobson 1971) describes stress assignment not in terms of syllables, but in terms of moras. Moraic accounts of stress in Arabic include Cantineau (1960: 240), for whom stress in El-Hamme of Gabes (Tunisia) is placed: "on the third mora of the word . . . on the fourth if the third corresponds to a consonant," accounting for penultimate stress in: [k'tibtu] 'you (PL) wrote', and antepenultimate stress in: ['madrasa] 'school'. The mora is referred to in the informal expression of Abdo's (1969) post-*SPE* (Chomsky and Halle 1968) account, where Classical Arabic stress is assigned to the vowel preceding the last two moras (the third or fourth mora from the right-edge). Within metrical theory, it later returns as a full-fledged element of the representation, firmly embedded within the prosodic hierarchy.

3.2 *Generative approaches*

Generative approaches to stress in Arabic have followed contemporary approaches in generative phonology, with a few landmark changes in orientation. In the *SPE* segment-based approach adopted by Abdo (1969), Brame (1970, 1973, 1974), Broselow (1976), Johnson (1979), and Weldon (1980), stress is encoded as a phonological distinctive feature, [+stress], assigned to a [+syllabic] segment in a particular segmental context. Essential variables are included in the vocabulary of phonological rules. Take Palestinian as an example, for which the basic stress rules are:

(2) a. Stress a final superheavy syllable: [ba-'ʃūf] 'I see', [bi-'ḥuṭṭ] 'he puts'.
 b. Otherwise stress the rightmost non-final heavy syllable: [ba-'ʃūfiʃ] 'I don't see', [ka'tabti] 'you (FEM SG) wrote', ['miktaʃif] 'discovering'.
 c. Otherwise stress the first syllable (up to the antepenult): ['katab] 'he wrote', ['zalama] 'man'.

Under this approach, stress is assigned by the following rule (Brame 1974), where C_0 indicates an arbitrary number of consonants, including zero, and C_0^1 either zero or one consonant.

(3) *Stress assignment*
 $V \rightarrow [+\text{stress}] \; / \; \_\_ \; C_0((VC)VC_0^1)]$

This rule abbreviates three disjunctively ordered sub-rules:

(4) $V \rightarrow [+\text{stress}] \; / \; \_\_ \; C_0VCVC_0^1]$ e.g. ['zalama], ['miktaʃif]
 $V \rightarrow [+\text{stress}] \; / \; \_\_ \; C_0VC_0^1]$ e.g. [ka'tabti], ['katab]
 $V \rightarrow [+\text{stress}] \; / \; \_\_ \; C_0]$ e.g. ['ḥaṭṭ] 'he put', [ba-'ʃūf]

3.2.1 *The interaction of morphology and word stress*

Brame noted that the basic stress rules and their formalism in (2) fail to account for data such as [ka'tabit] 'I/you (MASC SG) wrote' (cf. ['katabit] 'she wrote'),

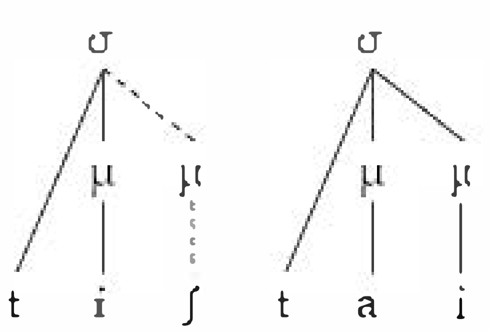
[ˈʔabilna] 'before us' (cf. [ʔˈbilna] 'we accepted'). These forms are derived from underlying /katab-t/ and /ʔabl-na/ through epenthesis (see CHAPTER 67: VOWEL EPENTHESIS), and are explained by Brame as the assignment of stress prior to epenthesis, giving intermediate [kaˈtab-t] and [ˈʔabl-na]. Epenthesis does not undo stress assignment, leading to opaque assignment of stress to a light penult in [kaˈtabit], and opaque non-assignment of stress to the heavy penult in [ˈʔabilna] (§3.2.3). These cases of opaque assignment, or lack of assignment, of stress were attributed initially by Brame (1970, 1973, 1974) and later by others (e.g. Kenstowicz and Abdul-Karim 1980; Kiparsky 1982, 2000, 2003) to the cycle and the preservation of metrical structure assigned in earlier cycles ((54); CHAPTER 85: CYCLICITY). The *SPE*-type approach to word assignment has since been superseded, but recognition of the role of the cycle and of the interaction of syncope and epenthesis with word stress assignment has not. As we shall see below (§4.2.3), within the stratal version of Optimality Theory (OT), opaque stress is attributed to inter-level constraint masking: if α is the constraint system of domain Y (e.g. stem), and β is the constraint system of a larger domain Z, then β's markedness constraints can render α opaque (Kiparsky 2000). Thus, opacity in dialects such as Levantine is attributed to word-level assignment of stress and postlexical epenthesis, which renders stress opaque.

3.2.2 *The prosodic hierarchy and representation of the syllable*

The post-*SPE* period formed an asyllabic interlude in the analysis of Arabic word stress. Most pre-generative and non-generative accounts made reference to the syllable, and within generative phonology there was increasing recognition that sounds grouped into syllables of differing prosodic weights, and that the syllable formed part of the prosodic hierarchy (Fudge 1969; Kiparsky 1979; McCarthy 1979; Selkirk 1980, 1982; Halle and Vergnaud 1987).

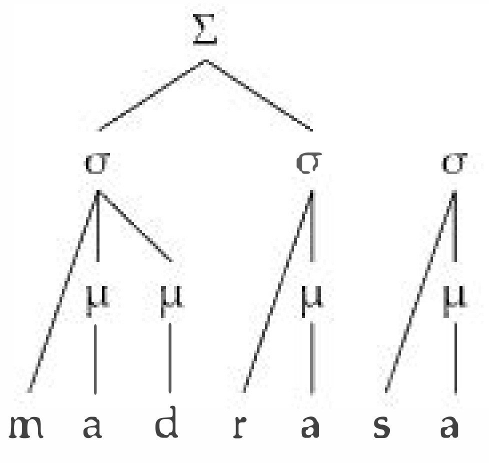
At this time it came to be recognized that the syllable formed a unit within a prosodic hierarchy, a hierarchy that recognized units of prosodic structure above the syllable – the foot (see CHAPTER 40: THE FOOT) and the prosodic word (see CHAPTER 51: THE PHONOLOGICAL WORD) – and a unit of prosodic structure below the syllable: the mora. Weight-based, rather than segment-based, models of the syllable representing the prosodic tier as a series of moras (e.g. McCarthy 1980; Angoujard 1990; McCarthy and Prince 1990) provide a model that reflects the role of prosodic weight in stress assignment by accounting for phonological positions, and by distinguishing between light (monomoraic) and heavy (bimoraic) syllables (Hayes 1989): short vowels contribute one mora, long vowels two moras, geminate consonants one mora, and coda consonants are assigned a mora through Weight-by-Position in languages such as Arabic, where CVC syllables count as heavy (quantity-sensitivity). In San'ani [tiʃtai] 'you (FEM SG) want', for example, the vowels contribute moras and /ʃ/ receives a mora through Weight-by-Position (indicated by a dashed line):

(5)



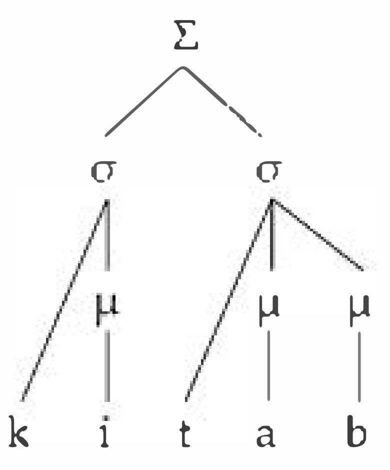
from the stressed syllable to the right-edge of the word, placing a word like ['madrase] 'school' within a single foot. McCarthy (1979, 1980) restricted the foot by measuring it in moras: in Damascene, the stressed mora plus, at most, two following moras. A word comprising three light syllables, such as ['darasu] 'they learnt', exhausts the foot, giving ['(darasu)], but in a word comprising a heavy plus two light syllables, such as ['madrase] 'school', the final light syllable is excluded from the foot: ['(madra)se].

(8)



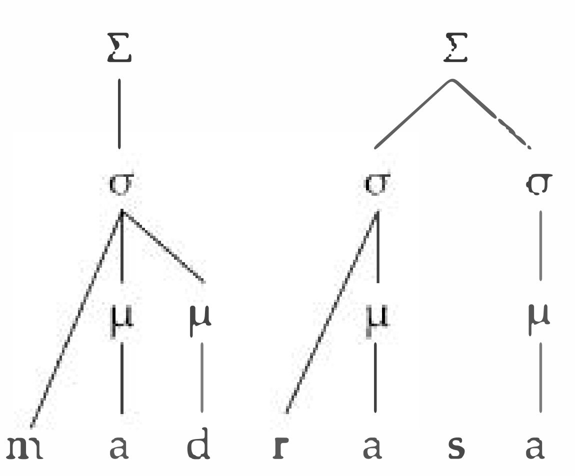
Since these accounts, bounded foot inventories have often excluded feet that require counting higher than two (but cf. e.g. Burzio 1994). Hayes (1989, 1995) argues for absolute binarity: the maximal and canonical iamb consists of a light syllable followed by a heavy syllable, as in Cyrenaican Bedouin [ki'tab] 'he wrote':

(9)



Trochees comprise two equal elements: syllables in the syllabic trochee, moras in the moraic trochee. In a moraic trochee dialect, [madrasa] 'school' comprises two moraic trochees; cf. (8):

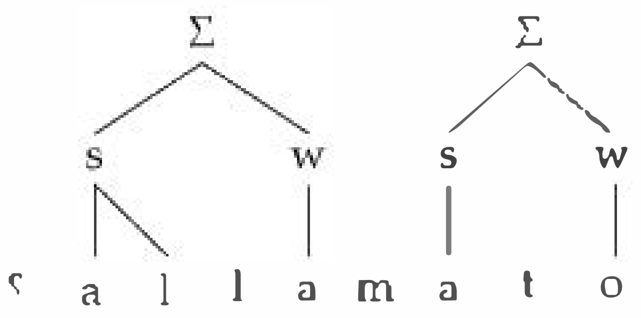
(10)



The uneven trochee, which comprises a heavy and a light syllable in moraic trochee systems, is ruled out by Hayes. It is, however, invoked by Irshied and Kenstowicz (1984), Angoujard (1990), and Kager (2009) to account for penultimate stress in

HLLL forms in trochaic Arabic dialects, as in Jordanian Bani-Hassan [ʕalla'mato] 'she taught him':

(11)



3.2.5 Constituency and the metrical grid

Metrical trees reflect constituency through sister nodes, but fail to represent in any transparent way typical stress characteristics of stress clash or alternating rhythm between strong and weak syllables. Prince (1983) and Selkirk (1984) argued that the metrical grid could better capture the rhythmic characteristics of stress, and that constituency into feet could be eliminated. Compare the pure grid representation of Cairene [mux'talifa] 'different (FEM SG)' below with the metrical tree representation in (6):

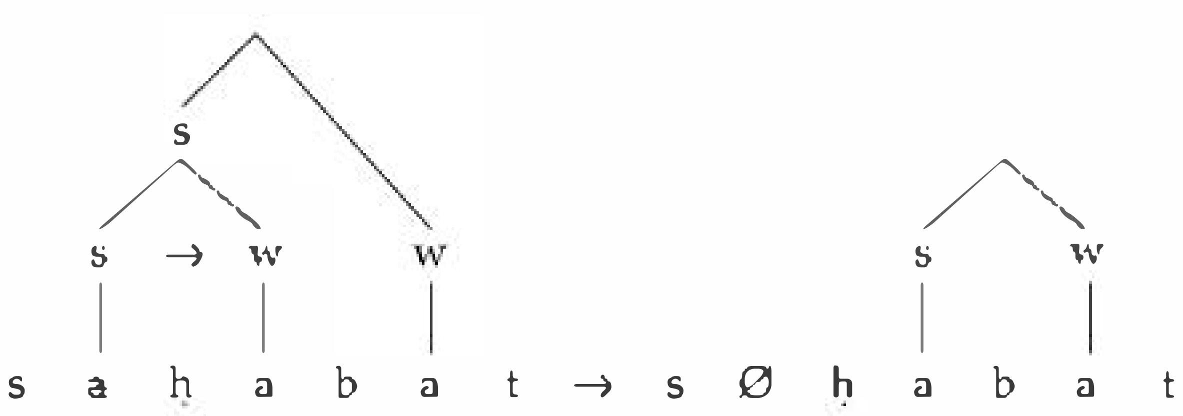
(12)

In 1985, data from an Arabic dialect appeared to challenge the effectiveness of the pure grid. In an account of Bedouin Hijazi Arabic (BHA), Al-Mozainy *et al.* (1985) analyzed alternations such as those in (13) as resulting from low vowel deletion and stress shift.

(13)	'saḥab	'he pulled'	s'ḥabat	'she pulled'
	'naxal	'palm trees'	n'xalah	'a palm tree'
	'salag	'hunting dogs'	s'ligah	'a hunting dog'

They argued the direction of shift was governed by constituent structure and that vowel deletion in BHA induces left-to-right stress shift to the sister node within the metrical tree:

(14)



Through eliminating constituency, the pure grid provided no explanatory account for the direction of stress shift in data such as these. The introduction of brackets within the grid (Halle and Vergnaud 1987; Halle and Kenstowicz 1991;

Hayes 1995), however, enabled the grid to handle such data, showing that on vowel deletion stress shifts within the bracketed constituent (in this case to the right):

```
(15) a.  (*       )      b.   (     *       )    word
         (*    .)             (     *)           foot
          *    *                    *            syllable
        s  a  ḥ  a  b        s  Ø  ḥ  a  b  a  t
```

Further work on BHA showed the initial findings to have been based on an incorrect analysis of the foot-type: essential paradigms had not been considered (Angoujard 1992; Paoli 2008), and BHA feet were re-analyzed as iambs rather than moraic trochees (McCarthy 2003); motivated by the tendency of iambic feet to maximize quantitative differences between the head and dependent syllables, deletion targets the unstressed vowel of an iambic foot. The basic conclusions that stress is preserved on deletion of stressed vowels and that the direction of stress shift is predictable from constituency remain valid, however, and are supported by data from other Arabic dialects: Cyrenaican Bedouin (Hayes 1995, data from Mitchell 1960) and San'ani (Watson 2002: 116) both exhibit stressed vowel deletion and stress shift[4] within the foot: leftwards in Cyrenaican, which exhibits iambic stress, and rightwards in San'ani, which exhibits trochaic stress.

```
(16) a. Cyrenaican [ki'tib] 'books' > ['kitbih] 'his books'
        (        *  )    (  *                )    word
          (.     *)      (  *          )          foot
           *     *          *                *    syllable
        k  i  t  i  b  >  k  i  t  Ø  b  i  h

     b. San'ani ['xaʃab] 'wood' > [x'ʃabih] 'a piece of wood'
        (  *        )    (        *          )    word
          (*     .)         (     *    )          foot
           *     *                *          *    syllable
        x  a  ʃ  a  b  >  x  Ø ʃ  a  b   i  h
```

3.2.6 *Unparsed syllables*

The findings of many researchers suggest metrical structure need not exhaust the string of syllables. Hayes (1995) claims syllables left over after foot construction are either unfooted or form degenerate feet. Degenerate feet are not optimal, but languages differ as to whether they impose a strong or a weak ban on them. Languages that permit words smaller than the canonical foot (a single mora in a moraic trochee system, or a single syllable in an iambic or syllabic trochee system) tend to invoke the weak ban, while languages that do not permit sub-minimal words invoke the absolute ban.

Sub-minimal content words are generally unattested in Arabic and sub-minimal grammatical words are expanded on suffixation to provide a minimal bimoraic base, as in Cairene:

[4] Optional in the case of San'ani: ['xaʃabih] ~ [x'ʃabih].

(17)	/min-u/	>	[minnu]	'from him/it (MASC)'
	/ʕan-kum/	>	[ʕannukum]	'about you (PL)'
	/fi-h/	>	[fī(h)]	'in it, there is'

Sub-minimal loanwords are typically expanded through vowel lengthening to match the minimal prosodic word, as in: [bār] 'bar' and [bāṣ] 'bus'. Some dialects, such as San'ani, however, do have stressable monomoraic content words, including [ab] 'father', [ax] 'brother', [yad] 'hand', [dam] 'blood', and a few sub-minimal function words that contrast with comparable bimoraic words in Cairene, and never lengthen, including [kam] 'how many' (cf. Cairene [kām]), [man] 'who' (cf. Cairene [mīn]), and [maʕ] 'with' (cf. Cairene [maʕa]) (Watson 2002: 88–89). Further evidence that degenerate feet are allowed in strong position in San'ani includes the exceptional stressing of peripheral light syllables, giving optional initial stress in forms such as ['tamām] 'good', and optional final stress in forms such as [ʃik'mih] 'party for parturient' (§4.2.2).

3.2.7 *Weak parsing*

Many languages stress the third syllable from the word edge. Hayes (1982, 1989, 1995) argues such systems can be accounted for not by expanding the universal inventory to include ternary feet, but by resorting to the independently motivated devices of extrametricality at the edge, destressing in clash, and the non-exhaustivity of foot construction. Non-exhaustivity of foot construction means syllables can be skipped through a device known as weak local parsing, potentially creating ternary alternation in longer strings (Hayes 1995: 308):

```
(18)  (x  .)    (x  .)    (x  .)
       ˘  ˘  ˘  ˘  ˘  ˘  ˘  ˘  ˘
```

This device enables Hayes to provide an account of Arabic dialects which dispenses with the uneven trochee. Bani-Hassan [ʕalla'mato], seen above in (11), is analyzed in moraic trochees with weak local parsing (Hayes 1995: 366):

```
(19)    (x)          (x     .)
         ¯       ˘     ˘     ˘
      ʕ  a  l  l  a  m  a  t  o
```

3.2.8 *Final consonants, syllables, and feet*

The right-edge of the word prompts exception in many languages: extra-long syllables are often restricted to the right-edge, and syllables that act as heavy non-finally often fail to attract stress in final position. In Cairene, the sequence [.tab.] is stressed penultimately in [ka.'tab.tu] 'you (PL) wrote', but not finally in ['ka.tab] 'he wrote'. To account for the asymmetric behavior of closed syllables and the invisibility of peripheral elements to stress rules, Liberman and Prince (1977) introduced the notion of extrametricality (see CHAPTER 43: EXTRAMETRICALITY AND NON-FINALITY), the rules for which were developed by Hayes (1979, 1982, 1989). Thus, rather than specify for relevant languages that CVC is light finally, but heavy non-finally, and that only CVCC and CVVC syllables are heavy finally, the rightmost consonant is analyzed as invisible to stress rules through extrametricality, making final CVC equivalent in weight to non-final CV (Hayes 1995: 57):

(20)	*Final*	*Non-final*
	CV	CV
	CV<C>	CV
	CVC<C>	CVC
	CVV	CVV
	CVV<C>	CVV

In several Arabic dialects, final pairs of light syllables also appear to be invisible to stress rules. In Palestinian, for example, stress is assigned to the rightmost bimoraic sequence in these words:[5]

(21)	'ʔana	'I'
	'katabu	'they wrote'
	bara'kitna	'our cow'
	ka'tabna	'we wrote'

However, in words comprising a heavy syllable followed by two light syllables, or four light syllables, stress is assigned to the initial syllable:

(22)	'barakito > 'bakarto	'his cow' (with High Vowel Deletion)
	'madrasa	'school'
	'ʃajaratun (Classical)	'tree (NOM)'

Hayes (1995) analyses such patterns as resulting from foot extrametricality, subject to the non-exhaustivity condition. Thus, ['madrasa] is parsed as [(mad)(rasa)], with two bimoraic feet; by not exhausting the word the peripheral foot is eligible for extrametricality, giving [(mad)<(rasa)>]. Stress is assigned to the rightmost visible (non-extrametrical) foot: ['(mad)<(rasa)>]. The characteristic Cairene pattern of stressing a light penult after a heavy antepenult, by contrast, is analyzed as resulting from lack of foot extrametricality: the rightmost visible foot is stressed in all dialects, but only in Cairene is the peripheral foot visible to stress rules: Cairene [(mad)'(rasa)] contrasts with Palestinian ['(mad)<(rasa)>].

3.2.9 CVXC syllables

Stress patterns in several languages, including Arabic, indicate that an extrametrical consonant does not deprive the rightmost foot of peripherality: by being contained within the peripheral foot, the peripheral consonant does not intervene between the foot and the right-edge. In San'ani ['katabatih] 'she wrote it (MASC)', for example, extrametrical /h/ falls within the rightmost foot, itself deemed extrametrical. Stress assignment to the rightmost visible foot gives ['(kata)<(bati<h>)>]. An analysis of extrametricality in the case of CVVC and CVCC strings, however, predicts the wrong results: extrametricality would render final C invisible to stress rules; as an extrametrical element it would fall in the adjacent foot, failing to deprive the foot of peripherality; as peripheral feet, CVCC and CVVC syllables would be invisible to stress rules in dialects such as Palestinian for which foot extrametricality holds, but not in dialects where foot extrametricality fails to apply. Such an analysis would predict a stress difference in words of the pattern CVVCVCC

[5] Data from Brame (1973, 1974), Abu-Salim (1980), and Kenstowicz and Abdul-Karim (1980).

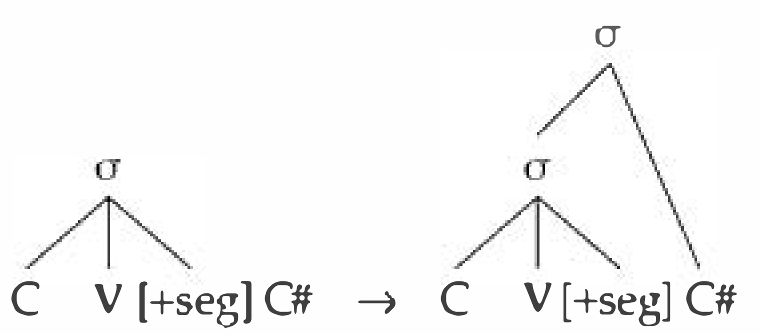
and CVCCVCC between dialects that exhibit foot extrametricality (23a), and dialects such as Cairene, which do not (23b):

(23) a. Consonant extrametricality: darrast → darras<t>
Foot extrametricality: (dar)(ras<t>) → (dar)<(ras<t>)>
Stress rightmost visible foot: *'(dar)<(ras<t>)> (= dar'rast)

b. Consonant extrametricality: darrast → darras<t>
Foot extrametricality: *n/a*
Stress rightmost visible foot: (dar)'(rast)

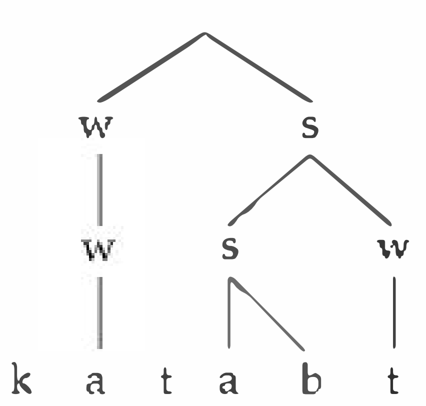
A significant step in research on Arabic stress concerns the analysis of these so-called "superheavy" syllables. Superheavy syllables are exceptional on two counts: they are the only syllable types that are always stressed in final position (although cf. §4.2.2), and they are restricted to domain-final position (at least in morphologically simple forms). Thus, in terms of stress rules, they behave like penultimate CVV or CVC (as do CVVC/CVCC syllables in English; Burzio 1994; Harris and Gussmann 2002). Several analyses of CVVC/CVCC strings have been proposed around the basic analysis of heavy syllable + element. This was expressed initially by McCarthy (1979: 453) through Chomsky-adjoining a word-final consonant to a heavy syllable:

(24)



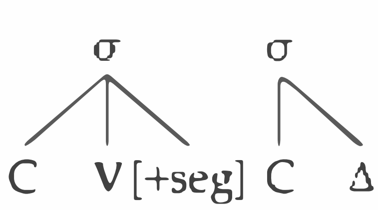
Halle and Vergnaud (1979) analyze final C in CVXC as the weak element in a branching foot:

(25)



In later accounts (Aoun 1979; Angoujard 1981, 1990; Selkirk 1981;[6] Burzio 1994), the rightmost C forms a degenerate syllable, i.e. a syllable with an empty nucleus or null vowel:

(26)



[6] Only for CVCC syllables. See §4.2.2.

For Hayes (1995: 126), final C in CVXC is not syllabified at the initial stage: CVXC is analyzed as heavy syllable + stray consonant, while CVC is analyzed as light syllable + extrametrical consonant:

(27) CVX.C
 CV<C>

Within OT, Al-Mohanna (2004) analyses final C of word-final CVXC as attached directly to the prosodic word node. Essentially, then, all accounts show that the rightmost C in CVXC sequences intervenes between the right edge and the heavy syllable, depriving the foot formed by the heavy syllable of peripherality. Thus, *darrast* (23) in both Cairene- and Palestinian-type dialects is analyzed as [(dar)'(ras).t].

This section has examined key approaches to Arabic stress within the generative paradigm, with particular focus on challenges raised by Arabic data for stress theory. §4 will provide sketches of the stress systems of Classical Arabic and three modern Arabic dialects, focusing on core similarities and differences between the dialects, and examining approaches invoked to account for cross-dialectal differences and apparent exceptions to the stress algorithms.

4 Stress algorithms

4.1 *Classical Arabic*

The early Arab grammarians provided detailed descriptions of segments and melodic phonological processes characteristic of readings of the Qur'an and certain Peninsula dialects; however, word stress is never mentioned. This led some researchers to believe that Classical Arabic exhibited no word stress (Birkeland 1954; Ferguson 1956; Garbell 1958), and others to assume it to have been similar to the rather fluctuating word-stress system found today in western dialects of the Arabian Peninsula.

The Classical Arabic stress patterns have since been reconstructed through comparison of modern dialect stress patterns (Janssens 1972), versification (Weil 1954; Wright 1971), and observation of the non-dialectal pronunciation of Classical Arabic in some regions (Abu-Fadl 1961; Mitchell 1993). There is now general consensus that Classical Arabic exhibited predictable stress. Disagreement exists, however, as to the leftmost limit of stress. It is agreed that penultimate CVC or CVV bore stress, or, if the penult was light, antepenultimate CVC or CVV. Where both the penult and antepenult were light, as in /masʔalat-un/, researchers differ: Erpenius (1656), Abdo (1969), Brame (1970), Bohas and Kouloughli (1981), and Angoujard (1990) argue that stress did not retract beyond the antepenultimate syllable, giving, in this case, [mas'ʔalat-un] 'problem (NOM)'. Brockelmann (1908), Wright (1971), Janssens (1972), and McCarthy (1979), by contrast, claim stress is assigned to the initial syllable in such cases, giving ['masʔalatun]. If this latter holds, this would mean Classical Arabic, in contrast to the modern dialects, exhibited unbounded metrical feet, constructing feet from one heavy syllable up to, but not including, the next heavy syllable. Under both analyses, lexical exceptions exist: dialect comparison and the non-dialectal pronunciation of Classical Arabic suggest stress was not assigned to the initial heavy syllable in derived verb forms VII and VIII, but to the light antepenult, giving: [in'faʕala] and [if'taʕala] rather than *['infaʕala] and *['iftaʕala].[7] Similarly, most particle prefixes are unstressed irrespective of their relative position (Mitchell 1960: 371, 1975: 77), leading to antepenultimate or penultimate stress in forms such as: [al-'walad-u] 'the boy (NOM)' (cf. Wright's ['madrasat-un] 'school (NOM)'), [al-'yad-u] 'the hand (NOM)', (cf. ['maktab-un] 'office (NOM)'), [wa-'yad-un] 'and a hand (NOM)' (cf. ['katabat] 'she wrote').

The stress algorithm for Classical Arabic is given in (28) (bracketed elements included in case stress retraction is limited to the antepenult):

(28) *Classical Arabic stress*

a. Stress a pre-pausal superheavy (CVVC, CVVCC, or CVCC) syllable: [ki'tāb] 'book', ['mādd] 'stretching (MASC SG)', [ʃa'ribt] 'I/you (MASC SG) drank'.

b. Otherwise, stress the rightmost non-final heavy (CVV, CVC, or CVVG) syllable (up to the antepenult): [da'rasnā] 'we learnt', [ṣā'būnun] 'soap (NOM)', ['maktabah] 'library', ['māddun] 'stretching (NOM)', ['maktabatun] 'library' (non-pause) (or [mak'tabatun]).

c. Otherwise, stress the leftmost CV syllable (or antepenult): ['kataba] 'he wrote', ['katabatuhu] 'library' (or [kata'batuhu]).

4.2 Arabic dialects

This section presents the basic stress algorithms for Cairene, San'ani, and Levantine, three dialects analyzed as exhibiting moraic trochaic stress. Each subsection considers some of the most significant data that has impacted on metrical theory and approaches invoked to handle this data. The section is concluded by a table summarizing the main stress and stress-related typological characteristics of each dialect and of Cyrenaican Bedouin, aspects of which we considered in §3.

4.2.1 Cairene

More generative accounts of word stress have been provided for Cairene than any other Arabic dialect. Cairene attracted attention due to its characteristic avoidance of a heavy antepenult in favor of a light penult, deletion of unstressed high vowels but, with few exceptions (Woidich 2006), not unstressed low vowels, reduction of unstressed long vowels, and its exceptions.[8] An initial stress algorithm for Cairene was provided by Harrell (1957, cf. also Mitchell 1956):

(29) *Cairene stress*

a. Stress-final CVV(C) or CVCC: [ka'tabt] 'I wrote', [ʔa'bū(h)] 'his father', [saka'kīn] 'knives', [ṭala'bāt] 'demands'.

b. Otherwise, stress the antepenult when the penult and antepenult are light, unless the pre-antepenult is light: ['ʔabadan] 'never', [mux'talifa] 'different (FEM SG)'. Cf. [kata'bitu] 'she wrote it (MASC)' with pre-antepenultimate CV.

c. Otherwise, stress the penult: [yik'tibu] 'they write', [ʕa'malti] 'you (FEM SG) did', [mar'taba] 'mattress', ['bētak] 'your (MASC SG) house'.

[7] An exception for Wright (1971), etc., but not for Angoujard (1990).

[8] Descriptive accounts of Cairene in the theoretical literature include Mitchell (1952, 1956, 1960, 1975), Harrell (1957, 1960), Tomiche (1964), Behnstedt and Woidich (1985), and Woidich (2006).

In /ʕālam-u/ > [ʕa'lamu], unstressed CVV is reduced to CV. In /ʔābil-u/ > ['ʔablu], the high vowel in the weak position of the foot is deleted, resulting in a CVVC syllable, impermissible in Cairene. The long vowel is subject to Closed Syllable Shortening (CSS) to prevent word-internal CVVC, giving ['ʔablu] 'they met'. Watson (2002), who follows Hayes (1995) in disallowing the uneven trochee, analyses forms such as /sāfir-it/ > ['safrit] as assignment of metrical structure prior to syncope, with re-assignment of metrical structure after the application of each phonological rule:

(37) a. Construction of moraic trochees from left to right: (sā)(firit).
b. Assignment of stress to the head of the rightmost foot: (sā'(firit)).
c. Reduction of unstressed CVV to CV: (sa'(firit)).
d. Refooting: ('(safi)rit).
e. Syncope of the high vowel: '(safrit).

Both these approaches deal with this data, but they miss the generalization that, with the exception of CiCiCa/CuCuCa plurals and the *-it* morpheme, all short high vowels in the position CVCVCV(C) are subject to syncope, even if they would be stressed by the normal stress algorithm (Kenstowicz 1980; Teeple 2009), namely: /kanakit-u/ > [ka'naktu] 'his coffeepot', /kanabit-u/ > [ka'nabtu] 'his sofa'. This data suggests word-internal short high vowels are subject to syncope prior to assignment of metrical structure, as long as the resulting syllable is permissible (cf. Broselow 1992: 36–37): /kanakit-u/ gives [kanaktu], but /mudarris-a/ fails to give *[mudarrsa] because word-internal [.darr.] (CVGG) is impermissible in Cairene.

Syncope is not restricted to the phonological word in Cairene; it also occurs within the phonological phrase: high vowels in word-initial CV syllables are subject to phrasal syncope after a word-final vowel, as in:

(38)	/ʔana fihimt/	>	'ʔana f'himt	'I understood'
	/ṭardi kibīr/	>	'ṭardi k'bīr	'my parcel is big'

Deletion fails to occur in (39), however, even though word-initial CV follows word-final CV:

(39)	/huwwa fihim/	>	'huwwa 'fihim	(*fhim)	'he understood'
	/huwwa wiḥiʃ/	>	'huwwa 'wiḥiʃ	(*w'ḥiʃ)	'he is bad'

In (38), the high vowel falls in an unstressed syllable in the citation form ([fi'himt] 'I understood', [ki'bīr] 'big'). By contrast, in (39), the high vowel falls in a stressed syllable in the citation form (['fihim] 'he understood', ['wiḥiʃ] 'bad'). Taken with data such as [ka'naktu] 'his coffeepot', this suggests two types of syncope occur in this dialect: lexical syncope, which targets word-internal CV-flanked high vowels prior to the assignment of metrical structure, and phrasal syncope, which targets unstressed word-initial CV-flanked high vowels after the assignment of metrical structure.

4.2.2 *San'ani*

The main interest in San'ani, the dialect of the old city of San'a, Yemen, lies in its stressing of peripheral light syllables, and the patterning of CVV syllables with syllables ending in the left leg of a geminate (CVG), but not with CVC syllables.[9]

The basic stress algorithm for Classical Arabic (28) applies to most word-types:

(40) a. Stress a final CVCC or CVVC syllable: [gi'rīt] 'I/you (MASC SG) read/learnt', [gam'bart] 'I/you (MASC SG) sat', [ki'tāb] 'book'.

b. Otherwise, stress the rightmost non-final heavy syllable (CVV or CVC) up to the antepenult: ['madrasih] 'school', [miǧ'sālih] 'launderette', [da'rastih] 'I/you (MASC SG) recited it'.

c. Otherwise, stress the leftmost CV syllable: ['libisat] 'she put on, wore', [mak'tabatih] 'his library', ['katabatih] 'she wrote it (MASC)'.

(40) fails to apply in San'ani, however, when the penult or antepenult is CVV or CVG. Here the rightmost non-final CVV or CVG syllable attracts stress from final CVCC or CVVC (Watson 2002: 81):

(41)	'ṣābūn	'soap'
	'xuṭṭāf	'clasp'
	mu'darrisīn	'teachers (MASC)'
	'xārijīn	'going out (MASC PL)'

Similarly, while the rightmost non-final CVC syllable is stressed iff in penultimate or antepenultimate position, CVV and CVG are stressable in pre-antepenultimate position:

(42)	mak'tabatī	'my library'
	mu'sajjilatī	'my recorder'
	'hākaḏahā	'like this'

(40) often fails to apply in post-pausal position: San'ani exhibits contextually fluctuating stress (Rossi 1939; Naïm-Sanbar 1994); post-pausally the initial syllable is usually stressed, irrespective of its weight or that of following syllables. Where initial CV is stressed before CVXC, this is analyzed in Watson (2002) as a degenerate syllable (§3.2.6):

(43)	'tamām	'okay'
	'katabt	'I/you (MASC SG) wrote'
	'ʔusbūʕ	'week'
	'baladīyāt	'municipality'

The following revised algorithm accounts for the data:

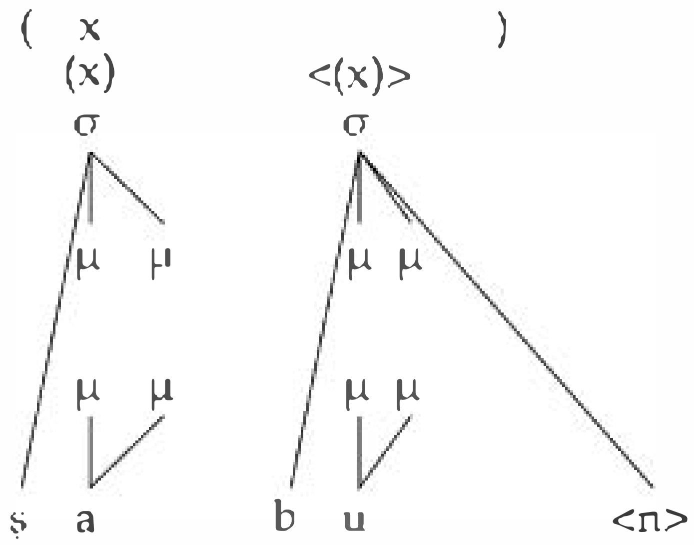
(44) *San'ani stress*

a. Stress the rightmost non-final CVV or CVG syllable: [ma'kātib] 'offices', [ba'sātīn] 'gardens', ['xārijīn] 'going out (MASC PL)', [mit'ʔaxxirāt] 'late (FEM PL)', ['hākaḏahā] 'like this', ['sāfart] 'I/you (MASC SG) traveled', ['dawwart] 'I/you (MASC SG) looked for'.

[9] Data cited in the theoretical literature include Rossi (1939), Goitein (1970), Naïm-Sanbar (1994), and Watson (2002).

'soap', two bimoraic feet are constructed on the upper moraic layer; the peripheral foot is extrametrical, and stress is assigned to the rightmost visible foot:

(49)



word layer construction (ERR)
foot extrametricality
consonant extrametricality

4.2.3 *Levantine*

Levantine has attracted a great deal of attention within generative phonology, due principally to its complex interactions between the morphology and stress assignment.[12] The dialects basically exhibit the Latin stress rule (McCarthy 1980: 79), with the exception of (50a), which Latin lacks:

(50) *Levantine stress*

a. Stress final CVCC or CVVC: [bi-t-'ḥuṭṭ] 'she/you (MASC SG) put', [ja'wāb] 'answer'.
b. Otherwise, stress a heavy penult: [ka'tabti] 'you (FEM SG) wrote', ['bārak] 'he blessed'.
c. In disyllables ending in CV or CVC, stress the initial syllable: ['ḍarab] 'he hit', ['bana] 'he built', ['ʔana] 'I'.
d. Otherwise, stress the antepenult: ['darasu] 'they learnt', [mut'taḥida] 'united', ['ʕallamat] 'she taught', ['madrasa] 'school'.

Hayes (1995: 128) accounts for Palestinian stress as follows:

(51)	a.	Consonant extrametricality	C → <C> /__]word
	b.	Foot construction	Form moraic trochees from left to right. Degenerate feet are forbidden absolutely.
	c.	Foot extrametricality	Foot → <Foot> /__]word
	d.	Word layer construction	ERR

These rules generate the following metrical structure:

```
(52)      (*                    )       word layer construction: ERR
          (*)    <(*         .)>        foot construction: moraic trochees, L > R,
                                        foot extrametricality
           μ  μ     μ       μ
       m   a  k  t  a   b   a   <k>     consonant extrametricality
```

[12] The many traditional descriptive-analytical accounts of Levantine include Feghali (1919), Bauer (1926), Cantineau (1939, 1946, 1960), El-Hajjé (1954), Cowell (1964), Jiha (1964), Grotzfeld (1965), Blanc (1970), Fleisch (1974), and Naim-Sanbar (1986).

Levantine-like opacity has continued to attract considerable interest within constraint-based models. Optimality-theoretic approaches towards opacity include: invoking constraints on the stressing of epenthetic vowels, maximizing paradigmatic contrasts,[13] and translating the notion of the cycle into lexical and postlexical strata. Within parallel OT, Kager (1999) accounts for Levantine opacity by invoking the constraint Head-Dep(O/I) ('every vowel in the output prosodic head has its correspondent in the input'), which prohibits stress on epenthetic vowels. By dominating constraints responsible for stress, Head-Dep(O/I) rules out penultimate stressed *[fi'himna] (input ['fihm-na]) and *[yi'kitbu] (input ['yiktib-u]), and other constraints select ['fihimna] and ['yikitbu]. Kiparsky, however, raises two objections to this constraint: first, its only remit is to prevent epenthetic vowels from being stressed, but epenthetic vowels are not simply unstressable, they are *invisible* to stress: words of the form CVCVCV(C) receive antepenultimate stress unless the final vowel is epenthetic, in which case the penult is stressed, acting as if the epenthetic vowel were not there, as in: [ka'tabit] 'I/you (masc sg) wrote'. Second, Head-Dep(O/I) fails to relate the opacity of stress to other word-level prosodic processes, thus missing the generalization "that *all* processes of word phonology ignore epenthetic vowels" (Kiparsky 2000: 353). For example, word-level closed CVVC syllables are shortened even though they are opened by postlexical epenthesis:

(55) /ʃāf-it/ > ʃāfit 'she saw'
/ʃāf-t/ > 'ʃifit (*ʃāfit) 'I saw'

Brame's insight (§3.2.1) that syncope is ordered before epenthesis in dialects such as Levantine is captured in stratal OT by allowing different constraint rankings in the lexical and postlexical strata (Kiparsky 2000, 2003). The relevant constraints here are the faithfulness constraint Max-'V, requiring the stressed vowel of the input to have a correspondent in the output, and the markedness constraints Reduce, which minimizes the number of non-final light syllables, and License-μ, which requires all moras to be licensed by syllables. Syncope takes place at the word level because Reduce outranks License-μ. At the post-lexical level, epenthesis is prompted by the promotion of License-μ, and Max-'V rules out candidates in which stress is assigned to the epenthetic vowel. Kiparsky's stratal OT analysis of /yiktib-u/ > ['yikitbu] is given in (56) and (57):

(56) VC *dialects: Word level*

Input: ['(yik).(ti.bu)]	Reduce	License-μ	…
☞ a. '(yik).tμ.bu	*	*	
b. '(yik).(ti.bu)	**!		

[13] Broselow (2008), for example, argues that the invisibility of epenthetic vowels in Iraqi Arabic is motivated by maximization of contrast between stems of different grammatical types.

Angoujard, Jean-Pierre. 1990. *Metrical structure of Arabic.* Dordrecht: Foris.
Angoujard, Jean-Pierre. 1992. Syncope et voyelles accentuées (parler du Hedjaz, Bani Hassan et hébreu biblique). *Languages Orientales Anciennes, Philologie et Linguistique* 4. 299–304.
Aoun, Youssef. 1979. Is the syllable or the supersyllable a constituent? *MIT Working Papers in Linguistics* 1. 140–148.
Bauer, Leonhard. 1926. *Das palästinische Arabisch.* Leipzig: J. C. Hinrichs'sche Buchhandlung.
Behnstedt, Peter & Manfred Woidich. 1985. *Die ägyptisch-arabischen Dialekte.* Wiesbaden: Ludwig Riechert.
Birkeland, Harris. 1954. *Stress patterns in Arabic.* Oslo: Jacob Dybwad.
Blanc, Haim. 1970. The Arabic dialect of the Negev Bedouins. *Proceedings of the Israeli Academy of Science and Humanities* 4. 112–150.
Bohas, Georges. 1975. Métrique arabe classique et moderne. Thèse de 3ème cycle, Université Paris VIII.
Bohas, Georges & Djamel Kouloughli. 1981. Processus accentuels en arabe (parlers du Caire, de Damas et arabe classique). *Analyses, Théorie* 1. 36–46.
Brame, Michael K. 1970. Arabic phonology: Implications for phonological theory and historical Semitic. Ph.D. dissertation, MIT.
Brame, Michael K. 1973. On stress assignment in two Arabic dialects. In Stephen R. Anderson & Paul Kiparsky (eds.) *A Festschrift for Morris Halle*, 14–25. New York: Holt, Rinehart & Winston.
Brame, Michael K. 1974. The cycle in phonology: Stress in Palestinian, Maltese and Spanish. *Linguistic Inquiry* 5. 39–60.
Brockelmann, Carl. 1908. *Grundriss der vergleichenden Grammatik der semitischen Sprachen*, vol. 1. Hildesheim: Georg Olm.
Broselow, Ellen. 1976. The phonology of Egyptian Arabic. Ph.D. dissertation, University of Massachusetts, Amherst.
Broselow, Ellen. 1992. Parametric variation in Arabic dialect phonology. In Ellen Broselow, Mushira Eid & John J. McCarthy (eds.) *Perspectives on Arabic linguistics IV*, 7–47. Amsterdam & Philadelphia: John Benjamins.
Broselow, Ellen. 2008. Stress–epenthesis interactions. In Bert Vaux & Andrew Nevins (eds.) *Rules, constraints, and phonological phenomena*, 121–148. Oxford: Oxford University Press.
Broselow, Ellen, Marie Huffman, Su-I Chen & Ruohmei Hsieh. 1995. The timing structure of CVVC syllables. In Mushira Eid (ed.) *Perspectives on Arabic linguistics VII*, 119–138. Amsterdam & Philadelphia: John Benjamins.
Broselow, Ellen, Su-I Chen & Marie Huffman. 1997. Syllable weight: Convergence of phonology and phonetics. *Phonology* 14. 47–82.
Burzio, Luigi. 1994. *Principles of English stress.* Cambridge: Cambridge University Press.
Cantineau, Jean. 1939. Remarques sur les parlers de sédentaires syro-libano-palesintinens. *Bulletin de la Société de Linguistique de Paris* 40. 80–88.
Cantineau, Jean. 1946. *Les parlers arabes du Hôrân.* Paris: C. Klincksieck.
Cantineau, Jean. 1960. *Etudes de linguistique arabe.* Paris: C. Klincksieck.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.
Cowell, Mark W. 1964. *A reference grammar of Syrian Arabic (based on the dialect of Damascus).* Washington, DC: Georgetown University Press.
El-Hajjé, Hassan. 1954. *Le parler arabe de Tripoli (Liban).* Paris: C. Klincksieck.
Erpenius, Thomas. 1656. *Grammatica Arabica.* Leiden.
Erwin, Wallace M. 1963. *A short reference grammar of Iraqi Arabic.* Washington, DC: Georgetown University Press.
Feghali, Michel. 1919. *Le parler de Kfar-Šabîda (Liban–Syrie).* Paris: C. Klincksieck.
Ferguson, Charles A. 1956. Review of Harris Birkeland, *Stress patterns in Arabic. Language* 32. 384–387.

Fleisch, Henri. 1974. *Etudes d'arabe dialectal.* Beirut: Dar el-Machreq.

Fudge, Erik C. 1969. Syllables. *Journal of Linguistics* 5. 253–286.

Garbell, Irene. 1958. Remarks on the historical phonology of an eastern Mediterranean dialect. *Word* 14. 303–337.

Goitein, Shelamo D. 1970. *Jemenica: Sprichwörter und Redensarten aus Zentral-Jemen mit zahlreichen Sach- und Worterläuterungen.* Leiden: Brill.

Grotzfeld, Heinz. 1965. *Syrisch-Arabische Grammatik (Dialekt von Damaskus).* Wiesbaden: Harrassowitz.

Halle, Morris & Michael Kenstowicz. 1991. The Free Element Condition and cyclic versus noncyclic stress. *Linguistic Inquiry* 22. 457–501.

Halle, Morris & Jean-Roger Vergnaud. 1978. Metrical structures in phonology. Unpublished ms., MIT.

Halle, Morris & Jean-Roger Vergnaud. 1979. Metrical phonology. Unpublished ms., MIT.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress.* Cambridge, MA: MIT Press.

Harrell, Richard. 1957. *The phonology of colloquial Egyptian Arabic.* New York: American Council of Learned Societies.

Harrell, Richard. 1960. A linguistic analysis of Egyptian radio Arabic. In Charles A. Ferguson (ed.) *Contributions to Arabic linguistics,* 3–77. Cambridge, MA: Harvard University Press.

Harris, John & Edmund Gussmann. 2002. Word-final onsets. *UCL Working Papers in Linguistics* 14. 1–42.

Hayes, Bruce. 1979. Extrametricality. *MIT Working Papers in Linguistics* 1. 77–86.

Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13. 227–276.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Hyde, Brett. 2006. Towards and uniform account of prominence-driven stress. In Eric Bakovic, Junko Itô & John McCarthy (eds.) *Wondering at the natural fecundity of things: Essays in honor of Alan Prince,* 139–183. Santa Cruz: Linguistics Research Center, University of California, Santa Cruz.

Irshied, Omar & Michael Kenstowicz. 1984. Some phonological rules of Bani-Hassan Arabic: A Bedouin dialect. *Studies in the Linguistic Sciences* 14. 109–147.

Jakobson, Roman. 1971. *Selected writings,* vol. 1: *Phonological studies.* The Hague: Mouton.

Janssens, Gerard. 1972. *Stress in Arabic and word structure in the modern Arabic dialects.* Leuven: Peeters.

Jastrow, Otto. 2007. Iraq. In Versteegh *et al.* (2008), 414–424.

Jiha, Michel. 1964. *Der arabische Dialekt von Bišmizzīn.* Beirut: Franz Steiner.

Johnson, C. Douglas. 1979. Opaque stress in Palestinian. *Lingua* 49. 153–168.

Kager, René. 1995. The metrical theory of word stress. In John A. Goldsmith (ed.) *The handbook of phonological theory,* 367–402. Cambridge, MA & Oxford: Blackwell.

Kager, René. 1999. *Optimality Theory.* Cambridge: Cambridge University Press.

Kager, René. 2009. Stress. In Versteegh *et al.* (2009), 344–353. Leiden: Brill.

Kenstowicz, Michael. 1980. Notes on Cairene Arabic syncope. *Studies in the Linguistic Sciences* 10. 39–54.

Kenstowicz, Michael & Kamal Abdul-Karim. 1980. Cyclic stress in Levantine Arabic. *Studies in the Linguistic Sciences* 10. 55–76.

Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10. 421–441.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm,* 3–91. Seoul: Hanshin.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–365.

Index

Index prepared by WordCo Indexing Services.
Page locators in bold indicate major blocks of information on a topic.

10 The Other Hand in Sign Language Phonology

Onno Crasborn

1 Introduction

Signed languages used by Deaf communities use the visual channel for communication. Multiple articulators are involved in producing all phonological material, including the head (its rotation around all three axes), many if not all aspects of facial expression (including eye gaze), the position of the upper body, and of course the two hands. It is only slowly becoming clear how linguistic the use of some articulations, e.g. eye gaze, really is (Metzger 1998; Meurant 2008). For most articulators and their phonological representations in the lexicon, we still have a very limited view on their actual use across the sign languages of the world (Eccarius and Brentari 2007).

The linguistic status of the two hands has been argued for from the very start of research on sign phonology, together with the features that make up their shape, rotation, location, and movement (CHAPTER 9: HANDSHAPE IN SIGN LANGUAGE PHONOLOGY; CHAPTER 24: THE PHONOLOGY OF MOVEMENT IN SIGN LANGUAGE; CHAPTER 56: SIGN SYLLABLES): Stokoe's (1960) ground-breaking analysis of the lexicon of American Sign Language (ASL) showed that some signs only have specifications for one hand, while others have a specification for two hands. That is, the distinction between one-handed and two-handed signs plays a role in the lexicon and needs to be specified there. While this has not been contested in any phonological analysis since, researchers have proposed different models of the two hands. This chapter will primarily focus on the possible phonological representations in the lexicon (§2 and §3), yet will also look beyond the lexicon to see to what extent the two hands can function independently in expressing separate morphemes simultaneously (§4). An overview of open questions is presented in §5.

1.1 *The two hands and the modality difference*

All sign languages that have been described so far show a lexical distinction between one-handed and two-handed signs, as a quick look through any (picture) dictionary will show. Some examples from NGT (*Nederlandse Gebarentaal*; Sign Language of the Netherlands) are presented in Figure 10.1.

Figure 10.1 Examples of one-handed and two-handed signs from NGT. (a) One-handed signs: TIRED, HOLIDAY, DOG; (b) two-handed signs: TEA, GIFT, PROCESS

In principle, the fact that there are two symmetrical articulators in the visuo-spatial modality used in signed languages is one of the most salient differences between sign and speech. Humans have one mouth, but two hands. The fundamental question one would like to answer is whether this does or does not have an impact on the linguistic structure of the two types of languages. It is clear that deaf signers do not "speak with two tongues" – their left and their right hand do not continuously convey independent messages. At a phonetic level, aside from the motoric difficulty of permanently controlling the two symmetrical limbs completely independently, it would be impossible to visually perceive two completely independent movements (CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY). At a higher cognitive level of processing as well, it is not easy to express two independent thoughts at the same time, and the same holds for parsing two information streams in perception. In social interaction between humans, there might be great advantages to being able to utter statements on a primary level while commenting on them on a "meta" level. One could thus communicate along two independent paths simultaneously, but this is clearly not how human interaction works in daily life, even though "tone of voice" may reveal some of our attitudes toward what is being said. The impact of having two symmetrical articulators on the linguistic structure of signed languages is thus likely to be more subtle in nature.

This chapter summarizes the different proposals that have been made to describe the lexical phonology of two-handed signs and their relation to one-handed signs, but it will also look at some of the other roles that can be played by the "other hand" outside the lexicon. The overall conclusion will be that, given the many possibilities for simultaneous activities of the two hands, primarily at the grammatical and discourse levels, many questions remain open for investigation. Thus we cannot yet give a satisfactory answer to the question of the possible impact

(1953) and Stokoe (1960) were devoted to making the case for signed languages as real languages. In phonology, this involved the investigation of the extent to which the structure of lexical signs could be characterized with the same types of structures proposed for spoken language phonology. While it is clear that the phonetic content of signed and spoken languages differs, the hypothesis was tested that grammars, as more abstract properties of phonological structure, are highly similar, if not identical. Thus, for example, Stokoe (1960) proposed the decomposition of the form of lexical items in terms of phonemes (giving them a different name, "cheremes," to emphasize the differences in phonetic content and the fact that they were structured simultaneously rather than sequentially), similar to the phonemic analysis of that time (see CHAPTER 11: THE PHONEME). Later, Sandler (1989) proposed a model of ASL phonology in terms of feature trees, similar to Clements's (1985) feature geometry model for spoken language phonology (see CHAPTER 27: THE ORGANIZATION OF FEATURES). In each of these models, the role of the second hand did not require any specific theoretical concept. It was one of the parameters or one of the features – but not something outside the spectrum of phonological constructions found up to that time. As Sandler and Lillo-Martin (2006) summarize, the grammars of signed languages studied so far strongly resemble those of spoken languages when one abstracts away from the concrete phonetic substance of the two modalities. For this reason, there is no doubt that signed languages are "true languages," whatever definition one likes to use. Still, it remains an interesting question how the particular phonetic substance of a second symmetrical articulator finds a place in a phonological model.

Stokoe (1960) considered the main phonological parameters of a lexical sign to be handshape ("dez"), location ("tab"), and movement ("sig"). Orientation was regarded as "minor parameter," because it typically does not fulfill any distinctive function and its phonetic form is predictable in many signs. Interestingly, although one-handed signs were clearly set apart from two-handed signs in the notation (using one *vs.* two handshape symbols), something like "handedness" was not explicitly considered to be a distinctive property of the phonology of the sign. Whether a sign has one or two hands needs to be specified, but this parameter was not assigned the status of the basic aspects, handshape, location, and movement. The lack of a large set of minimal pairs involving orientation may have been the reason for this choice.

In all studies since, whether on ASL or other signed languages, there has never been a claim that one- *vs.* two-handedness is predictable on the basis of other phonological properties, or of semantic (or iconic) properties of the sign. Thus it needs to be lexically specified, and as such is a distinctive phonological feature of lexical items (see CHAPTER 2: CONTRAST; CHAPTER 17: DISTINCTIVE FEATURES). In the languages studied to date, very few minimal pairs have been identified in support of this distinctive nature of two-handedness. A rare example from NGT is presented in Figure 10.2.

Stokoe introduced a basic distinction among two-handed signs that has been taken over by most phonological models since then: in some signs both hands move, while in others one hand appears to function as the location of the other hand ("double dez" and "tab-dez," respectively, in Stokoe's terminology). Some examples of such more symmetric and less symmetric signs from NGT are presented in Figures 10.3 and 10.4. When both hands move, this can be fully symmetric (Figure 10.3a), or in an alternating fashion (Figure 10.3b, c). Various

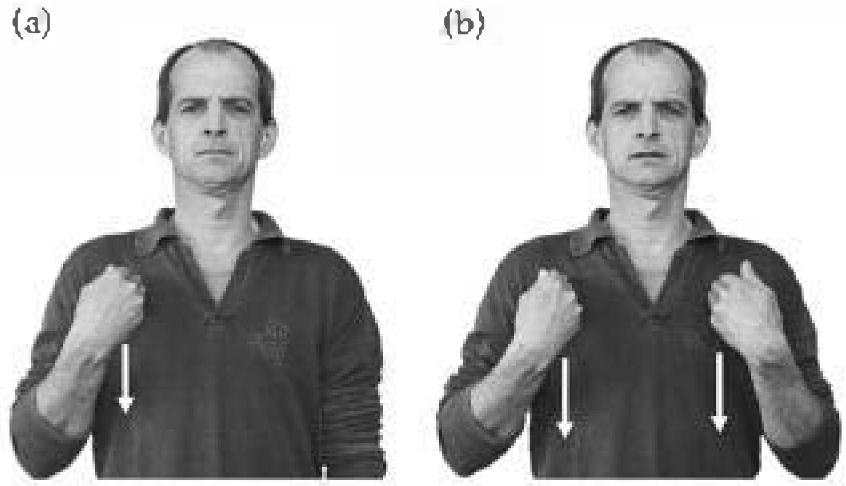


Figure 10.2 Minimal pairs in NGT based on number of hands: (a) HUMAN-BEING and (b) BEHAVIOR

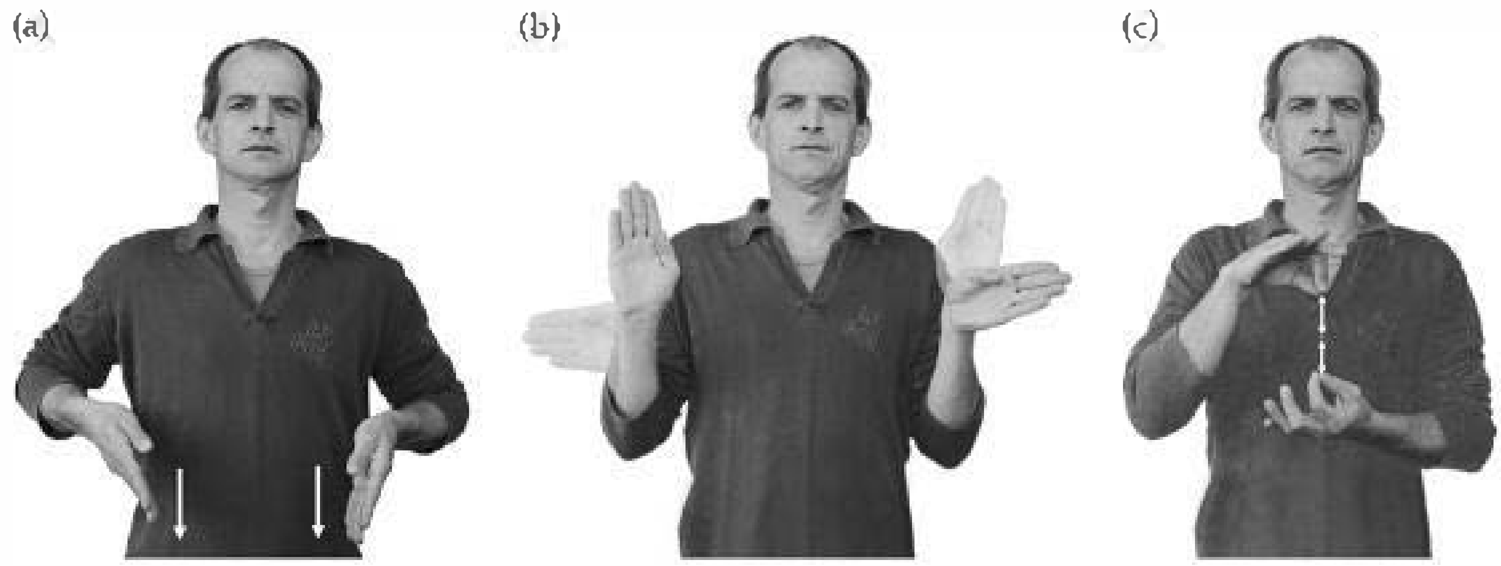


Figure 10.3 Examples of symmetrical signs: both hands move: (a) STAND, (b) TRAFFIC, (c) WEAKLING

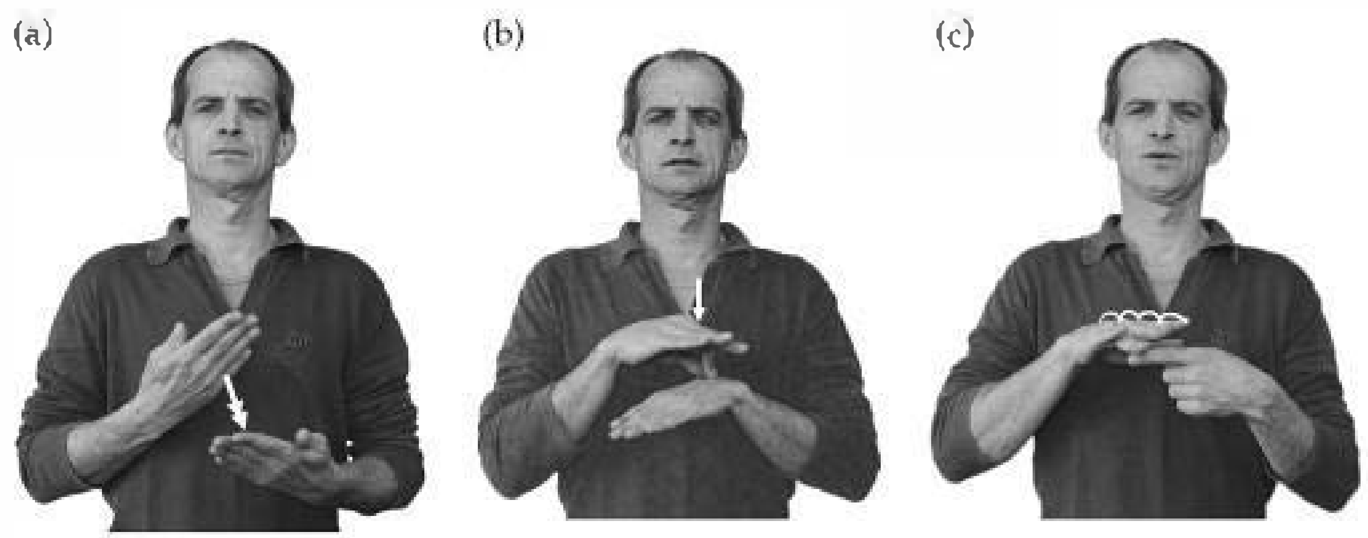


Figure 10.4 Examples of asymmetrical signs: one hand acts as the location for the other hand: (a) EVIDENCE, (b) ILL, (c) PHONOLOGY

types of asymmetry are further discussed in Padden and Perlmutter (1987) and Crasborn (1995).

2.2 *Battison's conditions*

Battison (1974, 1978) elaborated on the basic distinction between symmetric and asymmetric signs by observing two phonological regularities in the set of two-handed signs, which he characterized as morpheme structure constraints. They are the Symmetry Condition and the Dominance Condition, given in (1).

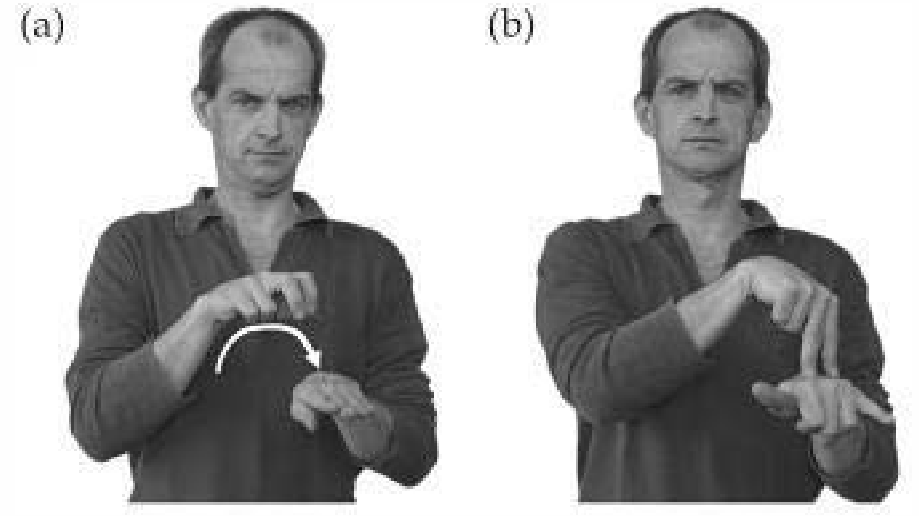


Figure 10.6 Classifier constructions from NGT. (a) 'Human or animal jumps onto a flat surface'; (b) 'Human stands on an airplane'

two-handed classifier constructions in many signed languages. These complex morphological constructions feature classifier hand configurations that each represent a separate morpheme, and (typically) the location and movement of each involves additional morphemic material. Two examples are presented in Figure 10.6.

While the example of a classifier construction from NGT in Figure 10.6a obeys the Dominance Condition for lexical items in (1b), the example in Figure 10.6b would be ruled out for monomorphemic signs, in ASL as well as in NGT: the Y handshape found in the non-dominant hand in Figure 10.6b does not occur on the weak hand in asymmetric signs in these languages. Eccarius and Brentari (2007) focus on this type of sign and classifier construction, to which Battison's Dominance Condition applied: the handshapes are different. Rather than looking at the two handshapes as a whole, they compare the selected fingers and their configuration within the handshapes. Each feature class can have marked or unmarked values (see CHAPTER 9: HANDSHAPE IN SIGN LANGUAGE PHONOLOGY). This allows the revision of the Dominance Condition as in (2), quoted from Eccarius and Brentari (2007: 1187), which also covers constructions such as the one in Figure 10.5b (italics in the original):

(2) *Revised Dominance Condition*
(a) If the two hands do not share the same specification for *both selected fingers and joints* (i.e., *the handshapes* are different), then (b) one hand must be passive while the active hand articulates the movement, and (c) *the form as a whole* (i.e., *selected fingers and joints for both hands*) *is limited to two marked phonological structures, only one of which can be on the passive hand.*

In terms of handshape features, the construction in Figure 10.6b shows a marked finger selection for the dominant hand (index and middle finger) and on the non-dominant hand (thumb and pinkie), making for a total of two marked structures in the whole form. All selected fingers are extended (unmarked), and thus do not add to the complexity beyond what the Revised Dominance Condition allows. A combination of the two hands where the dominant hand shows curved fingers, as in Figure 10.6a, is thus predicted to be unacceptable (a prediction that has yet to be tested for NGT). Thus, it would appear that the phonological complexity allowed for monomorphemic two-handed signs within the lexicon is not exceeded in morphologically complex forms.

Figure 10.9 The poem "Hope," by Wim Emmerik. Left hand: GUN, SHOOT-GUN, HUMAN-FALL; right hand: HOPE, DO, LIFE

finger for a certain time span ("theme buoy"). In an analysis of NGT poetry by Wim Emmerik, Crasborn (2006) shows that various types of buoy constructions occur in addition to the simultaneous realization of unrelated lexical items; this happens so frequently that one-handed articulations are virtually absent. When they are present, or when the two hands take turns in articulating a string of one-handed signs, this has a marked poetic effect. Figure 10.9 shows one of Emmerik's poems, in which each hand articulates three signs in sequence, the two hands articulating two different phrases at the same time. In this case, the left hand signs the optimistic Dutch expression *hoop doet leven* ("where there's hope, there's life"), while the other hand is pointing a gun at the head and shooting someone. Emmerik (1993) explains that this poem aims to express the two sides of the coin of a medical development (cochlear implants that promise to make deaf people hear) that was starting to gain popularity at the time. Similarly, Brentari (1998) discusses an example from a poem by the ASL poet Clayton Valli in which the hands act independently to express two different phrases simultaneously.

While this chapter mostly focuses on the phonological analysis of one-handed *vs.* two-handed lexical items, it is clear that the non-dominant hand also plays a role outside the lexicon, and that the uses of the two hands beyond the lexicon have been understudied. On the basis of the review of phonological representations in §3 and the discussion of some phonetic and phonological processes in §4, some of these open questions will be discussed in §5.

5 Conclusion

In summary, the striking modality difference of having two symmetrical articulators in signed but not in spoken languages does not appear to lead to wildly different lexical structures. Although in the signed languages that have been studied there is no prominent distinctive role in the lexicon for one-handedness *vs.* two-handedness, it is also clear that two-handedness has to be specified in the lexicon for each lexical item: whether one or two hands are involved in the articulation is not predictable from other properties of a sign. The actual phonological role of the second hand is rather restricted: it can be a copy of the dominant hand, or a place of articulation. The possibility of independence of the two hands is, however, exploited at the prosodic level, where spreading of the non-dominant articulator across one or more lexical items on the dominant hand can visually mark the size of certain prosodic units. In addition, the prolonged presence of certain morphemes can be exploited by the syntactic and discourse levels of the grammar, often yielding complex structures that await further investigation and analysis. These "simultaneous constructions," including the classifier constructions, reveal new questions about the relation between the lexical form of one and two-handed signs and their appearance in discourse. I briefly discuss some of these and related questions in the rest of this section.

While in the introduction it was stated that many signers by default use their preference hand as the dominant hand, it has also been observed that people can quite easily switch the left and the right hand. It is not yet clear when exactly this "dominance reversal" occurs in different languages, and to what extent its occurrence interacts with the lexical distinction between symmetric and asymmetric signs discussed in this chapter. Can some types of one-handed or two-handed sign perhaps block dominance reversal, or force it to occur at a different point in the discourse? Also, to what extent is the choice of dominant articulator influenced by the use of spatial locations on the left *vs.* on the right? Referents and parts of discourse can be localized by employing a left–right contrast (e.g. Friedman 1975; Lillo-Martin and Klima 1990); a question that arises is to what extent this triggers left-handed *vs.* right-handed articulations of sentences.

Several researchers have noted that classifier morphemes, as well as other types of articulations of the non-dominant hand, can be held at their (final) location for not just one Phonological Phrase (as with "meaningless," purely prosodic, spreading of the non-dominant hand), but for a number of sentences (Liddell 2003; Sandler 2006). How can we analyze such structures in terms of prosodic units? The largest unit in the prosodic hierarchy proposed by Nespor and Vogel (1986) is the Utterance, which can encompass several connected Intonational Phrases (or in syntactic terms, several clauses or even whole sentences joined by connectives). But is this unit large enough for the types of constructions that we are discovering in signed languages? Or is there a larger, discourse-level unit that simply forms a "Discourse Phrase," say, that can join several prosodic Utterances?

Longer simultaneous constructions at a morphosyntactic and discourse level also open up a new area of evidence for the study of Weak Drop: which types of signs block weak hand spreading because they do not allow for a one-handed articulation, and which do not? Are there circumstances where the process of weak hand spreading is so dominant that all two-handed signs are affected? The arrival of large corpora of semi-spontaneous signed language interaction promises to provide an important source of evidence for the investigation of these types of topics (Crasborn *et al.* 2007; Crasborn 2010; Johnston and Schembri, forthcoming).

Finally, as in other domains of sign language research, questions about linguistic typology remain largely open in the field of phonology. Although we have a reasonable number of elementary sign language dictionaries available from many non-Western signed languages, it is clear that most signed languages have not yet been analyzed in great detail, even at a lexical level. How universal are the (revised) Symmetry and Dominance conditions, for example? Eccarius and Brentari (2007) looked at Hong Kong Sign Language in addition to the Western languages ASL and Swiss German Sign Language, but this study is quite exceptional in this respect. The similarities across unrelated languages found so far, together with the strong phonetic origin of the phonological symmetry patterns,

Engberg-Pedersen, Elisabeth. 1994. Some simultaneous constructions in Danish Sign Language. In Brennan & Turner (1994), 73–88.
Ewen, Colin J. 1995. Dependency relations in phonology. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 570–585. Cambridge, MA & Oxford: Blackwell.
Friedman, Lynn A. 1975. Space, time, and person reference in American Sign Language. *Language* 51. 940–961.
Frishberg, Nancy. 1985. Dominance relations and discourse structures. In William C. Stokoe & Virginia Volterra (eds.) *SLR '83: Proceedings of the 3rd international symposium on sign language research*, 79–90. Rome: Consiglio Nazionale delle Ricerche.
Gijn, Ingeborg van. 1997. Vormkarakteristieken van gestures en gebaren. M.A. thesis, University of Leiden.
Hulst, Harry van der. 1993. Units in the analysis of signs. *Phonology* 10. 209–241.
Hulst, Harry van der. 1996. On the other hand. *Lingua* 98. 121–143.
Hulst, Harry van der & Wendy Sandler. 1994. Phonological theories meet sign language: Two theories of the two hands. *Toronto Working Papers in Linguistics* 13. 43–73.
Johnston, Trevor & Adam Schembri. 2007. *Australian sign language: An introduction to sign language linguistics*. Cambridge: Cambridge University Press.
Johnston, Trevor & Adam Schembri. Forthcoming. Corpus analysis of sign languages. In Carol A. Chapelle (ed.) *Encyclopedia of applied linguistics*. Oxford & Malden, MA: Wiley-Blackwell.
Kooij, Els van der. 2001. Weak drop in Sign Language of the Netherlands. In Valery L. Dively, Melanie Metzger, Sarah Taub & Anne Marie Baer (ed.) *Signed languages: Discoveries from international research*, 27–42. Washington, DC: Gallaudet University Press.
Kooij, Els van der. 2002. Phonological categories in Sign Language of the Netherlands: Phonetic implementation and iconic motivation. Ph.D. dissertation, University of Leiden.
Leeson, Lorraine & John I. Saeed. 2004. Windowing of attention in simultaneous constructions in Irish Sign Language. In Terry Cameron, Christopher Shank & Keri Holley (ed.) *Proceedings of the 5th High Desert Linguistics Conference*, 1–18. Albuquerque: University of New Mexico.
Liddell, Scott K. 2003. *Grammar, gesture and meaning in American Sign Language*. Cambridge: Cambridge University Press.
Liddell, Scott K. & Robert E. Johnson. 1986. American Sign Language compound formation processes, lexicalization and phonological remnants. *Natural Language and Linguistic Theory* 4. 445–513.
Liddell, Scott K. & Robert E. Johnson. 1989. American Sign Language: The phonological base. *Sign Language Studies* 64. 197–277.
Liddell, Scott K., Marit Vogt-Svendsen & Brita Bergman. 2007. A crosslinguistic comparison of buoys: Evidence from American, Norwegian, and Swedish Sign Language. In Vermeerbergen *et al.* (2007b), 187–216.
Lillo-Martin, Diane & Edward S. Klima. 1990. Pointing out differences: ASL pronouns in syntactic theory. In Susan D. Fischer & Patricia Siple (eds.) *Theoretical issues in sign language research*, vol. 1: *Linguistics*, 191–210. Chicago: Chicago University Press.
Metzger, Melanie. 1998. Eye gaze and pronominal reference in American Sign Language. In Ceil Lucas (ed.) *Pinky extension and eye gaze: Language use in deaf communities*, 170–182. Washington, DC: Gallaudet University Press.
Meurant, Laurence. 2008. *Le regard en langue des signes. Anaphore en langue des signes françaises de Belgique (LSFB): Morphologie, syntaxe, énonciation*. Namur: Presses Universitaires de Namur & Presses Universitaires de Rennes.
Miller, Christopher. 1994a. Simultaneous constructions in Quebec Sign Language. In Brennan & Turner (1994). 89–112.
Miller, Christopher. 1994b. Simultaneous constructions and complex signs in Quebec Sign Language. In Inger Ahlgren, Brita Bergman & Mary Brennan (eds.) *Perspectives*

11 The Phoneme

B. Elan Dresher

1 Introduction

The concept of the phoneme was central to the development of phonological theory. In the early twentieth century, phonological theory was all about the phoneme: how to define it, how to recognize it, how to discover it (see, for example, the articles selected for inclusion in Joos 1957 and Makkai 1972). The American structuralist term for phonology, phonemics, indicates to what extent the field was considered to be about the phoneme.

Things have now changed. The phoneme, to all appearances, no longer holds a central place in phonological theory. Two recent and voluminous handbooks devoted to phonology, edited by Goldsmith (1995) and by de Lacy (2007), have no chapter on the phoneme. It is barely mentioned in the indexes. This does not mean that the phoneme plays no role in modern phonology; closer inspection reveals that the phoneme is far from dead. However, it is not much talked about, and when it is, it is more often to dispute its existence than to affirm it.

Such a dramatic change in fortunes for a concept bears some looking into, and this chapter will be devoted to trying to understand what has happened to the phoneme in its journey into the twenty-first century, and what its prospects are for the future.

2 Origins of the term

S. R. Anderson (1985: 38) cites Godel (1957) and Jakobson (1971) as locating the origin of the term *phoneme* in the French word *phonème*, coined in the early 1870s by the French linguist Dufriche-Desgenettes. He proposed the term to substitute for the German *Sprachlaut* ("speech sound"), so it did not have the modern sense of phoneme, but rather corresponded to what we would now call "speech sound" or "phone." The term was taken up by Saussure (1879), who used it in yet a different sense, and from Saussure it was taken up by the Polish Kazan school linguists Jan Baudouin de Courtenay and Mikołaj Kruszewski.

S. R. Anderson (1985: 60–68) traces how the meaning of the term evolved from Saussure's use to the one that ultimately emerged from the Kazan school (for detailed accounts of the history of the phoneme see also Krámský 1974; Fischer-Jørgensen 1975). Saussure (1879) used it in his historical work on Indo-European to refer to a hypothesized sound in a proto-language together with its reflexes in the daughter languages, what we might call a "correspondence set." For example, if a sound that is reconstructed as *g in the proto-language has reflexes *g*, *h*, and *k* in three daughter languages, then the set {*g*, *h*, *k*} would constitute a "*phonème*" for Saussure.

Kruszewski recast the notion in synchronic terms to refer to a set of alternating elements; for example, if the same morpheme has a final [g] before suffixes beginning with a back vowel, a palatalized [g^j] before suffixes beginning with a front vowel, and a [k] when it is word-final, the alternation "[g] before a back vowel, [g^j] before a front vowel, and [k] when final" would constitute a "phoneme." Subsequently, Baudouin reinterpreted the term "phonemes" as referring to the abstract, invariant psychophonetic elements that alternate; in the above example, one could posit a phoneme /g/ that participates in the alternations that cause it to be realized as [g], [g^j], or [k], depending on the context.

In a final step, the term was extended also to sounds that do not alternate, thereby arriving at a conception of the phoneme as "the psychological equivalent of a speech sound" (Baudouin de Courtenay 1972: 152). It is in this sense that the phoneme entered phonological theory in Europe and North America.

3 General concept of the phoneme

The general concept of the phoneme preceded the term or its exact definition, which is a more difficult enterprise. The basic concept is that of the unity of sounds that are objectively different but in some sense functionally the same. As Twaddell (1935: 55) observes, this concept is not new: if a special term was not needed before the late nineteenth century, it is because in the absence of close phonetic observation, it is not necessary to distinguish between "phoneme" and "speech sound." Alphabetic writing systems tend to have separate letters only for sounds that have a distinctive function, though deviations from this principle occur (Krámský 1974: 10; Fischer-Jørgensen 1975: 4). In ordinary parlance one talks of the sound "d" or "k" as if each of these represents a single sound, rather than, as is the case, a range of sounds.

Parallel to the development of the phonemic concept as part of phonological theory mentioned above, British and French phoneticians who laid the foundations for what became the International Phonetic Association (IPA) arrived at a similar notion motivated by more practical concerns. According to Jones (1967: 256), Henry Sweet (1877) was the first to draw a distinction between "narrow" and "broad" transcription: narrow transcription aims (in principle) to record sounds in as much detail as possible, whereas broad transcription records only distinctive differences in sound. It was recognized early on that the goal of assigning a unique symbol to every sound in every language, even if it could be realized, would lead to transcriptions for particular languages that would be impractical and virtually illegible. Therefore, Paul Passy insisted in 1888 that only distinctive differences should be recorded, and called this principle *une règle d'or* ("a golden rule") from which one should never depart (cited in Jones 1967: 256). Thus, while the IPA is popularly known for developing a universal phonetic alphabet that is associated with phonetic ("narrow") transcription, its founders insisted on "broad" (i.e. phonemic) transcription for purely practical reasons. The practical strain remained influential in phonological theory, as attested by the subtitle of Pike's (1947) *Phonemics: A technique for reducing languages to writing.*

It is hard to imagine what linguistic description would be like without a phoneme concept of some sort. To take one entirely typical example, the Australian language Pitta-Pitta (Pama-Nyungan) is said to have three vowels, *i*, *a*, and *u* (Blake 1979: 187). In describing their pronunciation, Blake writes that they "are similar to the vowels of 'been', 'balm', and 'boot' respectively" (presumably [i], [ɑ], and [u]). Further reading reveals that this is only true in open syllables, and when stressed, and when near certain consonants. In a closed syllable, "they are similar to the vowels of 'bin', 'bun', and 'put'" ([ɪ], [ʌ], and [ʊ]). Further, the vowel *a* is pronounced [æ] in the vicinity of a palatal consonant, and unstressed *a* has a schwa-like pronunciation, [ɐ]. Objectively, then, Pitta-Pitta has at least eight different vowel sounds, and probably many more if we were to attend to further distinctions in different segmental and prosodic contexts, and in different situations and for different speakers.

This variation does not detract from the fact that there is an important sense in which this language has three vowels. In the distribution given above, we recognize that the variation is a consequence of the influence of context, and has no contrastive function: [i] and [ɪ] are variants of a phoneme we can designate as /i/, [u] and [ʊ] are variants of /u/, and [ɑ], [ʌ], [æ], and [ɐ] are variants of /a/. Put differently, in every slot where a vowel belongs we have only three choices in this language. If we are told that a word begins with the sequence *m*-vowel-*rr*-, we know that the vowel must be one of the variants of /a/ (e.g. *marra* 'open'), /i/ (e.g. *mirri* 'little girl'), or /u/ (e.g. *murra* 'stick').

4 Defining the phoneme

In the 1930s many linguists came to share the intuition that a concept like the phoneme is needed in phonological description.[1] Pinning down the definition of this concept proved to be difficult. Like other linguistic notions, such as "sentence," "syllable," and "topic," what starts out as a relatively unproblematic intuitive concept inevitably gets caught up in theory-internal considerations. In the case of the phoneme, three issues have been particularly contentious: (i) what sort of entity is the phoneme (physical, psychological, other); (ii) what is the content of the phoneme; and (iii) how does one identify phonemes?

4.1 *What type of entity is the phoneme?*

Twaddell (1935) surveyed the various definitions of the phoneme that were then in circulation, and classified them as being of two main types. One type assumes that the phoneme is a physical reality, and the other assumes that it is a psychological notion.

[1] Acceptance of the phonemic principle was by no means universal, however, particularly among traditional grammarians and writers of historical grammars. The phoneme does not appear in Campbell's (1959) *Old English grammar*, to the general applause of reviewers (see Dresher 1993 for discussion); its first appearance in a traditional-style Old English grammar is Hogg (1992).

be empirically tested. While it is no doubt correct that appealing to a vague and unknown "mind" cannot serve as an adequate explanation (*explanans*) of any phenomenon, the cognitive revolution that began in the 1950s has shown the fruitfulness of studying mental representations and processes as things to be explained (*explananda*).

4.1.3 The phoneme as a fiction

The consequence of rejecting both physical and psychological reality for the phoneme is that Twaddell (1935) is forced to conclude that the phoneme, though an "eminently useful" term, is a fictitious unit. There exist philosophies of science in which useful, indeed indispensible, units can be fictions, but most linguists since the 1950s have taken a "realist" view of linguistics (Chomsky 1980: 104–110). From this perspective, a unit that is required to give an adequate account of some phenomenon must be real at some level. Once we abandon empiricist assumptions about science and psychology, there is no obstacle to considering the phoneme to be a psychological entity.

4.2 *What is the content of the phoneme?*

It is one thing to locate the phoneme as a psychological (or physical) concept; it remains to try to characterize the content of the phoneme. What are phonemes made of? How are they represented? In this section I review some different approaches to these questions.

4.2.1 The phoneme as a set of contrastively underspecified features

Sapir's "point in the pattern." A particularly influential psychological conception of the phoneme was that of Sapir (1925, 1933). For Sapir (1925), each phoneme occupies a particular point in the sound pattern of a language. For example, he proposes that the hypothetical languages he calls C and D have the identical pattern, even though phonetic details differ. What is important is that each consonant in C has a corresponding consonant in D that occupies the same point in the pattern (the inventories in (1) maintain Sapir's arrangement, though I have updated his notation to modern IPA symbols).

(1) *Phonemes with identical patterning* (Sapir 1925)

a. *Pattern of C*

h		w	j	l	m	n
p	t		k	q		
b	d		g	c		
f	s		x	χ		

b. *Pattern of D*

h		v	ʒ	r	m	ŋ
pʰ	tʰ		kʰ	qʰ		
β	ð		ɣ	ʁ		
f	ʃ		ç	ħ		

In other terms, Sapir's "point in the pattern" refers to the contrastive status of a phoneme, and the way it relates to other phonemes in the system. Though Sapir did not assume a theory of distinctive features (CHAPTER 17: DISTINCTIVE FEATURES),

some such notion appears to be needed to make this notion explicit (Dresher 2009: 38–39). Thus, the series /b d g ɢ/ in language C can be characterized as being contrastively obstruent and voiced, properties it shares with the series /β ð ɣ ʁ/ in language D.

Sapir points out further that the sound pattern of a language is guided by the phonetics but may deviate from it. For example, /ʒ/ in language D is not classified with the voiced obstruents, but rather with the sonorants, corresponding to /j/ in language C. Thus, this sound is physically an obstruent, but psychologically and functionally a sonorant.

Sapir (1933) goes further in characterizing the phoneme as a psychological unit, arguing that "the phonemic attitude is more basic, psychologically speaking, than the more strictly phonetic one," setting off a debate about the psychological reality of phonemes that is still ongoing. He argues that perception in terms of phonemes accounts for difficulties native speakers have in grasping certain phonetic facts about their language, or perceiving "correctly" the objective sounds before them (see §6). Sapir's interpretation of these "errors" has been disputed over the years, but his work did much to establish the phoneme, and the "-emic attitude" more generally, as an important psychological and symbolic unit.

Prague School: Phonemic make-up or content. We observed that an explication of Sapir's notion of "point in the pattern" benefits from thinking of phonemes as possessing contrastive properties. This idea was carried further by phonologists of the Prague School, notably Jakobson and Trubetzkoy. The notion of opposition (or contrast between two phonemes) was central to their conception (CHAPTER 2: CONTRAST). Analysis of the nature of oppositions requires that phonemes be characterized as possessing features. The contrastive features necessary to distinguish a phoneme from others in the same system contribute to the phonemic make-up (Jakobson) or phonemic content (Trubetzkoy) of the phoneme.

Jakobson (1962) cites the observation of Hála that the simple vowels of Slovak are almost identical to the vowels of Standard Czech except for an additional short front vowel, /æ/, that occurs in dialects of Central Slovak (2).

(2) *Czech and Slovak vowel systems* (Jakobson 1962: 224)

a. *Standard Czech*			b. *Standard Slovak*	
i		u	i	u
e		o	e	o
	a		æ	a

Jakobson notes (1962: 224) that the presence of /æ/ in Slovak, though "a mere detail from a phonetic point of view . . . determines the phonemic make-up of all the short vowels." Thus all the short vowels in Standard Slovak come in pairs that contrast in the frontness/backness dimension, so that the vowels /i e æ/ are contrastively front (acute, in terms of Jakobson's features), and /u o a/ are contrastively back (grave). Lip rounding, though present phonetically in /u/ and /o/, is not contrastive and therefore does not enter into the phonemic make-up of these vowels.

In Czech the low vowel /a/ is not opposed to another low vowel. Therefore, even though it is almost identical to the Slovak /a/, Jakobson considers it to be neutral with respect to tonality, having no contrastive value except for its height.

Trubetzkoy (1969: 66–67) uses the term "phonemic content" to refer, like Jakobson, to those contrastive properties that characterize phonemes:

b. *Underlying representations*

$$\underset{\begin{bmatrix}+\text{high}\\-\text{round}\end{bmatrix}}{/\text{I}/} \quad \underset{[-\text{high}]}{/\text{A}/} \quad \underset{\begin{bmatrix}+\text{high}\\+\text{round}\end{bmatrix}}{/\text{U}/}$$

c. *Some realization rules*

i. $[\;\;] \rightarrow [+\text{tense}]$ / __ in an open syllable

ii. $[\;\;] \rightarrow [-\text{tense}]$ / __ in a closed syllable

iii. $\begin{bmatrix}-\text{high}\\ \alpha\text{tense}\end{bmatrix} \rightarrow \begin{bmatrix}\alpha\text{low}\\+\text{back}\\-\text{round}\end{bmatrix}$

iv. $\begin{bmatrix}+\text{high}\\ \alpha\text{round}\end{bmatrix} \rightarrow [\alpha\text{back}]$

v. $\begin{bmatrix}-\text{high}\\-\text{stress}\end{bmatrix} \rightarrow$ ɐ

d. *Sample derivations*

	'open'	'little girl'	'stick'
Underlying	/mAr.rA/	/mIr.rI/	/mUr.rA/
Stress	'mAr.rA	'mIr.rI	'mUr.rA
Rules (i)–(v)	'mʌr.rɐ	'mɪr.ri	'mʊr.rɐ

Other versions of underspecification theory have been proposed within generative grammar. In the 1980s, the most notable were Radical Underspecification (Kiparsky 1982, 1985; Archangeli 1984; Pulleyblank 1986) and Contrastive Specification (Steriade 1987). In the 2000s, a number of theories were proposed in which notions of contrast and phonological activity play key roles. Besides MCS, these include the minimalist theories of phonological representation of Hyman (2001a, 2001b, 2003) and Morén (2003, 2006), the theory of feature economy of Clements (2001, 2003, 2009), and the representational economy and underspecification proposal for laryngeal systems of Avery and Idsardi (2001). Other versions of phonological minimalism can be found in Dependency Phonology (Anderson and Ewen 1987; J. M. Anderson 2005; some of the papers in Carr *et al.* 2005) and Radical CV Phonology (van der Hulst 1995, 1996, 2005).

4.2.2 *The fully specified basic variant phoneme*

The model in (4), with each phoneme represented in the lexicon by a single underspecified representation, is not the only view of phonemic representation. S. R. Anderson (1985) traces it to subsequent interpretations of Saussure's notion that what is important in language is differences. Anderson (1985: 43f.) argues that this view, which he calls the "incompletely specified" theory of the phoneme, is not the only, or even the best, interpretation of what Saussure intended. He presents two alternative views. One is what he calls the "fully specified basic variant" phonemic theory. On this approach, one of the surface allophones of a phoneme is chosen as the basic underlying representation. That is, the representation of a phoneme is a full-fledged segment, with all its properties. A set of rules then changes the basic variant to its allophones in the appropriate contexts. Some on underlying representations, a principle known as Richness of the Base. In Pitta-Pitta, for example, the grammar would be required to derive the correct vowel allophones no matter what input vowels are presented to the grammar. A simplified set of constraints governing the Pitta-Pitta high vowel allophones is given in (6) and a sample evaluation is shown in (7).

(6) *Pitta-Pitta in OT*

Some constraints for high vowels

TnsOpen	Vowels are tense in open syllables.
LaxClosed	Vowels are lax in closed syllables.
Ident[high]	Preserve underlying values of [high].
Ident[round]	Preserve underlying values of [round].
Rd=Bk	The value of [back] must be the same as [round].
Ident[back]	Preserve underlying values of [back].
Ident[tense]	Preserve underlying values of [tense].

(7)

/mɨrrɪ/	TnsOp	LaxCl	Id[hi]	Id[rd]	Rd=Bk	Id[bk]	Id[tns]
a. mɨrrɪ	*!				*		
b. mɨrri					*!		*
c. murri				*!			*
d. mirri		*!				*	**
☞ e. mɪrri						*	*

The input in this example is /mɨrrɪ/, an impossible surface form in which both vowels are "wrong": the first vowel, /ɨ/, is not a possible surface vowel in Pitta-Pitta, according to the description above; the second vowel is possible in a closed syllable, but not in an open one. Therefore, the "faithful" candidate (a), which preserves both input vowels, violates two constraints, one for each vowel. Candidate (b) has a legal vowel in the open syllable but the back unrounded vowel is illicit. Candidate (c) repairs the unattested /ɨ/ by making it correspond to [ʊ], a possible sound in this context, but not in correspondence with an underlying [−round] vowel. Candidate (d) changes both vowels to [+tense], incurring a fatal violation of LaxClosed. Candidate (e) is the winner because it alone respects all the constraints in the highest tier, though violating the two lower-ranking constraints.

In this kind of grammar there is no representation of a phoneme /I/ or /i/, nor is there any statement to the effect that [i] and [ɪ] are allophones of a single phoneme. The latter is a consequence of the constraints and the way they interact, ensuring that any input vowel bearing the feature specifications [+high, −round] will surface as [i] in an open syllable and as [ɪ] in a closed syllable, whatever other specifications they start with.

This view of the phoneme has certain affinities with Exemplar Theory and related proposals (Johnson 1997; Bybee 2001; Pierrehumbert 2001; Välimaa-Blum 2009), whereby multiple copies of lexical items are stored in great detail, forming exemplar "clouds" of remembered episodes of individual experience. On this view, speech sounds, too, are stored in terms of exemplar clouds. Some exemplar theorists posit that there are exemplar clouds of phonemes as well as of words

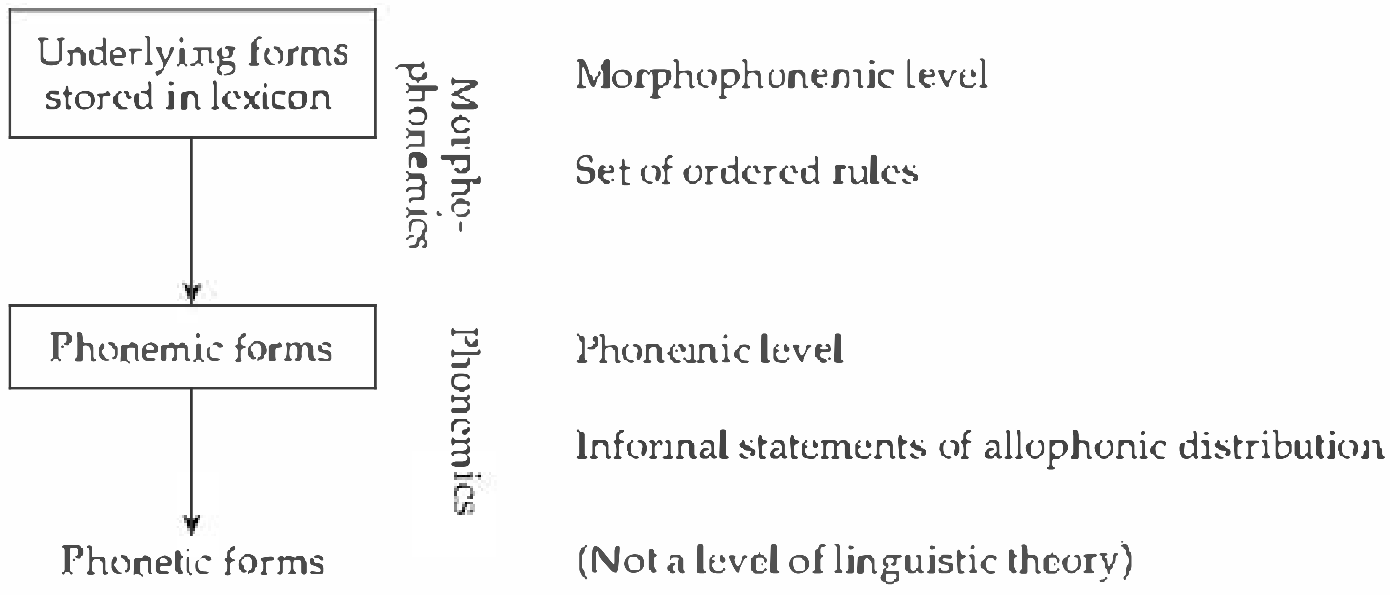


Figure 11.1 Levels in post-Bloomfieldian American structuralist phonology

final obstruent. Voicing is a contrastive feature in Russian that distinguishes pairs of obstruent phonemes: /t/ and /d/ have opposite specifications for the feature [voiced], as do /k/ and /g/, /s/ and /z/, and so on. Since RVA mainly turns one phoneme into another, it must apply in the morphophonemic component:

(13) *Russian regressive voicing assimilation*

a.	*Morphophonemic component*		
	Morphophonemes	//'mok bi//	//'mok lʲi//
	RVA	'mog bi	—
	Phonemes	/'mog bi/	/'mok lʲi/
b.	*Allophonic component*		
	Phonemes	/'mog bi/	/'mok lʲi/
	Other rules	'mog bɨ	—
	Phonetic form	['mog bɨ]	['mok lʲi]
		'were (he) getting wet'	'was (he) getting wet?'

In (13), the phrase 'were (he) getting wet' is realized as ['mog bɨ], where underlying /k/ voices to /g/ before voiced obstruent /b/ (compare [mok lʲi] 'was (he) getting wet?', with a /k/ preceding the sonorant /lʲ/). The rule that changes /k/ to /g/ changes one phoneme to another, and so it must be a morphophonemic rule. This result is forced in any phonemic theory that observes the constraint that allophones of different phonemes may not overlap: in this case, [k] may not be an allophone of both /k/ and /g/.

Halle (1959: 22–23) points out that there are Russian obstruents that do not have voiced counterparts, /ts ʧ x/ (that is, there are no contrasting phonemes /dz ʤ ɣ/). He observes that these phonemes participate in voicing alternations in the same way as other obstruents; in particular, they trigger and undergo RVA (14). Thus, we have ['ʒeʤ bɨ] 'were one to burn', where [ʤ] is the voiced counterpart of [ʧ] (compare ['ʒeʧ lʲi] 'should one burn?', with voiceless [ʧ] before [lʲ]). Because [ʤ] is not a phoneme in its own right, but exists only as an allophone of /ʧ/, this application of voicing is an allophonic rule, and must be assigned to the component that maps phonemic forms into phonetic forms.

Sapir's (1933) arguments for the phoneme as a unit of perception are early examples of this type. Fromkin (1971, 1973) argues that slips of the tongue are a window on linguistic representations and processes. Errors like *teep a cape* for *keep a tape* and [fuwt mijving] for *feet moving* show transpositions of individual segments cut out of the speech stream. Fromkin argues that errors where only one segment in a cluster is involved provide further evidence that individual segments are units of speech performance: examples are *fish grotto* > *frish gotto* and *sticky point* > *spicky point*.[8]

Language games exist in many languages and involve manipulations of various kinds of linguistic units (see Sherzer 1982 and Bagemihl 1995 for overviews). Games that pick out individual segments appear to presuppose a linguistic analysis in which such units are represented. For example, some games involve the exchange of segments: Tagalog /'dito/ > /'doti/ 'here', or Javanese /satus/ > /tasus/ '100' (cited in Bagemihl 1995: 704). Again, much of this does not specifically show evidence for phonemes as opposed to segments.

Neurolinguistic evidence is becoming increasingly influential in finding out about the sort of representations speakers have. Kazanina *et al.* (2006) report that magneto-encephalographic brain recordings reveal that Russian and Korean speakers react differently to tokens of [d] and [t]. In Russian, these sounds are contrastive, members of different phonemes, /d/ and /t/; in Korean, both sounds exist, but they are not contrastive and map into a single phoneme /T/. Russian speakers showed evidence of separating the sounds into two categories, whereas Korean speakers did not. Kazanina *et al.* (2006) conclude that a speaker's perceptual space is shaped not only by the phonetic distribution of sounds, but also by a more abstract phonemic analysis of speech sounds.

6.3 *Evidence from synchronic and diachronic patterning*

The most pervasive sort of evidence for phonemic representations comes from synchronic and diachronic phonological processes, which typically target individual segments, or classes of segments. It is hard to see how phonology could operate without some representations of the affected units. Of course, whether or not a phonemic representation is required depends on what alternate units are posited. Thus, some processes that apply to initial or final consonants could be recast as applying to syllable onsets or codas (CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE). Even in such cases, it may still be necessary to be able to identify individual phonemes, apart from their positions in syllables. As Idsardi (2010) points out, Russian /ivan/ 'Ivan' and /k ɨvanu/ 'to Ivan' have no syllables in common: 'Ivan' is syllabified /i.van/ and 'to Ivan' is syllabified /ki.va.nu/ (cf. Halle and Clements 1983: 149). A representation in which syllables are primitives would have difficulty showing how these words are related.

[8] As mentioned in the previous section, these types of tests are often ambiguous as to whether they target phonemes or just segments. Most of these speech errors show that individual segments can be isolated, but do not necessarily require a phonemic analysis. One interesting example Fromkin (1971: 31) cites is *split pea soup* becoming *plit spea soup*. The fact that *pl* (presumably [pʰl], though Fromkin does not explicitly say) surfaces rather than *bl* when the *s* is transposed could suggest that the speaker groups unaspirated stops following *s* with voiceless stops, rather than with voiced stops.

7 The phoneme in the twenty-first century

As the above survey shows, the phoneme has not disappeared from phonological theory. The fact that recent handbooks of phonology have no chapters devoted to it is not a sign of its demise; rather, it is a function of the development of phonological theory. The time is past when one can attempt to provide an exhaustive definition of the phoneme and its properties apart from elaborating a complete theory of phonology. Many current topics in phonology can be viewed as being about aspects of the phoneme, even though the phoneme is not invoked. For example, the content of the phoneme is studied in distinctive feature theory (CHAPTER 17: DISTINCTIVE FEATURES), feature organization ("geometry"; CHAPTER 27: THE ORGANIZATION OF FEATURES), underspecification (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION), markedness theory (CHAPTER 4: MARKEDNESS) and notions of contrast (CHAPTER 2: CONTRAST). Constraints on the relations between phonemes and phonetics on one side, and lexical representations on the other, are bound up with the question of the organization of the phonological grammar, whether parallel or derivational, or divided into lexical and post-lexical components, and the relation between lexical storage and production and perception.

When one reads the pioneering works of phonology in the late nineteenth and early twentieth centuries, one is struck at their sense of excitement and revelation when discussing the phoneme. This same feeling continues to exist in introductory courses, where the phoneme retains a central place. That phonological theory has subsumed it into more specialized issues and sub-theories does not detract from the fact that it remains, in the words of Krámský (1974: 7), "one of the most magnificent achievements of linguistic science."

ACKNOWLEDGMENTS

I would like to thank Ewan Dunbar, Liisa Duncan, Daniel Currie Hall, Bill Idsardi, and an anonymous reviewer for discussion and suggestions. This work was partially supported by research grant 410-08-2645 from the Social Sciences and Humanities Research Council of Canada.

REFERENCES

Anderson, John M. 2005. Old English *i*-umlaut (for the umpteenth time). *English Language and Linguistics* 9. 195–227.

Anderson, John M. & Colin J. Ewen. 1987. *Principles of dependency phonology*. Cambridge: Cambridge University Press.

Anderson, Stephen R. 1985. *Phonology in the twentieth century: Theories of rules and theories of representations*. Chicago: University of Chicago Press.

Archangeli, Diana. 1984. Underspecification in Yawelmani phonology and morphology. Ph.D. dissertation, MIT. Published 1988, New York: Garland.

Avery, Peter & William J. Idsardi. 2001. Laryngeal dimensions, completion and enhancement. In T. A. Hall (2001), 41–70.

Bagemihl, Bruce. 1995. Language games and related areas. In Goldsmith (1995), 697–712.

Baudouin de Courtenay, Jan. 1972. An attempt at a theory of phonetic alternations. In Edward Stankiewicz (ed. and trans.) *A Baudouin de Courtenay anthology: The beginnings of structural linguistics*, 144–212. Bloomington: Indiana University Press. First published 1895.

Bermúdez-Otero, Ricardo. 2003. The acquisition of phonological opacity. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.) *Variation within Optimality Theory: Proceedings of the Stockholm Workshop*, 25–36. Stockholm: Department of Linguistics, Stockholm University.

Bermúdez-Otero, Ricardo. Forthcoming. *Stratal Optimality Theory*. Oxford: Oxford University Press.

Blake, Barry J. 1979. Pitta-Pitta. In R. M. W. Dixon & Barry J. Blake (eds.) *Handbook of Australian languages*, vol. 1, 183–242. Canberra: Australian National University Press.

Bloch, Bernard. 1941. Phonemic overlapping. *American Speech* 16. 278-284. Reprinted in Joos (1957), 93–96. Also in Makkai (1972), 66–70.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Blumstein, Sheila E. & Kenneth N. Stevens. 1981. Phonetic features and acoustic invariance in speech. *Cognition* 10. 25–32.

Brunner, Karl. 1953. The Old English vowel phonemes. *English Studies* 34. 247–257.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Campbell, Alistair. 1959. *Old English grammar*. Oxford: Clarendon Press.

Carr, Philip, Jacques Durand & Colin J. Ewen (eds.) 2005. *Headhood, elements, specification and contrastivity: Phonological papers in honour of John Anderson*. Amsterdam & Philadelphia: John Benjamins.

Chao, Yuen-Ren. 1934. The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica* 4. 363–397. Reprinted in Joos (1957), 38–54.

Chomsky, Noam. 1964. Current issues in linguistic theory. In Jerry A. Fodor & Jerrold J. Katz (eds.) *The structure of language*, 50–118. Englewood Cliffs, NJ: Prentice-Hall.

Chomsky, Noam. 1980. *Rules and representations*. New York: Columbia University Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 2001. Representational economy in constraint-based phonology. In T. A. Hall (2001), 71–146.

Clements, G. N. 2003. Feature economy in sound systems. *Phonology* 20. 287–333.

Clements, G. N. 2009. The role of features in phonological inventories. In Eric Raimy & Charles Cairns (eds.) *Contemporary views on architecture and representations in phonological theory*, 19–68. Cambridge, MA: MIT Press.

de Lacy, Paul (ed.) 2007. *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press.

Dresher, B. Elan. 1985. *Old English and the theory of phonology*. New York: Garland.

Dresher, B. Elan. 1993. Review of Hogg (1992). *Phonology* 10. 146–157.

Dresher, B. Elan. 2005. Chomsky and Halle's revolution in phonology. In James McGilvray (ed.) *The Cambridge companion to Chomsky*, 102–122. Cambridge: Cambridge University Press.

Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.

Dresher, B. Elan, Glyne L. Piggott & Keren Rice. 1994. Contrast in phonology: Overview. *Toronto Working Papers in Linguistics* 13: iii–xvii.

Fischer-Jørgensen, Eli. 1975. *Trends in phonological theory: A historical introduction*. Copenhagen: Akademisk Forlag.

Fromkin, Victoria A. 1971. The non-anomalous nature of anomalous utterances. *Language* 47. 27–52.

Fromkin, Victoria A. 1973. *Speech errors as linguistic evidence*. The Hague: Mouton.

Gleitman, Lila R. & Paul Rozin. 1977. The structure and acquisition of reading I: Relations between orthographies and the structure of language. In Reber & Scarborough (1977), 1–53.

Godel, Robert. 1957. *Les sources manuscrits du cours de linguistique générale de F. de Saussure*. Geneva: Librairie Droz.
Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.
Gruzdeva, Ekaterina. 1998. *Nivkh*. Munich: Lincom Europa.
Hall, Daniel Currie. 2007. The role and representation of contrast in phonological theory. Ph.D. dissertation, University of Toronto.
Hall, Kathleen Currie. 2009. A probabilistic model of phonological relationships from contrast to allophony. Ph.D. dissertation, Ohio State University.
Hall, T.A. (ed.) 2001. *Distinctive feature theory*. Berlin & New York: Mouton de Gruyter.
Halle, Morris. 1959. *The sound pattern of Russian: A linguistic and acoustical investigation*. The Hague: Mouton.
Halle, Morris & G. N. Clements. 1983. *Problem book in phonology*. Cambridge, MA: MIT Press.
Harris, Zellig S. 1951. *Methods in structural linguistics*. Chicago: University of Chicago Press.
Hockett, Charles F. 1942. A system of descriptive phonology. *Language* 18. 3–21.
Hockett, Charles F. 1959. The stressed syllabics of Old English. *Language* 35. 575–597.
Hogg, Richard M. 1992. *A grammar of Old English*, vol. 1: *Phonology*. Oxford: Blackwell.
Hulst, Harry van der. 1995. Radical CV phonology: The categorial gesture. In Jacques Durand & Francis Katamba (eds.) *Frontiers of phonology: Atoms, structures, derivations*, 80–116. London & New York: Longman.
Hulst, Harry van der. 1996. Radical CV phonology: The segment–syllable connection. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, 333–361. Salford: ESRI.
Hulst, Harry van der. 2005. The molecular structure of phonological segments. In Carr *et al.* (2005), 193–234.
Hyman, Larry M. 2001a. Vowel harmony in Gunu. *Studies in African Linguistics* 30. 147–170.
Hyman, Larry M. 2001b. The limits of phonetic determinism in phonology: *NC revisited. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 141–185. San Diego: Academic Press.
Hyman, Larry M. 2003. "Abstract" vowel harmony in kàlɔ̀ŋ: A system-driven account. In Patrick Sauzet & Anne Zribi-Hertz (eds.) *Typologie des langues d'Afrique et universaux de la grammaire*. 85–112. Paris: L'Harmattan.
Idsardi, William J. 2010. Phonemic segmentation in speech perception – what's the evidence? (comment posted March 14, 2010). *Talking brains: News and views on the neural organization of language*, moderated by Greg Hickok and David Poeppel. Available (July 2010) at http://talkingbrains.blogspot.com/2010/03/phonemic-segmentation-in-speech.html.
Jakobson, Roman. 1962. Phonemic notes on Standard Slovak. In Roman Jakobson, *Selected writings I: Phonological studies*, 221–230. The Hague: Mouton. First published 1931 in Czech in *Slovenská miscellanea (Studies presented to Albert Pražak)*. Bratislava.
Jakobson, Roman. 1971. The Kazan' school of Polish linguistics and its place in the international development of phonology. In Roman Jakobson, *Selected writings II. Word and language*, 394–428. The Hague: Mouton. Translation of Polish original, 1960.
Jakobson, Roman & Morris Halle. 1956. *Fundamentals of language*. The Hague: Mouton.
Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.
Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson & John W. Mullenix (eds.) *Talker variability in speech processing*, 145–165. San Diego: Academic Press.
Jones, Daniel. 1967. *The phoneme: Its nature and use*, 3rd edition (with an Appendix on the history and meaning of the term "phoneme"). Cambridge: Cambridge University Press.
Joos, Martin (ed.) 1957. *Readings in linguistics*, vol. 1. Washington, DC: American Council of Learned Societies.

Kaisse, Ellen M. & Patricia A. Shaw. 1985. On the theory of Lexical Phonology. *Phonology Yearbook* 2. 1–30.

Kazanina, Nina, Colin Phillips & William J. Idsardi. 2006. The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences* 103. 11381–11386.

Kenstowicz, Michael & Charles W. Kisseberth. 1979. *Generative phonology: Description and theory*. New York: Academic Press.

Kiparsky, Paul. 1973. Phonological representations. In Osamu Fujimura (ed.) *Three dimensions of linguistic theory*, 3–136. Tokyo: Taikusha.

Kiparsky, Paul. 1982. From Cyclic Phonology to Lexical Phonology. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part I, 131–175. Dordrecht: Foris.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–365.

Kiparsky, Paul. 2003. Syllables and moras in Arabic. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*, 147–182. Cambridge: Cambridge University Press.

Kiparsky, Paul. Forthcoming. *Paradigms and opacity*. Stanford: CSLI.

Krámský, Jiří. 1974. *The phoneme: Introduction to the history and theories of a concept*. Munich: Wilhelm Fink.

Kuhn, Sherman M. (ed.) 1965. *The Vespasian Psalter*. Ann Arbor: University of Michigan Press.

Langacker, Ronald W. 1988. A usage-based model. In Brygida Rudzka-Ostyn (ed.) *Topics in cognitive linguistics*, 127–161. Amsterdam: John Benjamins.

Liberman, Isabelle Y. 1971. Basic research in speech and lateralization of language: Some implications for reading disability. *Bulletin of the Orton Society* 21. 71–87.

Makkai, Valerie Becker (ed.) 1972. *Phonological theory: Evolution and current practice*. New York: Holt, Rinehart & Winston.

Mohanan, K. P. 1986. *The theory of Lexical Phonology*. Dordrecht: Reidel.

Mompeán, José A. 2006. The phoneme as a basic-level category: Experimental evidence from English. *International Journal of English Studies* 6. 141–172.

Morais, José, Luz Cary, Jésus Alegria & Paul Bertelson. 1979. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* 7. 323–331.

Morén, Bruce. 2003. The parallel structures model of feature geometry. *Working Papers of the Cornell Phonetics Laboratory* 15. 194–270.

Morén, Bruce. 2006. Consonant–vowel interactions in Serbian: Features, representations and constraint interactions. *Lingua* 116. 1198–1244.

Nathan, Geoffrey S. 2006. Is the phoneme usage-based? Some issues. *International Journal of English Studies* 6. 173–194.

Peseisky, David. 1979. Russian morphology and lexical theory. Unpublished ms., MIT.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam & Philadelphia: John Benjamins.

Pike, Kenneth L. 1947. *Phonemics: A technique for reducing languages to writing*. Ann Arbor: University of Michigan.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology*. Dordrecht: Reidel.

Read, Charles, Zhang Yun-Fei, Nie Hong-Yin & Ding Bao-Qing. 1986. The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition* 24. 31–44.

Reber, A. S. & D. L. Scarborough (eds.) 1977. Toward a psychology of reading: The proceedings of the CUNY conference. Hillsdale, NJ: Lawrence Erlbaum.

Rozin, Paul & Lila R. Gleitman. 1977. The structure and acquisition of reading II: The reading process and the acquisition of the alphabetic principle. In Reber & Scarborough (1977), 55–141.

12 Coronals

T. A. Hall

1 Introduction

Numerous studies in the past two decades have argued that coronals like /t/ and /n/ have properties that differentiate them from labials like /p/ and /m/ and dorsals like /k/ and /ŋ/ (e.g. Paradis and Prunet 1991b). For example, in many languages coronals are the only sounds that undergo place assimilation or rules of epenthesis. Sounds like /t/ and /n/ are likewise often the output of rules neutralizing place contrasts. The present chapter summarizes the special properties of coronals and evaluates a number of theoretical and empirical claims made in the literature concerning these sounds.

In (1) I provide some background information for readers unfamiliar with the term "coronal." We can observe here several common IPA symbols for coronal consonants for four manners of articulation. Note that there are four major places, namely labial, coronal, dorsal, and glottal. The present chapter concerns itself with the sounds in the box in (1).

(1)

	labial		coronal				dorsal		glottal
stops	p b		t d		ʈ ɖ	c ɟ	k g		ʔ
fricatives		f v	s z	ʃ ʒ	ʂ ʐ		x ɣ	χ ʁ	h
nasals	m		n		ɳ	ɲ	ŋ		
liquids			l r						

From the point of view of phonetics, coronal consonants include at least the following four places of articulation: (denti-)alveolar (e.g. /t d/), post-alveolar (e.g. /ʃ ʒ/), retroflex (e.g. /ʈ ɖ/) and palatal (e.g. /c ɟ/).[1] From the perspective of

[1] Coronals also uncontroversially include (inter)dentals like /θ ð/. A lesser-known coronal place of articulation is alveolo-palatal (e.g. /ɕ ʑ/, as in Polish and Mandarin). Palatal stops like /c ɟ/ are considered to be coronal, although there is some controversy concerning whether or not palatal fricatives like /ç ʝ/ are coronal or dorsal (or both). For various views on this issue the reader is referred to Hume (1992), Hall (1997), and Robinson (2001). In addition to the coronal consonants like the ones in (1), a number of linguists have argued that front vowels like /i e/ are coronal, e.g. Clements (1976), Hume (1992), and Clements and Hume (1995). I use the term "coronal" in this chapter to refer only to consonants.

phonology there is general consensus that the boxed sounds in (1) form a natural class that is captured with the privative feature [Coronal]. Within that class, it is often useful to draw a distinction between "anterior coronals" like /t d s z n l/ and "non-anterior coronals" like /ʈ ɖ c ʃ ʒ ʂ ʐ ɳ ɲ/.[2]

As noted above, a number of studies have argued that coronals like the ones in (1) display a unique set of properties not shared by labials or dorsals. Put simply, the properties of sounds like /t/ have led many linguists to the conclusion that coronal is the least marked major place of articulation (e.g. Kean 1975; Paradis and Prunet 1991a). By contrast, much recent work has argued that the unmarked major place is glottal (e.g. Lombardi 2002; de Lacy 2006), dorsal (e.g. Trigo 1988, who uses the term "velar") or even labial (Hume and Tserdanelis 2002; see also CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION). In light of these apparently conflicting claims, it is natural to ask whether or not coronals are indeed endowed with special properties, or if coronal is simply another place on par with dorsal and labial (Blevins 2004). It is the purpose of the present chapter to evaluate the special status of coronals and to show where the areas of controversy lie.

The special status of coronals is clearly connected to the question of how markedness is defined in general and in phonology in particular (see Hume 2003, de Lacy 2006, and Rice 2007 for recent discussion). Rice draws a distinction between properties of "natural markedness" and those of "structural markedness," where the former term refers specifically to non-phonological criteria, e.g. frequency, acquisition, and saliency, and the latter term to phonological criteria.

The present chapter focuses only on phonological arguments for markedness (i.e. structural markedness). According to Rice (2007), three of these criteria are assimilation, neutralization, and epenthesis. This means that coronals like /t/ are unmarked if they undergo place assimilation, if they are the output of place neutralization, or if they can be epenthesized.[3]

The remainder of this chapter is organized as follows. In §2 I summarize some of the general questions concerning the status of coronals that are currently in dispute. All of these questions will be referred to extensively in the following sections. The following three sections concentrate on the role of coronals with respect to the three properties of structural markedness referred to above, namely assimilation (§3), epenthesis (§4), and place neutralization (§5). In §6 and §7 I consider two additional properties of structural markedness, namely the role of coronals in syllable structure and the transparency of coronals in place spreading. In §8 I provide a brief summary of each of the main theoretical and empirical issues addressed in the preceding sections.

[2] From the point of view of articulatory phonetics, coronals are usually defined as sounds produced with the tip blade (including the tip) of the tongue (see Keating 1991: 30). I adopt the definition proposed by Halle and Stevens (1979: 346), according to which coronal sounds involve the raising of the "central" portion of the tongue, i.e. the part of the tongue connecting the blade with the tongue body. In referring to the "central" part of the tongue, this revised definition is therefore able to include palatals as coronal sounds. From the point of view of acoustic phonetics, coronals are characterized by a concentration of energy in the upper frequencies of the spectrum (Jakobson *et al.* 1952).

[3] For reasons of space, I do not discuss the other two diagnostics for structural markedness mentioned by Rice (2007), namely coalescence and deletion. Put simply, it is typically assumed that unmarked features are more prone to coalescence or deletion than marked features. A property of structural markedness not mentioned by Rice (2007) is dissimilation. For an argument that coronals are unmarked in place dissimilation in Tashlhiyt Berber the reader is referred to Alderete (2004: 400ff.).

2 General research questions

In the following sections I discuss the five very general questions in (2) pertaining to coronals, all of which are open to debate.

(2) a. Markedness: Are coronals endowed with special properties that differentiate them from the other major places and, if so, what are they?
b. Universality: Are the properties characterizing coronals (or other places of articulation) universal or language-specific?
c. Representations: To what extent do the properties of coronals (or other places of articulation) require unique phonological representations?
d. Manner: Why do unmarked coronals belong to certain manner categories and not to others?
e. Coronal subplaces: Why do unmarked coronals belong to certain subplaces and not to others?

While the unmarkedness of coronals (i.e. (2a)) is the position usually taken in the literature, this question is still a controversial one. It also needs to be clarified whether or not the special properties for coronals can vary within a single language. For example, if coronals are unmarked relative to labials and dorsals in Language A with respect to assimilation, will they also be unmarked in Language A with respect to epenthesis or neutralization? If not, how does one account for this asymmetry? One of the most controversial questions concerning markedness in linguistics pertains to universality (i.e. (2b)): are coronals less marked than labials and dorsals in all languages without exception? It is sometimes assumed that this is the case, but much current work has questioned this assumption. If the unique properties of coronals are language-specific, does this imply that there are no constraints on possible targets for place assimilation and outputs of place neutralization?

Regardless of what one's answer is to the question of markedness and universality, one needs to provide a formal analysis for the unmarkedness of coronals with respect to other major places of articulation in at least some languages (i.e. (2c)). Many phonologists argue that the special status of coronals in such languages requires that they have a particular representation that makes them unique from other major places of articulation. By contrast, much recent work in Optimality Theory (OT) has denied the role of representations. It will be clear below that many adherents of OT argue that the special properties of coronals (i.e. their unmarkedness) should be captured not with representations, but instead with a markedness hierarchy. Is the rejection of unique representations for coronals warranted?

It will be clear below that for those languages in which coronals are unmarked, it is only certain sounds within that class that display these properties. For example, many languages have place assimilations that affect coronals, but these assimilations will only apply to nasals but not to fricatives (i.e. (2d)), and only to alveolars and not to palatals (i.e. (2e)). What kind of cross-linguistic variation is attested with respect to manner and coronal subplaces? And how should generalizations concerning manner and coronal subplaces be captured theoretically?

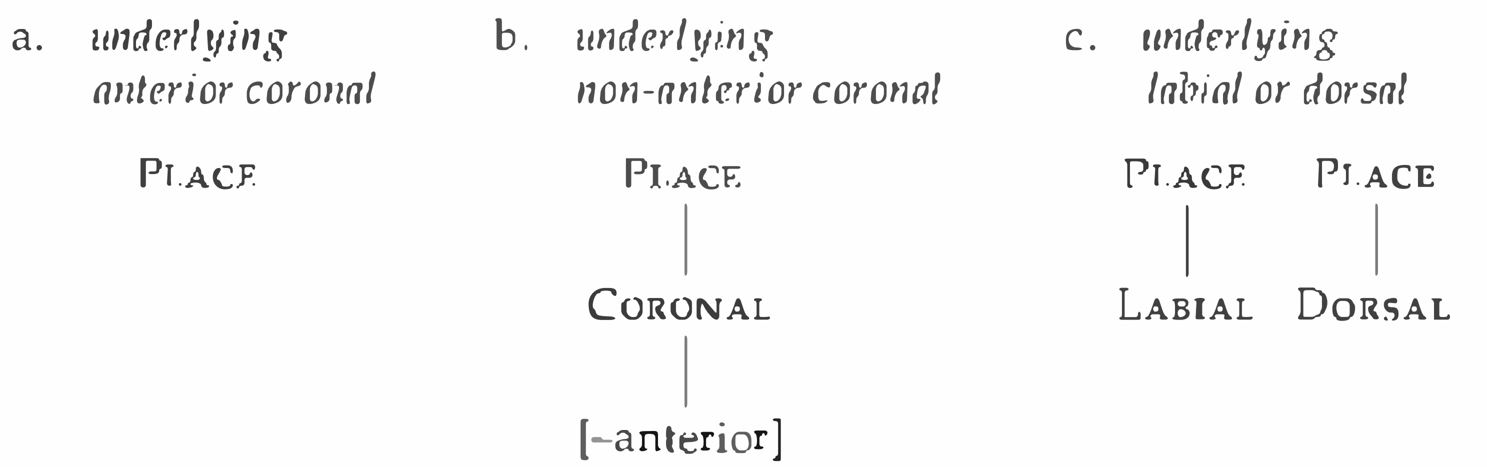
3 Coronals as the target for place assimilation

A number of linguists have argued that sounds like /n/ and /t/ are more prone to place assimilation than the corresponding labials or dorsals (e.g. Kiparsky 1985; Avery and Rice 1989). Consider the Catalan data in (3) below from Kiparsky (1985: 95), whose source is Mascaró (1976).

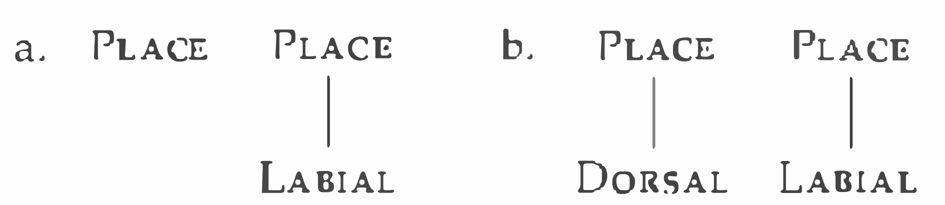
(3) a. *so*[n] *amics* 'they are friends'
so[m] *pocs* 'they are few'
so[ŋ] *grans* 'they are big'
b. *so*[m] *dos* 'we are two'
ti[ŋ] *pa* 'I have bread'
a[ɲ] *feliç* 'happy year'

In (3a) we can observe in the second and third example that the /n/ in the morpheme 'they' assimilates to the place of articulation of a following labial or dorsal. By contrast, none of the other three underlying nasals (i.e. /m ŋ ɲ/) assimilates to the place features [Labial] or [Dorsal]; see (3b).

One way of capturing the special status of /n/ is to posit a representation for that segment which is underspecified for all place features (e.g. Kiparsky 1985: 97–99). This idea is captured in the representations in (4), where (4a) is the structure for /n/ and (4c) for labials and dorsals, and (4b) for non-anterior coronals like /ɲ/. I refer to (4) henceforth as the "coronal underspecification" approach, although it should be clear from (4) that it is only anterior coronals that are underspecified.[4]

(4)


An important assumption made in the coronal underspecification approach is that nasal assimilation in Catalan involves the spreading of a feature to a segment lacking that feature. Given the structure-building nature of nasal assimilation, the rule can spread [Labial] to the left in (5a) but not in (5b).

(5)


[4] Kiparsky (1985) assumes that palatals like /ɲ/ are dorsal, but I follow the consensus view that they are coronal. His representations differ slightly from the ones in (4) because he does not assume the existence of a [Place] node. Structures like the ones in (4) are endorsed by several of the authors in Paradis and Prunet (1991b), although opinions differ on whether or not anterior coronals like /t/ and /n/ lack [Coronal] as in (4a), or [Place] as well as [Coronal].

The model in (7) can account for the pattern of assimilation in Korean in (6). Unmarked coronals like /t/ assimilate to a following labial or dorsal. Significantly, labials can assimilate to dorsals (by the spreading of [DORSAL]), but dorsals cannot serve as the targets of place assimilation.

Mohanan (1991), McCarthy and Taub (1992) and Hall (1995) all point out a potential problem with the coronal underspecification approaches in (4). It will be clear below that these criticisms also hold for (7). In many of the languages in which sounds like /n/ are argued to be underspecified for place features, we also find rules that crucially require /n/ to be marked for place features. For example, like Catalan, English and German /n/ assimilates to a following labial or dorsal. Since this assimilation is post-lexical, it would be necessary to analyze /n/ as underspecified for place features at that late stage in the derivation. The problem is that there are lexical rules (or static constraints) that require that coronal sounds like /n/ be specified for place features. For example, in German there is a lexical rule of schwa epenthesis that inserts a schwa only between a coronal stop and a coronal obstruent in verbs (e.g. *arbeit-et* [aɐbaitət] 'works' *vs. lern-t* [lɛɐnt] 'learns'). Clearly, the rule of schwa epenthesis can only go into effect if sounds like /t/ are [CORONAL]. In American English, initial coronal consonant plus [ju] is prohibited. Thus the constraint banning '[CORONAL] [ju]' sequences must make reference to the feature [CORONAL].

The pattern of place assimilation in Sri Lankan Portuguese Creole is problematic for both of the representational models discussed above because labials and dorsals but not coronals serve as the target. Consider the data from that language (from Hume and Tserdanelis 2002) in (8), which illustrate the pattern of place assimilation across morpheme boundaries:

(8)		*nom. sg.*	*gen. sg.*	*dat. sg.*	*verbal noun*	
	a.	[vaːrzim]	[vaːrzin-su]	[vaːrzim-pə]	[vaːrsiŋ-ki]	'harvest'
		[maːm]	[maːn-su]	[maːm-pə]	[maːŋ-ki]	'hand'
		[miːtiŋ]	[miːtin-su]	[miːtim-pə]	[miːtiŋ-ki]	'meeting'
	b.	[bataan]	[bataan-su]	[bataan-pə]	[bataan-ki]	'button'
		[siːn]	[siːn-su]	[siːn-pə]	[siːn-ki]	'bell'

In the first column of (8a) we can observe monomorphemic words in the nominative singular ending in a labial or dorsal nasal. Before the genitive singular suffix /-su/ in the second column, both the labial and dorsal assimilate to the place of articulation of the following /s/. The third column is important because it illustrates that the dorsal nasal assimilates to the place of the following /p/. Finally, we can see in the fourth column that /m/ assimilates to dorsal before /k/. The data in (8a) can be contrasted with the examples in (8b), which show that the coronal nasal /n/ resists place assimilation to both a labial and a dorsal.

What the data in (8) tell us is that in Sri Lankan Portuguese Creole, labials and dorsals undergo place assimilation but that coronals do not. Hume and Tserdanelis (2002) argue that the pattern of assimilation in (8) poses a serious problem for the coronal underspecification approach in (4). The reason is that coronal underspecification presupposes that only the marked features (e.g. the ones that resist place assimilation) are underlyingly specified for place features, while the unmarked features (e.g. the ones that undergo place assimilation) are the ones a list of other languages in which palatals fail to undergo place assimilation, but impressionistically, this generally appears to be the case.[7] If the constraint ranking approach taken in (8) were to be extended to include the constraints referring to the subplaces within the coronal range, then one would expect there to be a language in which the palatal stop /c/ or the retroflex stop /ʈ/ undergoes place assimilation, while the alveolar stop /t/ does not. To the best of my knowledge, such a pattern of assimilation is not attested, but the question of whether or not "everything goes" among coronal subplaces clearly needs to be left open for further research.

4 Coronals as epenthetic segments

There is general agreement in the literature that epenthesis is a valid markedness diagnostic. Glottals ([ʔ h]) are the most common epenthetic consonants from a cross-linguistic point of view, but coronals have been argued to be less marked than labials or dorsals in terms of epenthesis. For example, in their defense of the claim that coronal is the least marked place, Paradis and Prunet (1991a: 20) write: "In initial and intervocalic positions, coronals (including fricatives) are inserted more often than noncoronals." The language most often cited with coronal epenthesis (i.e. the unmarked voiceless stop /t/) is Axininca Campa, although sometimes other languages are cited as well (e.g. Gokana and Amharic).

Consider first languages with epenthetic glottals (i.e. [ʔ]). According to the survey provided in Lombardi (2002), there are languages attested in which [ʔ] is epenthesized initially (presumably to avoid vowel-initial syllables), medially (as a means of resolving hiatus), or finally. In many languages with an epenthetic glottal, the sound being epenthesized is not phonemic, but there are languages attested in which a glottal is epenthesized even if that sound belongs to the underlying consonants (e.g. Selayarese; Lombardi 2002). Some of the languages with epenthetic glottal include Arabic (word-initial) and Selayarese (between identical vowels). The coda context for glottal epenthesis is illustrated in the Cupeño examples in (13) from Lombardi (2002: 229), who cites Crowhurst (1994).

(13)	/ʧi/	[ʧiʔ]	'gather'
	/kʷa/	[kʷaʔ]	'eat'

According to the sources cited above, epenthesis of a consonant in final position in Cupeño is necessary to satisfy a bimoraic minimal word requirement. The reason [ʔ] is epenthesized, as opposed to some other consonant, is that glottal is seen as the least marked place. The unmarkedness of place of articulation is accomplished with constraints similar in function to the Preserve constraints discussed in (9)–(11) above. In particular, the constraints penalize the relevant major place category (i.e. *Lab, *Cor, *Dors, *Phar). Lombardi (2002) assumes

7 Hungarian is yet another example of a language in which palatals fail to undergo place assimilation (Siptár and Törkenczy 2000: 207). Interestingly, these authors remark that Hungarian /ɲ/ assimilates to [ŋ] before dorsals in certain dialects (2000: 211). In the literature on Korean it is usually assumed that "palatal" /c/ assimilates to a following dorsal (i.e. /ck/ → [kk]), but not to a following labial; however, an anonymous reviewer points out that Korean /c/ is closer to a strident affricate (IPA /tɕ/) than a palatal stop.

that the constraint penalizing glottals is *Phar, because in her system glottals have pharyngeal place features (i.e. [Phar, +glottal]). Lombardi (2002) also operates under the assumption that glottals are obstruents and not sonorants. As I point out below, the treatment of glottals as obstruents is not universally accepted.

Lombardi's (2002) markedness hierarchy is presented in (14). In contrast to the markedness hierarchy posited in the previous section, the one in (14) is assumed to be universally fixed. See de Lacy (2006) for a similar markedness hierarchy.

(14) *Dors, *Lab >> *Cor >> *Phar

According to the hierarchy in (14), glottals are the optimal epenthetic consonants because their place markedness is even lower than that of the relatively unmarked coronals. Consider the evaluation of the first example in (13), given the hierarchy in (14):

(15)

Input: /ʧi/	MinWd	*Dors	*Lab	*Cor	*Phar
a. [ʧi]	*!				
☞ b. [ʧiʔ]					*
c. [ʧit]				*!	
d. [ʧik]		*!			
e. [ʧip]			*!		

In this tableau, candidate (15b) is selected over candidates (15c–15e) because the latter three insert consonants that have a place of articulation other than glottal.

In other languages a coronal is epenthetic. The intuition behind the ranking in (14) is that coronal is the next best place in terms of epenthesis. Thus, if there is some independent reason for why a glottal cannot be epenthesized, then a coronal (and not a labial or a dorsal) will be.

Consider now the Axininca Campa examples in (16) from Lombardi (2002), who cites McCarthy and Prince (1993) and Payne (1981). This example is usually considered to be a paradigm case for a language with an epenthetic [t]. /N/ represents a nasal consonant unspecified for place features.

(16) /i-N-koma-i/ [iŋkomati] 'he will paddle'
/i-N-koma-aa-i/ [iŋkomataati] 'he will paddle again'

These examples illustrate that [t] is epenthesized intervocalically in order to ensure that every syllable has an onset. Given the markedness hierarchy in (14), the challenge posed by these examples is that they require that a coronal rather than a glottal be epenthesized.

Both Lombardi (2002) and de Lacy (2006) argue that a coronal, rather than a glottal, is epenthesized for reasons of sonority. The generalization is that onsets with a stop are more harmonic than those with a fricative, nasal, or liquid, because stops are the least sonorous segment type on the sonority hierarchy. Assuming that de Lacy (2006) is correct in his analysis of glottals as sonorants, [t] is selected over glottals because the latter make worse onsets than the former. This intuition is captured in the following tableau:

For example, de Lacy (2006) discusses languages with an epenthetic rhotic [r], e.g. English and Southern Tati, and shows that this process applies only in the context of low vowels. Consider the latter language. In Southern Tati the approximant [r] is epenthesized after low vowels, but after non-low (unrounded) vowels (e.g. /i e/) the epenthetic segment is [j], and after non-low (rounded) vowels (e.g. /y u/) it is [w]. In this language, epenthetic consonants copy the place and manner features of the preceding vowel; hence, [j] appears after /i e/ because all of these sounds have the same value of [sonorant] and [approximant] (and the place feature [Coronal]). The sound [w] likewise appears after sounds like /y u/, because all of these segments share the same manner and place features. After /a/ the epenthetic consonant is the approximant [r], because the latter sound copies the preceding vowel's manner features (i.e. [sonorant] and [approximant]). Significantly, the epenthetic consonant cannot copy the preceding vowel's place features because /a/ is neither [Labial], [Coronal], nor [Dorsal] (de Lacy 2006: 101–102, after Clements and Hume 1995).

Since there appear to be no languages in which a non-anterior coronal like [j] can be freely epenthesized regardless of the quality of an adjacent vowel, I conclude that the approach described by de Lacy (2006) makes the correct predictions. This approach could potentially be falsified by a language in which [j] (or some other non-anterior coronal) is epenthesized in the context of any vowel (and not simply in the neighborhood of /i e/). Whether or not there are such languages is a question that clearly needs to be left open for further study.

5 Coronals as the output of place neutralization

Many languages are attested in which a multiple place of articulation contrast is neutralized in coda position, but these languages can vary concerning the output sound. Some languages have been argued to involve a place neutralization to coronal, thereby providing evidence for the claim that coronal in these languages is less marked than labials or dorsals. Some of the languages cited with a place neutralization to coronal are Finnish (Yip 1991), Greek, Italian, Croatian dialects, as well as Fante (Kiparsky 1995: 669), Castilian Spanish, and Koyukon (Rice 1996). Many well-known languages are attested in which the output of place assimilation is glottal, i.e. debuccalizations converting sounds like /s/ to [h] (in dialects of Spanish) or /t/ to [ʔ] (in dialects of English). More interesting for purposes of the present chapter are languages reported to neutralize place contrasts to dorsal (Selayarese) or labial (Manam).

Languages referred to above which debuccalize sounds like /s/ or /t/ to glottal are unproblematic given the markedness hierarchy in (14). Given an input like /tat/, the most harmonic output is [taʔ], because the candidates [tak], [tap] and [tat] incur violations of the high-ranking markedness constraints *Dors, *Lab and *Cor. If the faithfulness constraint militating against the change from /t/ to some other sound is ranked lower than *Phar, then the output [taʔ] will be selected as optimal. An approach to neutralizations that accepts coronal underspecification would need to explain why the output of /tat/ is [taʔ] and not [tat]: since debuccalization is usually assumed to delete all and only place features, one might expect the output to be a coronal if coronals lack place features. However, the problem might be solvable if glottals are analyzed as having even less structure than

In Rice's (1996) representational approach, final nasals in Selayarese – like the final coronals in languages like Finnish – have the representation in (7a), which lacks [Place]. In Finnish a default rule applies specifying that consonants unspecified for [Place] are [Coronal]. Unlike Finnish, Selayarese does not have the default rule assigning [Coronal] to placeless consonants, and thus the consonant that lacks [Place] is interpreted phonetically as velar.

Languages that neutralize place of articulation to dorsal can be captured in an OT approach with a freely rankable markedness hierarchy. The correct output can be achieved by ranking the constraint militating against dorsals lower than the constraints militating against coronals or labials. By contrast, in approaches to OT with a fixed markedness hierarchy (as in (14)), it is not clear how the examples in (20) can be accounted for. De Lacy (2006) ultimately argues that place neutralizations can have only two possible outputs, namely coronal and glottal. He argues that the cases of neutralization to dorsal can be explained in other ways; recall from §4 that he re-analyzes dorsals that are apparently epenthetic as being morphological in nature. In the case of Selayarese in (20) (and other languages), de Lacy (2006) argues that the dorsal nasal is really glottal. In Oceanic languages the neutralization of /t/ to [k] is a common historical process, but de Lacy (2006) claims that there are no languages in which such neutralizations can be motivated as synchronic processes on the basis of alternations. Whether or not de Lacy's (2006) re-analysis of apparent cases of neutralization to dorsal can be extended to other languages is – in this writer's view – doubtful. See also Rice (2008) for criticisms. Rice also points to Manam as a language that neutralizes a multiple place contrast to labial, but it is not clear how these data can be accounted for with the fixed markedness hierarchy in (14).

Not much attention seems to have been paid in the literature to questions (2d) and (2e). Given that there are languages in which the output of place neutralization is a coronal, why is the subplace anterior and why is the manner a stop or nasal? Could a language with /p t c k/ neutralize the four-way contrast to the palatal stop [c]? While it is conceivable that such a neutralization might occur in the neighborhood of a front vowel (since front vowels are often assumed to be [Coronal] and (redundantly) [−anterior]), there appear to be no attested cases of this type of a neutralization that are not assimilatory. Could a contrast among fricatives /f s x/ be neutralized to [s]? Again, this type of neutralization appears not to be attested.

6 Coronals and syllable structure

A number of linguists have observed that there are languages in which coronal obstruents have a privileged position in syllable structure in the sense that they can occur in contexts where labial and dorsal obstruents cannot. Examples like the ones to be discussed below can be thought of as a particular type of neutralization to coronal, although they do not involve active neutralizations like the one discussed above for Pawnee. In this section I discuss a familiar case study from German and English syllable structure illustrating the special status of coronals.

Many writers have made the following observation for German and English: if a syllable ends in a sequence of two obstruents and if the first one is a stop, then the second one must be a coronal obstruent (see Hall 2002 and references representation as in (22a). Thus, on this view, all words are fully syllabified, as in (22b). In order to explain the apparent SSG violations in (21), Hall restates the SSG as follows: "A sonorant consonant in the syllable coda may only be preceded by segments of higher sonority." Given this constraint (which Hall refers to as SON), the final sequences in (21) have a well-formed sonority profile because SON does not refer to sequences of adjacent obstruents.

One way of capturing the fact that the underlined sounds in (21) are coronal is to adopt the following context-sensitive markedness constraints (modified slightly from Hall 2002: 47):

(23) *Context-sensitive markedness constraints*

a. *TP/TK: $\begin{bmatrix} -\text{son} \\ -\text{cont} \\ \text{CORONAL} \end{bmatrix} \begin{bmatrix} -\text{son} \\ \text{PERIPHERAL} \end{bmatrix}]_\sigma$

b. *PK/KP: *[PERIPHERAL] [PERIPHERAL]]$_\sigma$

The constraint *TP/TK in (23a) is necessary to rule out sequences like [atk] and [atp], and the constraint *PK/KP in (23b) likewise rules out sequences like [apk] and [akp]. Note that (23a) and (23b) cannot be combined into a single constraint (i.e. *[−son, −cont] [PERIPHERAL]]$_\sigma$), because the first segment referred to in (23a) is a stop only (cf. [sk] clusters, e.g. *ask*), but the two segments referred to in (23b) can be any peripheral segment.

The constraints in (23) are advantageous because they are also able to account for the fact that sequences like [tk tp] are repaired in other languages by processes like epenthesis, deletion, or metathesis. A typical example is discussed by Kenstowicz (1994: 300), who shows that in Lebanese Arabic word-final sequences like /tk db/ are split up by vowel epenthesis (see (24b)), but the mirror image sequences /kt bd/ are not (see (24a)). Lebanese Arabic also provides evidence that sequences of two peripheral obstruents are marked as well, since they are separated by an epenthetic vowel (see (24c)).

(24)	a.	/sakt/	[sakt]	'act'
		/ʕabd/	[ʕabd]	'slave'
	b.	/fatk/	[fatik]	'eradicating'
		/nadb/	[nadib]	'wailing'
	c.	/ʕakf/	[ʕakif]	'leaning to the letter'
		/rikb/	[rikib]	'riding'
		/ħabk/	[ħabik]	'weaving'

Consider now the data in (25) from German and English, which illustrate a slightly different generalization than the ones in (21): after VXC, where X = a short vowel or a consonant, i.e. after a short vowel plus two consonants, a long vowel plus one consonant or a diphthong plus one consonant, the only type of consonant that can follow is a coronal obstruent. In (25a) the coronal and the preceding consonant are tautomorphemic and in (25b) they are heteromorphemic. Generalizations similar to the ones in final position also hold in initial position as well, but I concentrate below only on data like the ones in (25). As in (21), the word-edge sounds in (25) subsume anterior coronals and non-anterior coronals.

In at least three languages (Guere, Fula, and Mau), vowel harmony has been shown to apply across coronals but not across labials and dorsals (or non-anterior coronals), thereby implying that coronals are less marked than the latter two major places (Paradis and Prunet 1989). Paradis and Prunet argue that the aforementioned languages require coronal underspecification: only coronals lack [Place] (and hence [Coronal]), but labials and dorsals are specified for [Place] and their dependent articulator features [Dorsal]/[Labial]. Given representations like the ones described, only coronals should allow vowel harmony to apply across them, because vowel harmony is argued to spread [Place] (and the dependent feature [Dorsal]). Vowel harmony is blocked when a dorsal or a labial (or non-anterior coronal) intervenes, because these segments all have [Place], which inhibits spreading. A potential problem for this representational approach would be a language in which labials or dorsals but not coronals are transparent to spreading. Paradis and Prunet (1991c, 1994) discuss apparent examples of languages with transparent dorsals (e.g. Choctaw, Chinook, Ennemor) and re-analyze the data in ways that are compatible with coronal underspecification. A question that clearly needs to remain open is whether or not the same strategy holds for all languages with transparent dorsals. A second question is whether or not there are languages in which labials are transparent, but not coronals or dorsals. Finally, it remains to be seen how an OT analysis might account for languages with transparent coronals.

8 Conclusion

The present chapter has examined a number of theoretical and empirical issues with respect to the special status of coronals. Emphasis was placed on the behavior of coronals from the point of view of structural markedness, in particular epenthesis, assimilation, neutralization, syllable structure, and transparency. In (2) I posited five general research questions, which I repeat in (26) for convenience:

(26) a. Markedness: Are coronals endowed with special properties that differentiate them from the other major places and, if so, what are they?
  b. Universality: Are the properties characterizing coronals (or other places of articulation) universal or language-specific?
  c. Representations: To what extent do the properties of coronals (or other places of articulation) require unique phonological representations?
  d. Manner: Why do unmarked coronals belong to certain manner categories and not to others?
  e. Coronal subplaces: Why do unmarked coronals belong to certain subplaces and not to others?

Concerning point (26a), the examples discussed in the preceding sections have demonstrated that there are undeniably languages in which coronals have a special status, e.g. languages like Catalan, in which only /n/ and not other nasals undergo place assimilation. However, one cannot conclude that coronals are unique in all languages (i.e. (26b)). I draw this conclusion in light of some of the examples discussed above. For example, in Sri Lankan Portuguese Creole labials are unmarked with respect to place of assimilation. It also appears to be the case

13 The Stricture Features

Ellen M. Kaisse

The late 1980s and early 1990s saw the publication of a large number of articles treating features and feature geometry, following on the highly influential article by Clements (1985) that had argued for the organization of features into a hierarchically arranged tree. But if one takes a look at the features whose position and organization were so often debated, they fall into only a few groups: first, the vocalic features typically involved in vowel harmony – features of height, backness, rounding, or tongue root position; second, laryngeal features such as glottalization, voicing, and aspiration; and third, the place features labial, dorsal, and coronal and their dependents, especially those of coronal (see CHAPTER 27: THE ORGANIZATION OF FEATURES; CHAPTER 21: VOWEL HEIGHT; CHAPTER 19: VOWEL PLACE; CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION; CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION; CHAPTER 12: CORONALS). Such a concentration can be traced back in part to the material covered in Clements's seminal article. But it also reflects those features that seem most often and most unequivocally involved in assimilation, or, to follow McCarthy's (1988) diagnostics, the trio of assimilation, dissimilation, and reduction or deletion. There was also notable interest in the position and behavior of [lateral] (CHAPTER 81: LOCAL ASSIMILATION; CHAPTER 60: DISSIMILATION; CHAPTER 79: REDUCTION; CHAPTER 31: LATERAL CONSONANTS). But most of the traditional manner and major class features – [continuant], [consonantal], [sonorant] – received short shrift.[1] These three, along with the less widely adopted feature [approximant], are sometimes called stricture features (Kenstowicz 1994: 480ff.), because they specify how the articulators are brought together but are not themselves inherently associated with any one articulator. They are "articulator-free," in the terminology of Clements and

[1] Chomsky and Halle (1968: 299–300) distinguish five kinds of features: major class, cavity, manner of articulation, source, and prosodic. The major class features are [sonorant] and [consonantal] (and the quickly eliminated [vocalic]). The manner features are [continuant], [delayed release], [tense], and several features related to suction and ejection. Major class features, they explain, mainly have to do with closing and opening and with pressure buildup and release. While [continuant] was not categorized as a major class feature by Chomsky and Halle, it is the one manner feature that strongly relates to how closely the articulators are approximated. Padgett (1991) and Kenstowicz (1994), amongst others, thus group [consonantal], [sonorant], and [continuant] together as stricture features.

Hume (1995) and must be associated with some articulator to be fully interpreted. The major class features [consonantal] and [sonorant] do not participate often, if at all, in the archetypical phonological processes of spreading and delinking, and they do not form nice bundles of features that beg to be organized under a node of the feature geometry. The other major class features of Chomsky and Halle (1968), [syllabic] and [vocalic] (the latter already tentatively withdrawn in a footnote in Chomsky and Halle 1968: 302), have over the decades been subsumed in representations, with [syllabic] coming to mean head or nucleus of the syllable and not being specified as a distinctive feature of any segment.[2] As early as 1974, Hankamer and Aissen suggested that the major class features might be superfluous, and in a more elaborated proposal, Selkirk (1984) proposed doing away with both [consonantal] and [sonorant] by replacing them with a multivalued sonorancy feature (see also CHAPTER 8: SONORANTS). Nonetheless, the consensus view seems to be that these features are still needed. Their necessity is taken for granted in many significant works on feature theory of the last decade or two, such as Clements (2003), to mention one representative example. As an independent issue, the proper definition of [continuant] has always been problematic: it has not been obvious exactly what natural classes should be delineated by this feature, nor how it should be placed on the feature tree (see also CHAPTER 28: THE REPRESENTATION OF FRICATIVES).

In this chapter I will review some of the proposals to keep, redefine, or locate on the feature tree the problematic stricture features [consonantal], [sonorant], and [continuant], or to do away with them altogether. In the spirit of the *Companion*, we will spend most of the time looking at the phonological and phonetic phenomena upon which these various proposals rested. The chapter is organized as follows. We begin with a discussion of the arguments for and against the existence of [consonantal]. In the second section, we turn a similar spotlight upon [sonorant]. The third section considers arguments that bear on the need for both features, hinging on the sonority sequencing principle, weight, and lenition. The final section treats the definition and feature geometry of [continuant].

1 The major class feature [consonantal]

Chomsky and Halle (1968: 301–302) distinguished a group of binary "major class features" that segregate segments into the large groups of consonants and vowels, sonorants and obstruents, glides and liquids, and syllabic and non-syllabic sounds. The binary feature [consonantal] makes the most obvious cut among segments, capturing the distinction between vowels and consonants.

> Consonantal sounds are produced with a radical obstruction in the midsagittal region of the vocal tract; nonconsonantal sounds are produced without such an obstruction. (Chomsky and Halle 1968: 302)

[2] Padgett (2008) argues that [vocalic] should be reinstated to describe the difference between vowels, which have no frication, and the type of glide that shows some frication. For Padgett, both [consonantal] and [vocalic] should be employed so as to capture a fine hierarchy of degree of stricture.

(1) a. *Cypriot Greek consonantalization*
The glide [j] becomes a voiceless palatal stop [c] after a consonant.
b. *Bergüner Romansh consonantalization*
The glides [j] and [w] are realized as the voiced stop [g] before a consonant.
c. *Uyghur vowel consonantalization*
Devoiced high vowels become syllabic fricatives before a consonant. /i/ is realized as [ʃ̩], /u/ as [ɸ̩ʷ] and /y/ as [ɸ̩ᶣ].
d. *Ahtna vocalization*
Certain consonants become vowels before a consonant. /b/ and /χʷ/ are realized as [u], while /ɣ/ and /c/ appear as [i].
e. *Halland Swedish vocalization*
The uvular fricative /ʁ/ is realized as the offglide [ɑ̯] before any consonant.

Kaisse (1996) adds the debuccalization of [s] to [h] in Argentinian Spanish, which only takes place when a coda /s/ is preceded by a vowel or glide. Thus the rule appears to be a rightward spreading of the feature [−consonantal].

Let us consider the first two examples, Modern Cypriot Greek and the Bergüner dialect of Romansh, in a little more detail. These two have mirror-image environments that make it unlikely that alternative analyses will succeed by identifying a specific position in the syllable, be it onset or coda, as the cause for consonantalization. On the other hand, Cypriot also shows the sorts of complications that offer a wedge to alternative analyses that do not use the spread of [consonantal].

Newton (1972a) describes a defining characteristic of all Cypriot dialects: what would appear as the glide [j] in other dialects is realized as a voiceless palatal stop after most consonants or as a voiceless velar stop after [r]. Numerous alternations in Cypriot support this consonantalization as a productive process. (The cognate form in non-consonantalizing dialects is the intermediate form shown to the right of the first arrow in the examples below.)

(2)	a.	aðerfi	'brother'
		/aðerfi+a/ → aðerfja → aðerfca	'brothers'
	b.	vari	'heavy'
		/vari+ume/ → varjume → varkume	'I am bored'
	c.	mati	'eye'
		/mati+a/ → matja → matca *or* maθca	'eyes'

This last example shows the continuancy adjustment that complicates the phenomenon, and which Kaisse regards as orthogonal to it. In virtually all dialects of Modern Greek (Newton 1972b), sequences of obstruents are adjusted so that the first is a fricative and the second a stop (see also CHAPTER 28: THE REPRESENTATION OF FRICATIVES). Though spread of [consonantal] seems the most obvious approach to the Cypriot alternation, J. Harris (1996) points out that the case is not straightforward, since we must devoice and obstruentize the glide, not merely turn it into a consonant.

Bergüner Romansh (Kamprath 1986) displays a process of "glide hardening" that is practically the mirror image of the Cypriot case: glides become velar stops before any consonant. Unlike Greek, Romansh has a [w] as well as a [j]. Both glides become [g], which may then devoice before voiceless consonants.

(3) a. kreja /krej+a/ 'believes'
krekr̩ /krej+r/ 'to believe'
b. /skrejver/ → skregver̥ 'to write'
c. /lavowra/ → lavogra 'works'

Note that both for Cypriot and Romansh, there is no place, continuancy or sonorancy assimilation. Thus the most likely first analysis is that this is a simple spread of the feature [consonantal], which is indeed how Kamprath formalizes the rule.

Since onsets tend to favor strengthening of consonants (CHAPTER 55: ONSETS), the Romansh case is useful in showing that position in onset is not a necessary condition for consonantalization of glides. The Greek case shows it is not sufficient, either, since a word-initial or post-vocalic glide does not strengthen, lacking as it does a preceding consonant.

(4) jerakos 'falcon'
lojazo 'pay attention to'

Cho and Inkelas (1993) attempt to re-analyze Cypriot and Romansh without [consonantal] spreading, and object that all the examples cited by Kaisse must refer to structural information in addition to features. This seems a peculiar objection, as many voicing and place assimilations, for example, affect only segments in coda position. Moreover, it is not true of all Kaisse's examples: Ahtna and Uyghur involve only features, not position in the syllable. And we have just seen that the Cypriot Greek case involves segments in the onset, while its mirror image in Romansh involves those in the coda. Nonetheless, with so few examples in the literature, there is indeed a danger that they can all be accounted for either by referring to syllabic position without [consonantal] or by replacing [consonantal] with some other active feature, and allowing consonantality to come along as a side-effect. Cho and Inkelas suggest that in Cypriot, consonantalization is a side-effect of the imposition of the [+cont][–cont] template that Kaisse argues must be imposed on all Modern Greek consonant clusters.

Hume and Odden (1996) take a more ambitious and thoroughgoing approach in attempting to dispose of Kaisse's examples. Their ultimate goal is to remove [consonantal] from the inventory of features altogether, by demonstrating that it is superfluous. They give three types of arguments.[4] First, no segments are distinguished only by the one being [+consonantal] and the other [–consonantal]. Thus, the difference between [w] and [β] is not only one of consonantality but also one of sonority, the difference between [j] and [ʎ] lies in laterality as well as consonantality, and so forth. Second, [consonantal] is never employed crucially to describe a natural class. (The class would be vowels, glides, and possibly

[4] Hume and Odden's line of attack follows from the same sort of reasoning as that given in Harris (1996). He argues that we can motivate the need for a particular distinctive feature in four ways: features code lexical contrasts, define natural classes of segments, determine syllabification, and guarantee phonetic interpretability.

Hume and Odden discuss two classes of phenomena where the class [−consonantal] does seem at first glance to be required. These are nasalization and the calculation of sonority. The first is exemplified by Arabela and several other languages, where nasality spreads to vowels, glides, and laryngeals (CHAPTER 78: NASAL HARMONY). However, Hume and Odden point out that the general hierarchy of segments able to be nasalized parallels the sonority hierarchy (see CHAPTER 49: SONORITY), so that in some languages, liquids are added to the class of nasalizable segments, in others voiced stops, and so forth. We will see in the next section that later research has suggested that the sonority hierarchy may in fact be a derivative notion. Hume and Odden anticipate this line of attack, arguing that the sonority hierarchy need not refer to [consonantal]; it can be unified with the nasalizability scale using the notion of impedance, "the resistance offered by a sound to the flow of air through the vocal tract above the glottis." Since laryngeals have no impedance, they are vowel-like and nasalizable, but they are inadequate syllable peaks, because some impedance is required for a syllable peak.

The remaining cases of consonantal spread offered in Kaisse (1992) are re-analyzed by Hume and Odden as fortitions involving other features: continuancy in the case of Cypriot Greek, voicing in the case of Uyghur, and perhaps some combination of structural position and continuancy dissimilation for Romansh.

J. Harris (1996) also proposes to eliminate the major class features and takes on a re-analysis of the Cypriot case as his major example. His argument is that the Cypriot process specifically targets a glide in the onset position of a coda-onset cluster. I do not think this particular solution will work, since the rule applies word-initially:

(7)	a.	/θios/ → θjos → θcos	'uncle'
	b.	pi	'drink'
		/na pi+o/ → na pjo → na pco *or* na fco	'that I drink'

Greek is particularly coda-averse and accepting of complex onsets, as judged by the intuitive syllabifications of native speakers and the wide variety of complex onsets *vs.* the very strict restrictions on coda consonants. Consequently, native speakers do not judge that a word-initial consonant like the [f] in [na.fco] is syllabified into the coda of a preceding vowel-final word (Joseph and Philippaki-Warburton 1987: 241ff.).

In conclusion, the cases of consonantal spread advanced by Kaisse (1992) constitute a reasonable body of data that suggest that [consonantal] may be an active feature that is useful in describing cases of consonantalization of glides and vowels and vocalization of consonants. But there are not enough cases that are immune from re-analysis to allow phonologists to rest comfortably in the knowledge that the feature is required in our theory. The rest of this section considers other evidence for the superfluity or redundancy of [consonantal].

Clements and Hume (1995) and Levi (2004) contain excellent demonstrations that [consonantal] in itself is inadequate to account for phenomena involving the transparency of consonants to vowel harmony and the distinction between underlying vowels and true underlying glides; what we need is a feature-geometric distinction between vowels, which have a vocalic place node, and consonants, which do not (unless they happen to have a secondary vowel-like feature such as palatalization). Levi (2004; CHAPTER 15: GLIDES) shows that if basic are best described using a multivalued [sonorant] feature, dispensing with the other major class features altogether.

Parker (2002) takes on the frequent complaint that [sonorant] has vague or conflicting phonetic definitions; he shows that the traditional feature (for which he ultimately supports a scalar rather than a binary-valued feature) and the traditional sonority hierarchy have regular phonetic correlates, namely intensity and, to a lesser extent, intra-oral pressure.

Sonority certainly does seem to be an important concept in speech perception, as one would expect if it is based in loudness. Recall that Stevens and Keyser (1989) pick it out as one of three perceptually salient features that can be expected to be used and enhanced in many segment inventories. This perceptual argument is reinforced from another angle in more recent work by Bates *et al.* (2007), who show that the feature [±sonorant] plays an important role in modeling pronunciation for speech recognition. Whether these perceptually based observations constitute arguments that impinge on the actual universal inventory of phonological features, however, is not entirely obvious. As we have noted already, the gold standard for motivating a phonological feature is its activity in one or more processes of assimilation, dissimilation, or deletion. Even less than [consonantal], [sonorant] does not seem to behave actively in such phonological processes. For this reason, McCarthy's (1988) survey article likewise locates it as an annotation on the root node. Furthermore, [sonorant] is subject to the same critique as [consonantal] when it comes to its use in defining the position of a segment in the sonority hierarchy and in sonority sequencing. We shall see in the next section that many phoneticians have argued that universals of segment ordering are derivative from perceptual factors. If that argument is accepted, the feature [sonorant] may be superfluous.

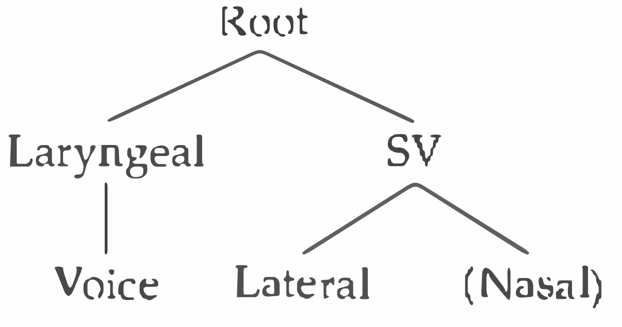
I am aware of only two direct defenses of [sonorant] as an active feature that participates in an assimilation process. Milliken (1988) suggests that flapping in North American English (see CHAPTER 113: FLAPPING IN AMERICAN ENGLISH) can be seen as the spread of [+sonorant]. In this process, a coronal stop becomes a voiced flap, roughly in inter-sonorant position. Olson and Schulz (2002) bring forward a much less familiar case, which thus bears presenting in some detail: the alternation of the 3rd person singular suffix /-ɲa/ in the Nilo-Saharan language Bilaala. This suffix appears as [ɲa] after nasals, liquids, glides, and vowels. That is, it is a nasal sonorant when adjacent to a sonorant. However, it appears as an obstruent following an obstruent, but maintains its place of articulation, ruling out spread of the entire root node. As the data below show, the process is not simply spread of [−sonorant]; the suffix also takes on the voicing and continuancy of the preceding obstruent, suggesting possible if unwieldy lines of re-analysis for someone attempting to dispose of this case.

(9)	a.	mon-ɲa	'his child'
		ɓɛr-ɲa	'his slave'
		kuhul-ɲa	'his hip'
		kaw-ɲa	'its length'
		ʧe-ɲa	'his mother'
	b.	got-ʧa	'his place'
		bob-ʤa	'his father'
		gurus-ʃa	'his money'

Olson and Schulz have certainly found a very relevant case for the question of whether [sonorant] is an active feature. Still, this process is a morphologically circumscribed one in Bilaala. It applies only to this pronoun; the nasal-initial 1st person singular, /-ma/, is invariant. One would like more robust cases on which to base a decision. As Olson and Schulz themselves say, until more cases are found, the question of the cross-linguistic phonological activity of [sonorant] remains unresolved. They offer and argue against a less elegant solution, where [−nasal] rather than [−sonorant] spreads. Such a solution is hardly ideal, requiring that nasal be a binary, not a privative feature, but it would allow a way out of the current case.

Rice (1993) attacks the traditional definition of [sonorant] from another direction. Summarizing and expanding work by Piggott (1990) and by Rice and Avery (1989), she proposes eliminating [sonorant] as an annotation on the root note and replacing it with a feature [spontaneous (or sonorant) voice], with dependents [lateral] and (predictable) [nasal].

(10)



Rice's proposal is based on the fact that even voiced obstruents in some systems, such as that of Southern Barasano (Piggott 1992) and Rotokas, can act like sonorants, take their voicing from sonorants or alternate with sonorants. These obstruents possess a spontaneous voicing node. In such languages, the basic division between consonants seems to fall not between the class of obstruents and the class of sonorants but rather between voiceless obstruents and everything else. Rice's proposal also encompasses the well-known fact that in many languages the voicing of obstruents acts very differently phonologically from that of sonorants. For instance, in Japanese, voiced obstruents block the voicing process known as Rendaku, but voiced sonorants do not block voicing. Rice shows that the familiar way of explaining such facts – namely that sonorant voicing only becomes available post-lexically – is not empirically adequate, motivating her alternative proposal.

3 The features [sonorant] and [consonantal] in sonority sequencing, weight, and lenition

Another strong motivation for retaining the features [consonantal] and [sonorant] may be that they are critical in describing sonority sequencing, the general tendency of syllables to have the most obstructed, consonant-like segments at the margins (the onset and the coda) and to progress to more vowel-like segments at the peak or nucleus. Clements (1990: 292) adds up the major class features [consonantal], [sonorant], and the less generally recognized [approximant] to give a sonority value for non-syllabic segments, resulting in the widely known scale.

of segments maximizes the robustness of each segment's cues. The phonotactics of human language are arranged to maximize redundancy of cues, strong auditory impact of each cue and resistance of cues to masking by surrounding noise. Since some segments contain internal cues that do not depend on their flanking segments, they are expected to figure in apparent exceptions to segment sequencing generalizations. Noisy (strident) fricatives, for instance, can be recognized without a following vowel or sonorant consonant because the noise of frication allows them to be uniquely identified. Thus /s/ figures repeatedly in clusters where sonority plateaus or is reversed (CHAPTER 38: THE REPRESENTATION OF SC CLUSTERS). Similarly, nasal consonants contain nasal poles and zeros that allow their manner of articulation, though not their place, to be reconstructed without a flanking vowel. Thus, sonority-reversing onset clusters like [nd] are frequently found in the world's languages, and not all can be explained away as prenasalized stops.

Stops, on the other hand, do not contain internal cues – their formant transitions and noise bursts are best perceived if a vowel or sonorant follows. Therefore, strings of stops not flanked by vowels or liquids are hard to recover (CHAPTER 46: POSITIONAL EFFECTS IN CONSONANT CLUSTERS). The ideal encoding for all segments results from the alternation of vowels and consonants. This is not only for the sake of the consonants. Though cues to vowels are the formants of the vowel itself, in natural speech, where a steady state is frequently not achieved, the formant transitions from a preceding consonant can add valuable information. CV is an optimal syllable not only because the V maximizes the cues to the C's identity but also because the C maximizes the cues for the V.

Clements (1990) encodes the naturalness of CVCV sequences with his construct of the "demisyllable": a sequence of adjacent tautosyllabic consonants + vowel or of adjacent tautosyllabic vowel + consonants (CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE). When conjoined with his Dispersion Principle, which favors maximal changes in sonority between adjacent segments in a syllable, the alternation of consonants and vowels becomes optimal. But without perceptual underpinning (CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY), we are still short of an explanation of why demisyllables with large sonority jumps are preferable.

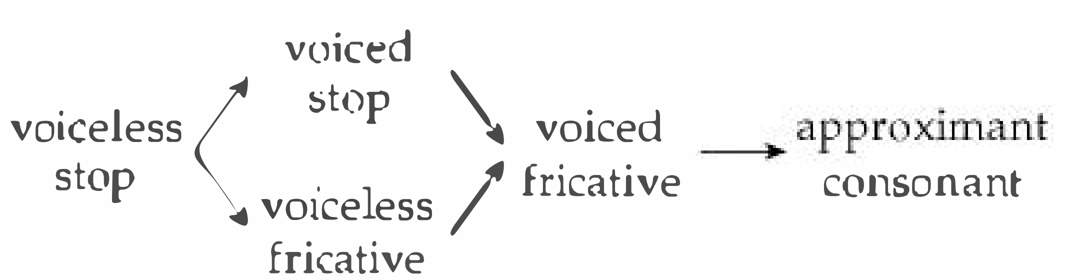
A cue-based theory can also offer an explanation for the preference which languages demonstrate for onsets over codas, for obstruent onsets over sonorant onsets, and for sonorant codas over obstruent codas, tendencies that do not fall out from the sonority sequencing principle without additional stipulations. Clements (1990) stipulates that steep dispersion in the CV demisyllable and shallow dispersion in the VC demisyllable are favored. And indeed this observation was a significant step forward. But it need not be a stipulation. Wright (2004) explains that the response pattern of the auditory nerves favors CV sequences over VC because of the perceptual boost that occurs at the beginning of a stimulus. An onset consonant, especially an obstruent, receives a boost but a coda consonant tends to be masked because the nerve fibers are already saturated by the sound of the vowel. (The nerve fibers recover quickly, in a matter of a few milliseconds, so that even the brief silence that occurs during a stop closure allows for the whole boost phenomenon to re-set; see Moore (1989). Therefore, even the burst and transitions of a medial C in a CVCV sequence get a boost, due to the silence created by the closure of that consonant.) Nasals are not particularly boosted in onset position, because the nerve fibers are saturated before the nasal ends, so nasals are less favored in onsets than obstruents are, yielding an apparent favoring of steep sonority dispersion in CV onsets and a shallow one in VC codas. Codas should be sonorous, because only sonorant consonants and fricatives, but not stops, can encode their own cues without needing release into a vowel. Cue-based sequencing explanations thus also cover the "Syllable Contact Law" (Murray and Vennemann 1983), which prefers a large fall in sonority between the coda of one syllable and the onset of an adjacent syllable. Even the tendency of nasals to lose their place of articulation in codas is encompassed in the cue-based view, since nasals have good voicing and manner cues internal to their realization, but not good place cues. The apparent preference for gradual sonority slopes from nucleus to coda is thus a by-product of the poorness of the internal cues of stops.

To summarize our discussion of the relation between the major class features and phonotactic constraints, the current view of many phonologists is that [consonantal] and [sonorant] are needed to describe sonority sequencing. The sonority feature may be binary (Clements 1990; Zec 2007) or multivalued (Selkirk 1984; Parker 2002). But there are also strong arguments that sonority sequencing is epiphenomenal, and that cue robustness covers the empirical ground better (Wright 2004). The last view, if accepted, undercuts one of the major arguments that we need [sonorant] or [consonantal] as a distinctive feature.

One might also think that the feature [sonorant] would be required to define syllable weight in languages where this concept is needed to describe the distribution of stress or contour tone or to define the minimal word. Zec (1995), for instance, makes major strides in demonstrating how something like sonority can operate in defining what constitutes a heavy syllable in a given language. However, Gordon (2002, 2004) argues that weight is a non-uniform, phonetically driven concept, not best captured with a feature. The definition of a heavy syllable can differ within even a single language, with the stress system, for instance, treating one kind of syllable as heavy while the tone system employs a different definition. Gordon finds that the energy within the syllable rhyme correlates with the tendency of a stress system to treat it as heavy, while the portion of the rhyme characterized by sonorant energy best determines whether it can bear contour tone. (For more on this discussion see CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE and CHAPTER 57: QUANTITY-SENSITIVITY.) So once again, the issue of whether [sonorant] is required remains without complete consensus.

Finally, the features [sonorant] and [consonantal], along with [continuant] and [voice], have been repeatedly argued to participate in weakening or lenition processes, especially those occurring intervocalically (CHAPTER 66: LENITION). The typical lenition pattern is summarized by Lass and Anderson (1975) as follows:

(13)



For instance, the pre-Old English phoneme /g/ weakens to /ɣ/ in Old English and to /w/ in Middle English before being lost altogether in late Middle English. While such developments are both common and intuitively natural, they are difficult to describe in a unified or coherent fashion using the binary distinctive features. Voicing and continuancy are kinds of manner features, while sonorancy and consonantality are major class features. Dependency phonologists (see for instance van der Hulst and Ewen 1991) have proposed that what is going on is the gradual addition of a |V| component, a sonorancy feature that is more present in voiced sounds than voiceless ones, more in fricatives than stops, more in sonorants than obstruents, and more in vowels and approximant consonants than in non-approximants. A similar solution has been proposed within Government Phonology by J. Harris (1990), which sees lenition as a loss of the elements representing occlusion and noise.[7]

4 The manner feature [continuant]

Three major issues dog the discussions of [continuant] in the literature (see also CHAPTER 16: AFFRICATES; CHAPTER 28: THE REPRESENTATION OF FRICATIVES). First, there is debate on the proper definition of the feature and thus about its values for laterals. Second, there is the question of what kinds of processes [continuant] participates in and what these tell us about where is it located on the feature tree. In particular, is it a direct and independent daughter of the root node, or is it somehow implicated in a daughter relation with Place? Or is it, perhaps, the organizing node of the tree, as proposed by Steriade (1993)? The last issue deals with the structure of affricates: are they a sequence of [−continuant] and [+continuant] features characterizing a single root (Sagey 1986), an unsequenced complex containing both values (Hualde 1987; Lombardi 1990), or basically stops whose release is fricated rather than strictly open (Steriade 1993)?

The definition of [continuant] is intimately bound up with an empirical question: do lateral consonants form a natural class with stops, which are [−continuant], or with fricatives and other [+continuant] sonorants such as rhotics? (See also CHAPTER 30: THE REPRESENTATION OF RHOTICS; CHAPTER 31: LATERAL CONSONANTS.) Chomsky and Halle (1968: 317) adopt the following definition:

(14) In the production of continuant sounds, the primary constriction in the vowel [sic] tract is not narrowed to the point where the air flow past the constriction is blocked; in stops the air flow through the mouth is effectively blocked.

Since laterals have airflow along one or both sides of the tongue, they are [+continuant] by this definition, and we will want to adopt it if it can be demonstrated that the great majority of phenomena bearing on the question categorize laterals with fricatives rather than with stops.

Halle and Clements (1983: 7), on the other hand, consider laterals to be noncontinuants, and therefore modify the definition.

(15) Continuants are formed with a vocal tract configuration allowing the airstream to flow through the midsagittal region of the oral tract; stops are produced with a sustained occlusion in this region. (Vowels, glides, *r*-sounds, fricatives *vs.* nasal and oral stops, laterals.)

7 Within Optimality Theory, Kirchner (2004) proposes that lenition is not a phonological process of spreading that refers to distinctive features at all. Rather it is achieved through the high ranking of a constraint LAZY, which penalizes biomechanical effort and works on a highly detailed phonetic scale of such effort.

as well as obvious strengths. In a survey of some 6,000 languages, it is not practical to investigate each case in anything approaching the detail I was able to lavish on twelve cases. But sometimes the devil is in the details.

In Harris and Kaisse (1999: 142) we analyzed this process as spread of the feature [−continuant]:

(17) X [COR, −son, +voice]
[−cont]

Navarro Tomás (1965) describes virtually the same process for Castilian Spanish. Since this example manifestly involves a change from [+continuant] to [−continuant], Kaisse (2000) considered it to be the best kind of evidence for the continuancy of /l/. Mielke's Finnish case is not so obviously related to continuancy. It could, for instance, have to do with the unmarked status of /t/. Nonetheless, it seems unlikely that such a stance could successfully dispose of all 30 of Mielke's examples. But since Finnish is the only [+continuant] case he reports in any detail, the matter cannot yet be fully resolved.

Mielke also discusses a different and intriguing kind of ambiguous behavior on the part of laterals. As has been much discussed, in Catalan, Spanish, and Basque, the lateral acts as [+continuant] before consonants at some places of articulation and [−continuant] before other places. In a related but distinct kind of case, some speakers of Hungarian add /l/ to the class of palatalizable non-continuants while others do not (Abondolo 1988: 64, cited in Mielke 2005). Again, let us consider the Spanish version of this phenomenon, as it is both well-known and frustratingly complicated. The process in question distributes the voiced obstruents, which come in stop and fricative variants. (J. W. Harris 1985; Kenstowicz 1994). But the case is really too vexed to draw any firm conclusion from. The judgments for these alternations are unstable – the same speaker can produce the stop and the fricative variant in the same environment – there are variations amongst idiolects and dialects that are not understood, and the theoretical bases for some of the analyses that have been proposed are shaky. Almost every analysis of the phenomenon one could think of has been vigorously defended by an excellent phonologist, and no consensus has emerged. The process works as follows. Spanish has no underlying contrast between voiced stops and voiced fricatives. Descriptions generally simplify the distribution as follows: after vowels, fricatives, and *r*, the voiced fricative allophones [β ð ɣ] appear. The voiced stops appear initially, after a pause, and after stops and nasals.

(18)	tu βeso	'your kiss'	besos	'kisses'
	los βesos	'the kisses'	um beso	'a kiss'
	por βesos	'through kisses'	futbol	'football'

The behavior of *l* as a trigger is more complex: following the lateral, we find the stop [d] but the fricatives [β ɣ].

(19)	el βeso	'the kiss'
	el dinero	'the money'
	el ɣato	'the cat'

Exactly what one is to make of even these simplified facts has been the subject of some debate, well summarized in Kenstowicz (1994: 487–489) Apparently, the process involves the spread of [continuant]. If this is so, is the lateral showing its true colors in contact with the homorganic obstruent [d] or with the non-homorganic [β ɣ]? We saw in the affrication process described earlier that some speakers may require homorganicity in order for a consonant to spread its continuancy to a palatal fricative. This suggests that it is in clusters with coronals that /l/ shows its continuancy – it is [-continuant] – and that fricatives are the underlying members of the voiced obstruent phoneme, emerging unscathed after *l* when not homorganic with it. On the other hand, the mixed environments in which the stop variants occur (after pause as well as after nasals and stops) suggests that the stops are underlying and the voiced fricatives are derived by spread of [+continuant]. Following that line of reasoning would suggest that /l/ is [+continuant].

But the facts are actually even more complicated. The rule is highly variable, and even utterance-initial fricatives are occasionally produced. I have had the same speaker produce a fricative and a stop in the same context, one after the other, in elicitation. Technically, the proposal of spread of [+continuant] is hard to execute in theories where non-contrastive features are not present in the lexical phonology, because vowels would have to be represented as [+continuant] at the point in the derivation where spreading occurs, while voiced obstruents were still [0 continuant] at that point (so as to be able to take on the continuancy of the preceding segment). The upshot is that we would do better to base our decision on the continuancy of laterals on less daunting cases. Mielke in fact uses this case to argue that even within a single process of a single language, the continuancy of a lateral may differ. But one stands on shaky ground at this point in drawing conclusions from this phenomenon. Similar comments can be made about two other Iberian languages, Catalan (Mascaró 1984, 1991) and Basque (Hualde 1991), whose continuancy alternations are remarkably similar to those of Spanish, probably due to areal influences. (See also CHAPTER 28: THE REPRESENTATION OF FRICATIVES.)

An appealing recent proposal is that features are emergent, rather than innate (Pulleyblank 2003; Mielke 2004, 2005, 2008). Speakers are not born with a definition for continuant and a categorization of segments according to that feature (see also CHAPTER 17: DISTINCTIVE FEATURES). Rather, they may group sounds together as a phonological category because those sounds have acoustic and articulatory similarities or because in the language they are learning, they act as a natural class. As Mielke puts it, stops and fricatives are archetypically and unmistakably [–continuant] *vs.* [+continuant]. So we almost always find them patterning as such in phonological processes, whatever language they occur in. But laterals and nasals have ambiguous cues and may be categorized as [+continuant] in one language and as [–continuant] in another. Even more iconoclastically, perhaps, Mielke's survey of 571 languages found 6,077 phonologically active classes of sounds, only 71 percent of which were describable by any set of distinctive features at all.

We turn now to the question of where the feature [continuant] should be located in the feature geometry. Kenstowicz (1994) contains an excellent summary of the issues. As mentioned at the beginning of this chapter, there are comparatively few phonological processes which make use of the feature [continuant], so while the need to employ it to differentiate contrasting segments has not been challenged, On the other hand, a rule like Zoque post-nasal voicing, which is looking at the left edge of the affricate, would treat it as a [−continuant] and class it with the stops, which voice after nasals, rather than with the fricatives, which do not:

(21)	min-pa	→	minba	'he comes'
	pʌn-ʧʌki	→	pʌn-ʤʌki	'figure of a man'
	winsaʔu	→	winsaʔu	'he received'

Unfortunately, it turns out that there are also phonological processes that class affricates and stops together even if they are looking at the wrong edge of the affricate, such as a deletion rule in Basque (Hualde 1987) that deletes a stop before another stop but turns an affricate to a fricative in the same environment, as it were deleting the [−continuant] portion of the affricate:

(22)	bat paratu	→	ba paratu	'put one'
	hots bat	→	hos bat	'a cold'

Kenstowicz (1994: 499–503) contains an excellent summary of the various moves theoreticians made after Hualde's demonstration that affricates were not simply contour segments with the [−continuant] segment ordered first. The most comprehensive solution yet proposed is that of Steriade (1993). As mentioned above, Steriade treats affricates as a species of stop whose release feature has the same degree of stricture as a fricative's. Thus affricates are generally expected to pattern with stops, as in the Basque and Zoque cases. However, the adjacent fricative releases that would result in English if the [s] or [z] allomorph was employed after an affricate prevent affricates from acting with the stops in this case.

To summarize this section, a number of proposals for the proper position of [continuant] in the feature geometry have been proposed, many with positive empirical consequences. However, no consensus has been reached which would allow us to say that one of them has been the most widely adopted by the phonological community. On a similar note, the proper extension of the feature to laterals remains troubled and may be a reflection of the emergent nature of features, which have no universal or innate definition. Finally, the behavior of affricates is not well explained by regarding them as segments with ordered minus and plus [continuant] values (see CHAPTER 16: AFFRICATES).

REFERENCES

Abondolo, Daniel. 1988. *Hungarian inflectional morphology*. Budapest: Akadémiai Kiadó.

Anderson, Stephen R. 1974. *The organization of phonology*. New York: Academic Press.

Aronoff, Mark & Richard T. Oehrle (eds.) 1984. *Language sound structure*. Cambridge, MA: MIT Press.

Bates, Rebecca, A., Mari Ostendorf & Richard Wright. 2007. Symbolic phonetic features for modeling of pronunciation variation. *Speech Communication* 49. 83–97.

Bladon, Anthony. 1986. Phonetics for hearers. In Graham McGregor (ed.) *Language for hearers*, 1–14. Oxford: Pergamon Press.

Blevins, Juliette. 1995. The syllable in phonological theory. In Goldsmith (1995), 206–244.

Byrd, Dani. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24. 209–244.

Cho, Young-Mee & Sharon Inkelas. 1993. Major class alternations. *Proceedings of the West Coast Conference on Formal Linguistics* 12. 3–18.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. 1987. Phonological feature representation and the description of intrusive stops. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 29–50.

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 283–333. Cambridge: Cambridge University Press.

Clements, G. N. 1991. Place of articulation in consonants and vowels: A unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5. 77–123.

Clements, G. N. 2003. Feature economy in sound systems. *Phonology* 20. 287–333.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

de Lacy, Paul (ed.) 2007. *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press.

Deligiorgis, Ioanna. 1988. Glides and syllables. Ph.D. dissertation, University of Iowa.

Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.

Gordon, Matthew. 2002. A phonetically driven account of syllable weight. *Language* 78. 51–80.

Gordon, Matthew. 2004. Syllable weight. In Hayes *et al.* (2004), 277–312.

Hall, T. A. 2007. Segmental features. In de Lacy (2007), 311–334.

Halle, Morris. 1992. Features. In William Bright (ed.) *Oxford international encyclopedia of linguistics*, vol. 3, 207–212. New York: Oxford University Press.

Halle, Morris & G. N. Clements. 1983. *Problem book in phonology*. Cambridge, MA: MIT Press.

Hankamer, Jorge & Judith Aissen. 1974. The sonority hierarchy. *Papers from the Annual Regional Meeting, Chicago Linguistic Society: Parasession on natural phonology*, 131–145.

Harris, James W. 1985. Autosegmental phonology and liquid assimilation in Havana Spanish. In Larry D. King & Catherine A. Maley (eds.) *Selected papers from the 13th Linguistic Symposium on Romance Languages*, 127–148. Amsterdam: John Benjamins.

Harris, James W. & Ellen M. Kaisse. 1999. Palatal vowels, glides and obstruents in Argentinian Spanish. *Phonology* 16. 117–190.

Harris, John. 1990. Segmental complexity and phonological government. *Phonology* 7. 255–300.

Harris, John. 1996. Phonological output is redundancy-free and fully interpretable. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, vol. 1, 305–332. Salford: ESRI.

Hayes, Bruce. 1986. Inalterability in CV phonology. *Language* 62. 321–351.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.

Hualde, José Ignacio. 1987. On Basque affricates. *Proceedings of the West Coast Conference on Formal Linguistics* 6. 77–89.

Hualde, José Ignacio. 1991. *Basque phonology*. London & New York: Routledge.

Hulst, Harry van der & Colin J. Ewen. 1991. Major class and manner features. In Pier Marco Bertinetto, Michael Kenstowicz & Michele Loporcaro (eds.) *Certamen Phonologicum II: Papers from the 1990 Cortona Phonology Meeting*, 19–41. Turin: Rosenberg & Sellier.

Hume, Elizabeth & David Odden. 1996. Reconsidering [consonantal]. *Phonology* 13. 345–376.

Hyman, Larry M. 1985. *A theory of phonological weight*. Dordrecht: Foris.

Joseph, Brian D. & Irene Philippaki-Warburton. 1987. *Modern Greek*. London: Croom Helm.

Kaisse, Ellen M. 1992. Can [consonantal] spread? *Language* 68. 313–332.

Kaisse, Ellen M. 1996. The prosodic environment of s-weakening in Argentinian Spanish. In Karen Zagona (ed.) *Grammatical theory and Romance languages*, 123–134. Amsterdam & Philadelphia: John Benjamins.

Kaisse, Ellen M. 1998. Are laterals [−continuant]? Paper presented at V Encuentro Internacional de Lingüística en el Noroeste, Hermosillo, Mexico.

Kaisse, Ellen M. 2000. Laterals are [−continuant]. Unpublished ms., University of Washington.

Kamprath, Christine. 1986. The syllabification of consonantal glides: Post-peak distinctions. *Papers from the Annual Meeting of the North East Linguistic Society* 16. 217–229.

Kawasaki, Haruko. 1982. An acoustical basis for universal constraints on sound sequences. Ph.D. dissertation, University of California, Berkeley.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kirchner, Robert. 2004. Consonant lenition. In Hayes *et al.* (2004), 313–345.

Ladefoged, Peter. 1964. *A phonetic study of West African languages*. Cambridge: Cambridge University Press.

Ladefoged, Peter. 1982. *A course in phonetics*. 2nd edn. New York: Harcourt Brace Jovanovich.

Lass, Roger & John M. Anderson. 1975. *Old English phonology*. Cambridge: Cambridge University Press.

Levi, Susannah V. 2004. The representation of underlying glides: A cross-linguistic study. Ph.D. dissertation, University of Washington.

Lindblom, Björn. 1983. Economy of speech gestures. In Peter F. MacNeilage (ed.) *The production of speech*, 217–245. New York: Springer.

Lombardi, Linda. 1990. The nonlinear organization of the affricate. *Natural Language and Linguistic Theory* 8. 375–425.

Lozano, María del Carmen. 1979. Stop and spirant alternations: Fortition and spirantization processes in Spanish phonology. Ph.D. dissertation, Ohio State University.

Mascaró, Joan. 1984. Continuant spreading in Basque, Catalan, and Spanish. In Aronoff & Oehrle (1984), 287–298.

Mascaró, Joan. 1991. Iberian spirantization and continuant spreading. *Catalan Working Papers in Linguistics* 1991. 167–179.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

Mielke, Jeff. 2004. The emergence of distinctive features. Ph.D. dissertation, Ohio State University. Published 2008, Oxford: Oxford University Press.

Mielke, Jeff. 2005. Ambivalence and ambiguity in laterals and nasals. *Phonology* 22. 169–203.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Milliken, Stuart R. 1988. Protosyllables: A theory of underlying syllable structure in nonlinear phonology. Ph.D. dissertation, Cornell University.

Moore, Brian C. J. 1989. *An introduction to the psychology of hearing*. London: Academic Press.

Murray, Robert W. & Theo Vennemann. 1983. Sound change and syllable structure in Germanic phonology. *Language* 59. 514–528.

Navarro Tomás, Tomás. 1965. *Manual de pronunciación española*. Madrid: Consejo Superior de Investigaciones Científicas.

Newton, Brian. 1972a. *Cypriot Greek: Its phonology and inflections*. The Hague: Mouton.

Newton, Brian. 1972b. *The generative interpretation of dialect: A study of Modern Greek phonology*. Cambridge: Cambridge University Press.

Odden, David. 2005. *Introducing phonology*. Cambridge: Cambridge University Press.
Ohala, John J. 1990. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26(2). 319–338.
Olson, Kenneth S. & Paul H. Schultz. 2002. Can [sonorant] spread? In J. Albert Bickford (ed.) *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 46. Available (July 2010) at www.und.edu/dept/linguistics/wp/2002.htm.
Padgett, Jaye. 1991. Stricture in feature geometry. Ph.D. dissertation, University of Massachusetts, Amherst.
Padgett, Jaye. 2008. Glides, vowels, and features. *Lingua* 118. 1937–1955.
Parker, Steve. 2002. Quantifying the sonority hierarchy. Ph.D. dissertation, University of Massachusetts, Amherst.
Piggott, Glyne L. 1990. The representation of sonorant features. Unpublished ms., McGill University.
Piggott, Glyne L. 1992. Variability in feature dependency: The case of nasality. *Natural Language and Linguistic Theory* 10. 33–77.
Pulleyblank, Douglas. 2003. Covert feature effects. *Proceedings of the West Coast Conference on Formal Linguistics* 22. 398-422.
Rice, Keren. 1993. A re-examination of the feature [sonorant]: The status of "sonorant obstruents." *Language* 69. 308–344.
Rice, Keren & Peter Avery. 1989. On the interaction between sonorancy and voicing. *Toronto Working Papers in Linguistics* 10. 65–82.
Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.
Schein, Barry & Donca Steriade. 1986. On geminates. *Linguistic Inquiry* 17. 691–744.
Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In Aronoff & Oehrle (1984), 107–136.
Silverman, Daniel. 1997. *Phasing and recoverability*. New York: Garland.
Steriade, Donca. 1993. Closure, release, and nasal contours. In Marie K. Huffman & Rena A. Krakow (eds.) *Nasals, nasalization, and the velum*, 401–470. Orlando: Academic Press.
Steriade, Donca. 1995. Positional neutralization. Unpublished ms., University of California, Los Angeles.
Stevens, Kenneth N. & Samuel J. Keyser. 1989. Primary features and their enhancement in consonants. *Language* 65. 81–106.
Sulkala, Helena & Merja Karjalainen. 1992. *Finnish*. London & New York: Routledge.
Vennemann, Theo. 1972. On the theory of syllabic phonology. *Linguistische Berichte* 18. 1–18.
Weijer, Jeroen van de. 1993. The manner–place dependency in complex segments. *Linguistics* 31. 87–110.
Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Hayes *et al.* (2004), 34–57.
Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12. 85–129.
Zec, Draga. 2007. The syllable. In de Lacy (2007), 161–194.

14 Autosegments

William R. Leben

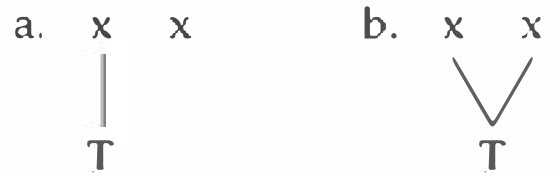
1 Tonal autosegments

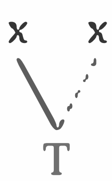
Autosegments are the central unit of representation in what generativists came to call non-linear phonology, an area described in earlier linguistic models under rubrics such as suprasegmentals (Trager and Bloch 1941), simultaneous components (Harris 1944), and prosodic analysis (Firth 1948). At the time autosegments were introduced by Goldsmith (1976), phonological and phonetic representations in the standard model of generative phonology were challenged for lack of adequate ways to express traditional prosodic features with the strictly segmental framework of Chomsky and Halle (1968; *SPE*). Even stress – widely regarded outside the generative school as a property of syllables – was treated as a feature on vowels in *SPE* (see CHAPTER 40: THE FOOT for more discussion), which in fact had no use for the syllable as a unit of representation, though it did posit a feature [syllabic] on individual segments (CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE).

If one of the early reasons for *SPE*'s huge success was the scope of what it promised to describe and explain, the framework's formal precision quickly made it apparent where to look for problems. Thus Woo (1969) hypothesized that contrastive lexical tones were level tones, like High, Mid, and Low, rather than contour (or "dynamic") tones like Rising and Falling, and that tones were expressed as features on vowels or on sonorants following the vowel in the same syllable. Together these two hypotheses made a dramatic prediction: "no dynamic tone can occur on a short vowel distinctively." After re-analyzing a few potential counterexamples to this generalization, she concluded that her prediction was essentially correct. New counterexamples to Woo's prediction arose in Leben (1971). That study provided independent support for Woo's analysis of rising and falling tones as level tone sequences and thus concluded that what must be wrong was Woo's hypothesis that each level component of a contour tone was assigned to its own unique segment.

A simple sign that contour tones are decomposed into level tone sequences comes from downstep, the gradual change in register that is triggered in many languages by a Low tone following a High tone (CHAPTER 35: DOWNSTEP). Leben (1971) noted that in a falling tone, the High component is subject to the same downstep effect as any level High tone, while the Low component triggers downstep just like any level Low tone. The same of course is true of the components of rising and rising–falling tones. Interestingly, this is true even in languages like Mende, where falling, rising, and rising–falling can all occur on single monosyllables. The conclusion is that contour tones do indeed behave like sequences of level tones, yet these level tones can be in a many-to-one relationship with a tone-bearing unit (TBU).[1]

Goldsmith (1976) supplies an equally elegant argument for one-to-many relationships between tones and tone-bearing units. The familiar case of tone spreading (Hyman and Schuh 1974), in which a toneless syllable is assigned the tone of an adjacent syllable, creates a one-to-many relationship between a tone and a set of tone-bearing units. Following Goldsmith, rightward tone spreading consists in the change from (1a) to (1b) and is schematized as in (2).

(1)


(2)


One-to-many and many-to-one relations among tones and tone-bearing units provided early motivations for the autosegmental model. Also persuasive were linguistically significant tone melodies going well beyond the individual tone-bearing unit. Drawing on work in some tone languages of West Africa by Welmers (1949), Rowlands (1959), and Edmondson and Bendor-Samuel (1966), Leben (1973) argued that word-length tone melodies motivated a level of representation in which tones were not represented on individual segments but on whole words or morphemes. A striking example is Mende, in which nouns exhibit the tone melodies H, HL, LH, L, and LHL. The fact that these melodies all occur on nouns of one, two, and three syllables argues that at some level the melody is independent of the number of syllables or segments in a word. Further, the non-occurrence of the melody HLH in words of any length strengthens the point. For Mende forms illustrating this point, see CHAPTER 45: THE REPRESENTATION OF TONE.

Hyman (1987) discusses a case "more like Mende than Mende" in a language from a very different group and region, namely Kukuya, analyzed by Paulian (1975). In the example below from Hyman (1987: 313–314), words of varying lengths are grouped by tone melody, given in the left-hand column. Like Mende, Kukuya has words of one, two, and three syllables, and attested tone melodies are combinations of High and Low. As in Mende, the melody HLH is not found. The second column shows the surface mappings schematically. A hyphen separates tones on adjacent syllables, e.g. L-L shows that an L appears

[1] Duanmu (1993, 1999) argues that Woo's original insight can be preserved by positing the mora as the universal TBU. For cases like Kukuya, where contour tones are assigned to syllables lacking phonological length, he suggests that *phonetic* length may suffice. Optimality Theory, with violable constraints, offers other alternatives for rescuing Duanmu's suggestion, even in the face of cases like Kukuya, with more tones than moras to bear them.

on the first and second syllable. Where two tones are not separated by a hyphen, they appear on the same syllable, e.g. LL-H indicates that both moras of the first syllable are L and the second syllable is H.

(3) *Kukuya* (Bantu, Congo)

underlying	*mapped*	*example*	
/L/	L	(ki).ba	'grasshopper killer'
	LL	(ki).baa	'jealousy'
	L-L	(kì).bala	'to build'
	LL-L	(kì).baala	'to cleave'
	L-L-L	(kì).balaga	'to change route'
/H/	H	(mà).ba	'oi. palms'
	HH	(mà).baa	'cheeks'
	H-H	(mà).baga	'show knives'
	HH-H	(lì).baama	'liana'
	H-H-H	(lì).balaga	'fence'
/LH/	LH	(mò).sa	'weaving knot'
	LH	(mò).saa	'seed necklace'
	L-H	(mò).sami	'conversation'
	LL-H	.saabi	'roofing'
	L-L-H	.m^{w}arəgi	'younger brother'
/HL/	HL	(ki).kaa	'to pick'
	HL	(ki).kaa	'to grill'
	H-L	(kì).kara	'paralytic'
	HL-L	(kì).kaara	'to be just right'
	H-L-L	(kì).karaga	'to be entangled'
/LHL/	LHL	(.ndí).bvi	'he falls'
	LHL	(.ndi).kaaj	'he loses weight'
	L-HL	(.ndí).pali	'he goes out'
	LH-L	(.ndí).baami	'he wakes up'
	L-H-L	(.ndí).kaləgi	'he turns around'

2 Compound segments

In another case that antedates and to some extent anticipates autosegments, Anderson (1974, 1976) explicitly questioned the adequacy of *SPE*-type segments in complex cases involving phenomena like nasality, aspiration, and intrusive stop formation. The essential insight was that they functioned as single segments, yet comprised a sequence of articulations like contour tones above. Illustrated below are prenasalized, postnasalized, and "medio-nasalized" stops (see CHAPTER 23: PARTIALLY NASAL SEGMENTS). The last-mentioned are alternants of prenasalized stops occurring between vowels in Kaingáng, a language of Brazil (Wiesemann 1972):

(4)

a. *prenasalized consonant* [^{m}b]	b. *postnasalized consonant* [b^m]	c. *"medio-nasalized" consonant* [^{m}b^m]
C	C	C
[+nas] [−nas]	[−nas] [+nas]	[−nas] [+nas] [−nas]

Drawing on a parallel between the level tone sequences making up rising and falling tones, autosegmental analyses began to represent prenasalized, post-nasalized, and "medio-nasalized" segments like the above as single consonants associated with sequences of autosegments with contrasting values for the feature [nasal].

It is easy to see parallels with other segments that phonology treats as single units even though they consist of sequence articulations, and by the late 1970s this two-tiered approach spread to affricates, which *SPE* had analyzed with the implicit sequence feature [delayed release].

3 The autosegmental formalism

The core of the autosegmental model is a set of multi-tiered linear sequences of feature bundles (segments), each tier linked to others by association lines. The "auto" of *autosegment* designates the relative autonomy or independence of the "segments" on one tier from those on another. Linking between one autosegmental tier and another is carried out by a general convention, the Well-Formedness Condition, supplemented by language-particular rules. Goldsmith's (1976: 48) version of the Well-Formedness Condition linked autosegmental tones to the segmental tier as follows, leaving aside a variety of possible adjustments:

(5) *Well-Formedness Condition* (WFC)[2]

a. All vowels are associated with at least one tone; all tones are associated with at least one vowel.
b. Association lines do not cross.

The WFC interacts with a set of language-specific rules and general operations linking one tier with the other. In the earliest analyses, the first tone of a melody would be assigned by a (possibly language-specific) rule linking a single tone with a single tone-bearing unit). Then, as we can see by inspection, the WFC would guarantee that all remaining tones and TBUs were linked. This elegant solution quickly became messier once it was applied to a variety of languages.

The first complication involved initial tone assignment. Since Williams (1976, circulated in 1971), the unmarked case in tone languages seemed to be that the first tone in a melody was assigned to the first TBU. But Goldsmith's (1977) analysis of English intonation, circulated in 1975, showed that an accented TBU would tend to attract a tone, whether or not the accented TBU or the tone it was to be paired with was leftmost in the string. Haraguchi (1975) established a similar result for lexical accents in Japanese, with additional complications (CHAPTER 120: JAPANESE PITCH ACCENT). The pairing of tones with accents was widely recognized by pre-generative traditions, as shown by the traditional term *pitch accent*, but the generative phonologist's need to account for this pairing in explicit terms led to new investigations of what aspects of tone association were and were not universal (CHAPTER 42: PITCH ACCENT SYSTEMS).

2 As McCarthy (2004) reminds us, (5) was one of the original generativist constraints, expressing a set of requirements to be satisfied rather than specifying an operation (see also Bird and Ladd 1991).

Kahn's (1976) dissertation explored the consequences of expressing the syllable as an autosegmental unit, with segments assigned to syllables by autosegmental association lines. Up to that time, syllable constituency had been represented through the placement of syllable boundaries among a linear string of segments. Kahn argued that the many-to-one associations typical of autosegmental phonology predicted the possibility of the ambisyllabic segment, a single segment simultaneously belonging to two adjacent syllables. This insight led to even more highly structured approaches to the syllable; see §8 and CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE.

Autosegmental association was also enlisted early on to account for phonological length (Leben 1977, 1980; McCarthy 1979; Hayes 1980; Ingria 1980; Clements and Keyser 1983; see also CHAPTER 20: THE REPRESENTATION OF VOWEL LENGTH; CHAPTER 37: GEMINATES). The basic idea, associating one length unit with a short segment and two with a long one, was incorporated into moraic theory (Prince 1983; Hyman 1985; McCarthy and Prince 1986; Itô 1989).

By applying autosegmental notions to vowel harmony in Akan, Clements (1977) accounted for some of the defining properties of vowel harmony. These included the one-to-many relationship between a phonological feature and the feature-bearing units of a phonological string and the tendency of such a feature to spread automatically to neighboring segments unspecified for that feature (CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS; CHAPTER 118: TURKISH VOWEL HARMONY; CHAPTER 123: HUNGARIAN VOWEL HARMONY).

These developments would eventually lead to a theory of feature geometry assigning structure to what had been "bundles" of unordered distinctive features in the *SPE* theory. The structure was a hierarchical arrangement into groups, reflecting ways in which one feature set affects others. In effect, as McCarthy (1988: 89) noted, feature geometry "constitutes a model of the phonologically relevant characteristics of the human vocal tract." We return to feature geometry in §6.

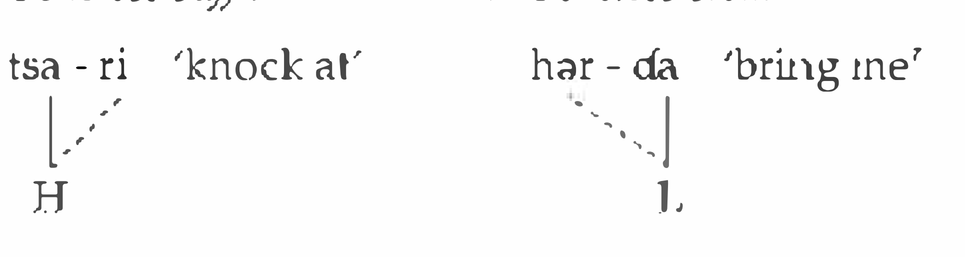
5 Autosegmental linking and spreading

As autosegmental analyses accumulated, conceptions altered about what the Well-Formedness Condition could capture as truly universal. As noted in §3, initial tone linking was found to differ quite a bit from language to language. The same is true of tone spreading, for which Pulleyblank (1986) observed three basic situations: some languages don't allow spreading at all, while others spread tone only to a single adjacent TBU and still others apply spreading to an unbounded sequence of TBUs.

Pulleyblank (1986) shows that in Margi, stems and suffixes can bear either High or Low tones or, as in the examples below, one or the other can be toneless. When either is toneless, it gets its tone from its neighbor, whether on the right or on the left. The dotted line indicates the change:

(7) *Margi* (Chadic, Nigeria)

a. *Toneless suffix* b. *Toneless stem*



Goldsmith's original formulation of the WFC accounted for this spreading automatically, but Pulleyblank (1986) proposed several modifications, observing that languages differ in how they treat tone spreading. Some don't allow spreading at all, while others spread tone only to a single adjacent TBU, and still others apply spreading to an unbounded sequence of TBUs. Pulleyblank (1986: 11) consigns spreading to language-specific rules and is left with (8) and (9) as the assumed universals. In the new formulation the only part of the process that is a constraint rather than an operation is (9).

(8) *Association conventions* (Pulleyblank 1986)

Map a sequence of tones onto a sequence of tone-bearing units,
a. from left to right;
b. in a one-to-one relation.

(9) *Well-Formedness Condition* (Pulleyblank 1986)
Association lines do not cross.

Here is how these two principles apply in Kukuya examples drawn from (3):

(10) *Kukuya*

a. baami 'he wakes up'
LHL

b. kaara 'to be just right'
HL

c. ɪnwarəgi 'younger brother'
LH

d. kaaj 'he loses weight'
LH L

In (10a) left-to-right mapping assigns a tone to each TBU. The fact that two tones are assigned to the bimoraic vowel /aa/ suggests that the mora is the TBU. In (10b) H is assigned to the first mora and L to the second. Kukuya-specific tone spreading needs to extend L to the final vowel. In another language, spreading might not happen, in which case Pulleyblank's approach would place a Low tone on the final vowel by default if no other rule assigned a tone beforehand.

In (10c) the first two tones are assigned to the first two TBUs. As in (10b), the rightmost TBU receives its tone by spreading from the left. However, this predicts the mapping LHH instead of the correct LLH. Another Kukuya-specific rule must be invoked – a very natural one, according to Hyman and Schuh (1974), which moves the transition from L to H in the sequence LHH one TBU to the right, giving LLH. An alternative that has received some attention would be to first associate the tones at the left and right edges of the melody with the TBUs closest to the corresponding edges of the segmental string, with later adjustments for any remaining TBUs or tones. Edge-In Association is illustrated in connection with example (16) below.

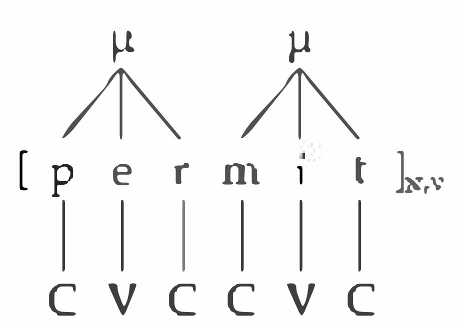
In (10d) the first Low tone is assigned to the first mora of /aa/ and High is assigned to the second. A Kukuya-specific process needs to assign the Low floating on the right to a TBU. This process is subject to the WFC, preventing the rightmost Low from going to the leftmost mora. As a result, the WFC guarantees that the last tone will go on the last TBU. For more detail on the WFC and autosegmental association, see Hammond (1988).

6 Non-concatenative morphology

Another significant outgrowth of autosegmental phonology was the extension of multi-tiered representations to morphology, beginning with McCarthy's (1979, 1981) templatic analysis of Arabic, applying the notion of long components (Harris 1951) to morphemes in general, and using autosegmental notation instead of the boundary symbols used by *SPE* to delimit morphemes. The examples in this section illustrate some developments in templatic morphology that fed back into phonology, including the addition of a CV tier and some new principles for linking one autosegmental tier with another (see also CHAPTER 105: TIER SEGREGATION; CHAPTER 108: SEMITIC TEMPLATES).

McCarthy illustrates the basic idea with this informal autosegmental representation of the English noun and verb *permit*:

(11)



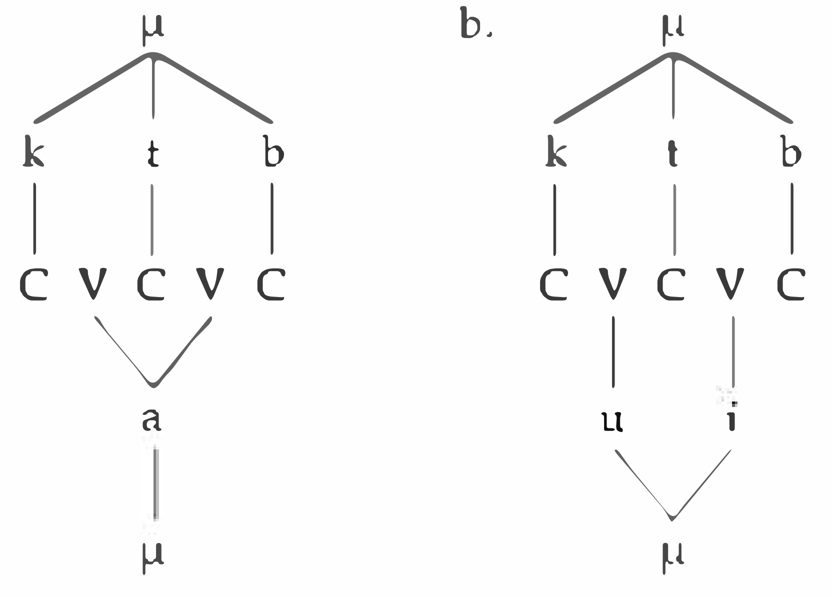
where μ is an autosegmental node representing a morphemic constituent lined by association lines to the sequences *per* and *mit*, which stand for the feature bundles that spell out the prefix *per-* and the root *mit*. The Cs and Vs at the bottom – shorthand for [–syllabic] and [+syllabic] segments respectively – represent the form's overall syllable pattern and are called the prosodic template.[4] In templatic morphology Cs and Vs are the analog of the TBUs in autosegmental representations of tone, while the morphemes – segment sequences like *per* and *mit* – are the analog of tonal melodies in our earlier examples.

This approach provides a simple way for McCarthy to represent the phonological and morphological structure of forms in a language like Arabic, where roots typically consist of consonants, and vowel patterns are used to differentiate inflectional categories like active/passive and perfective/imperfective. In the example

[4] This model was adapted and put to a variety of new uses by Halle and Vergnaud (1980) and by Clements and Keyser (1983). Halle and Vergnaud referred to the prosodic template as the CV skeleton, and Clements and Keyser called it the CV tier.

below, the triliteral root /ktb/ 'read' combines with the vowel pattern /aa/ to form the perfective active stem /katab/ and the perfective passive stem /kutib/.

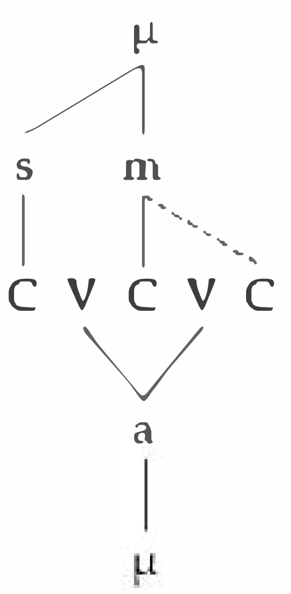
(12) a. b.



The tiered analysis expresses each stem's two discontinuous morphemes and at the same time its resulting phonological form. Not only are the separate tiers and the association lines borrowed from autosegmental phonology; to some extent, so too are the mapping principles. Note that the links between the morphemes in (12) and the C- and V-slots are consistent with principles (5), (8), and (9) above, as long as we require that [−syllabic] segments map onto C and [+syllabic] segments onto V. McCarthy makes a number of significant adjustments for more complicated cases, but the basic autosegmental framework provides an elegant start.

The need to associate every C position with a melody element leads to some interesting predictions. In (13) the bilateral root /sm/ combines with the vowel pattern /aa/ to form the perfective active stem /samam/. The structure is sketched in (13):

(13)

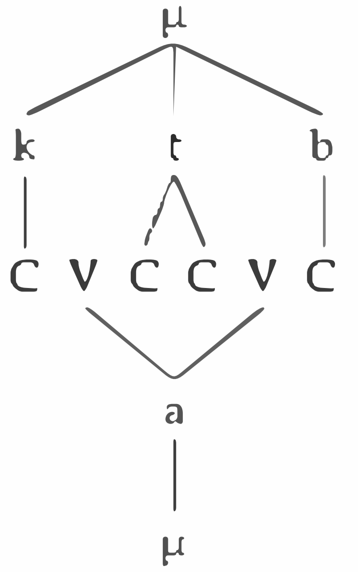


Surprisingly, the mapping procedure employed in (12) suffices for (13). Because the root contains only two consonants, and because the mapping procedure goes from left to right, we might expect the rightmost C position to remain empty. But the WFC requirement that all C positions must be filled will automatically add a link, and the prohibition against crossing association lines guarantees that the link will go to the second root consonant and not to the first.[5]

[5] As McCarthy notes, the absence of stems like */sasam/ is correctly predicted by this account, which maps from left to right. Another key component in this prediction is the Obligatory Contour Principle (see §9), which prohibits adjacent copies of a phonological element. This is yet another parallel between autosegmental accounts of phonological and morphological structure.

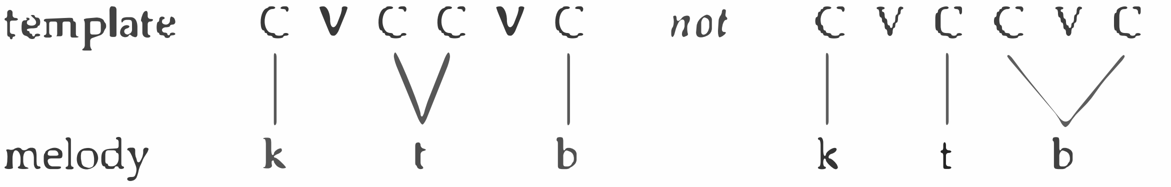
One complication in the autosegmental representation of morphological structure has led to a possible modification later applied to phonology. Consider how the proposal sketched in (12) would apply to the Arabic stem /kattab/, Form II of the active perfective Form I stem /katab/. Forms I and II have the same morphemic content and differ only in their prosodic template, CVCVC for Form I and CVCCVC for Form II (McCarthy 1981: 385–386). The second is shown in (14), after applying the mapping principles form above. Note that we get */katbab/ instead of the desired /kattab/.

(14)



For this purpose McCarthy devises a rule that results in a change from /katbab/ to /kattab/. This is fine, but one wonders whether a more general solution can be found. As an alternative, Yip (1988) proposes abandoning the left-to-right mapping principle for Arabic in favor of Edge-In Association, which would associate the leftmost and rightmost melody elements with the leftmost and rightmost Cs of the prosodic template:[6]

(15)



(16) gives a version of Yip's Edge-In Association:

(16) *Edge-In Association* (Yip 1988: 571)

a. *Anchoring*
Associate the outermost melodic elements to the outermost skeletal slots, one-to-one.

b. *Filling*
Associate the remaining melodic elements and the remaining slots in the same way.

c. *Template satisfaction*
Language-specific.

Yip goes on to mention tonal cases where (16) would be an improvement over left-to-right mapping, while allowing for the possibility that mapping tones

[6] As Yip notes, this rules out McCarthy's elegant analysis of /samam/ *vs.* */sasam/, and the account she offers in its place is unfortunately less general than the original one.

onto TBUs is governed by different principles from these (see also Archangeli and Pulleyblank 1994; Zoll 1997, 2003). Back in templatic morphology, Buckley (1990) shows that Edge-In Association predicts the correct mapping of Tigrinya roots onto noun and verb templates, for example in the key case of mapping triliteral roots onto the quadrilateral plural template, where other approaches fail.

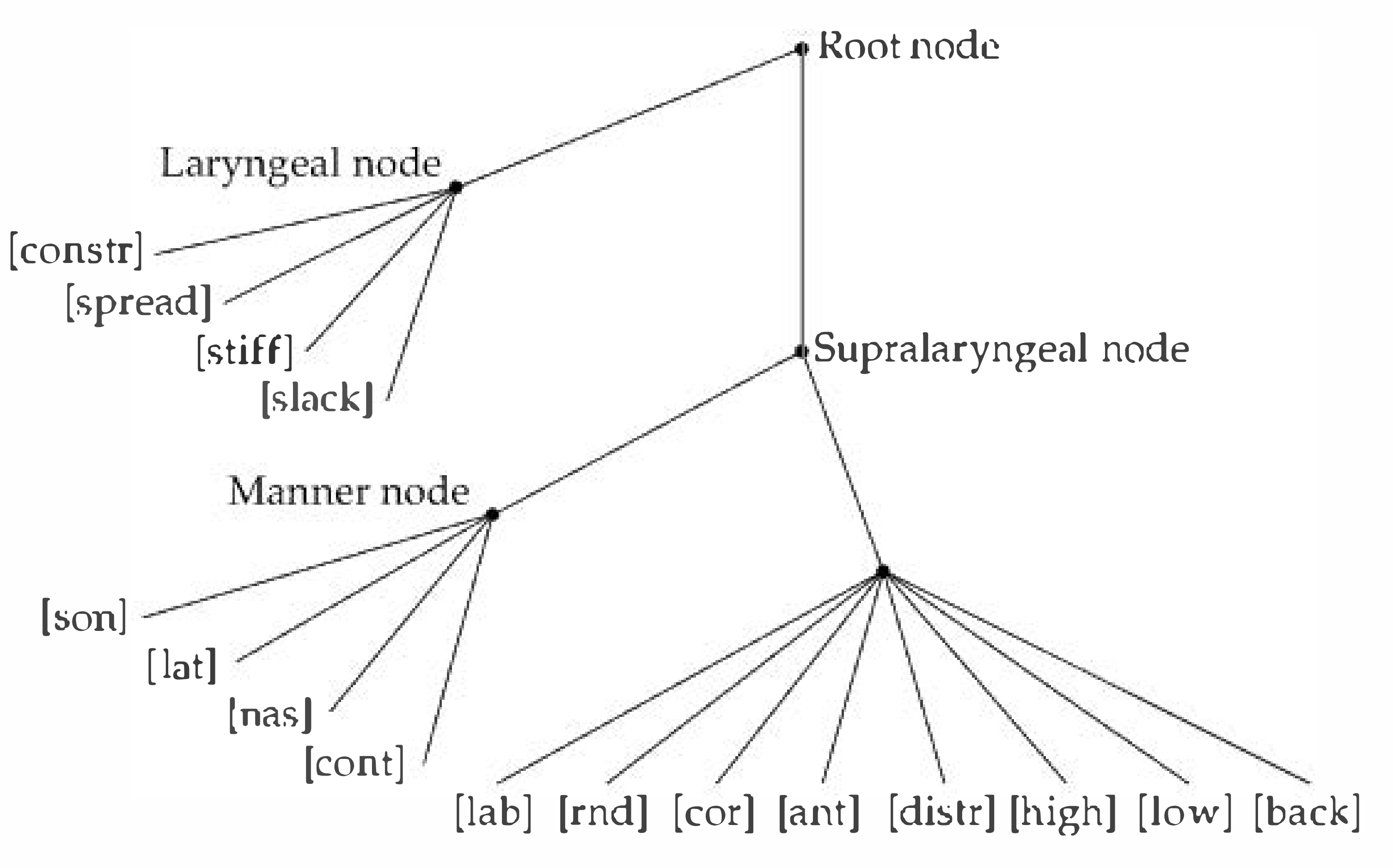
7 Feature geometry

In light of the autosegmental treatment of tone spreading and vowel harmony, which spread a given element across a string of feature-bearing units, consider now the standard assimilations of segmental phonology, such as voicing assimilation or nasal assimilation.

Phenomena like these could profit greatly from the autosegmental formalism if motivation could be found for abstracting features like [voice] and [nasal] from the segments.

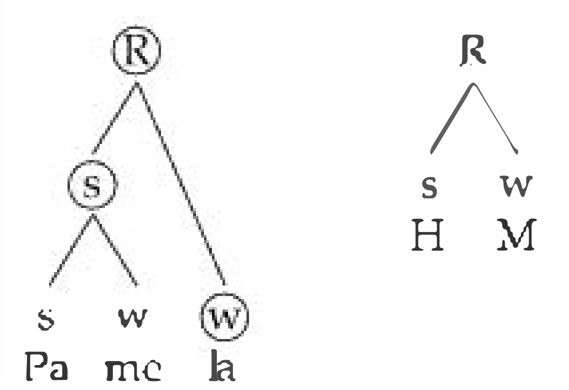
This is exactly what Goldsmith (1981) proposed, and the idea was pursued in rigorous detail by Clements (1985) and Sagey (1986). Clements (1985) noted that distinctive features grouped into neat classes, including place, manner, and supralaryngeal, which could be naturally represented by branching tree structures with phonological features as terminal nodes dominated by nodes representing feature classes (CHAPTER 27: THE ORGANIZATION OF FEATURES). Subclasses like place and manner could in turn be dominated by other classes, like supralaryngeal. In Clements's system, each feature and feature class was an autosegmental tier, and this tier in turn was linked to the tier above. (17) is a sketch adapted from McCarthy (1988: 89):

(17)



The motivation for each grouping lies in the cross-linguistic tendency of features within a group to behave together. For example, the place node above groups together features that tend to undergo assimilation together. Independent motivation postulating a metrical structure for autosegmental tunes[7] (illustrated on the right in (20)) and a procedure using this structure to fit the tune to metrically congruent parts of the text (illustrated on the left in (20)). The binary metrical feet can be related to early work in generative metrics, e.g. Halle and Keyser (1972). (The nodes labeled R are root nodes of metrical trees for text and tune.)

(20)



Liberman also included what he called a metrical grid to account for rhythmic structure. Across a fairly rich range of cases, these two additions enabled an English intonational melody to be linked to the right places in the intonational phrase. Liberman retained the autosegmental notion that complex tones were resolved as sequences of level ones, an idea pursued even further by others, beginning with Pierrehumbert (1980), as described in the following paragraphs.

On the right in example (20) above is an early autosegmental intonational "tune." To Liberman and Sag (1974), these tunes were meaningful – Liberman (1975) called them morphemes and stored them in a lexicon. This notion soon led to work on the internal structure of the melodies, beginning with Pierrehumbert (1980), Beckman and Pierrehumbert (1986), and Pierrehumbert and Beckman (1988). The constituents of intonational melodies included pitch accents, phrase accents, and boundary tones. Each constituent is a sequence of one or more autosegments, with a few enhancements to designate which autosegment was assigned to an accented syllable or to a given position in the phonological or intonational phrase. (An intermediate phrase was also posited for some languages, e.g. Japanese in Pierrehumbert and Beckman 1988.) This framework served as the basis for the widely used ToBI intonation transcription system (Beckman and Ayers 1994; Jun 2005).

In the example below, from Pierrehumbert and Hirschberg (1990), the instances of H* are pitch accents, each associated with the accented element in the text just above it. The symbols H and L not followed by an asterisk are called phrase tones: H is the phrase tone for the first phrase, and L is the phrase tone for the second. The L% at the end marks the utterance boundary (see also CHAPTER 32: THE REPRESENTATION OF INTONATION; CHAPTER 116: SENTENTIAL PROMINENCE IN ENGLISH; CHAPTER 50: TONAL ALIGNMENT).

(21) The train leaves at seven or nine twenty-five.
H* H* H* H H* H* L L%

[7] Such tunes are familiar from earlier American, British, Dutch, and other traditions, but the decomposition of tunes into sequences of level tones, drawn from the same inventory as the tones of lexical tone languages, was novel, if similar in some respects to earlier American approaches, notably Pike (1948).

point out, every intonational contour obeys the OCP, i.e. consists only of sequences of unlike tones.

Jun (2005: 434–435) provides a table summarizing the ToBI analyses of intonation in a sample of 11 languages that are typologically quite diverse. Listed in the table are all of the attested ToBi tone sequences for pitch accents, phrase tones, and boundary tones in these languages. Jun's table includes 22 cases of pitch accents comprising two or more tones, such as L+H*, H+↓H*, and H*+L. Three of Chickasaw's four pitch accents violate the OCP: LHHL, LL, and LHH. The German pitch accent H*↓H either conforms or does not conform, depending on whether we regard the downstep ↓ as a break between the adjacent H tones.[10] Aside from these cases, the remaining 18 sequences feature only unlike tones in adjacent positions. For phrase tones and boundary tones, the pattern is clear. The two bi-tonal phrase accents and 13 multi-tonal boundary tones all conform to the OCP.[11]

However, the literature is rife with examples of tone languages violating the OCP in one way or another (see Myers 1997). Some of the most enduring results on the question were captured in Odden (1986), with additional details in Odden (1995). One of these is that languages differ in how they deal with sequences of identical tones across prosodic boundaries. Odden (1995: 462) suggests that Kipare fuses High tones across phrase boundaries into a single High tone. The evidence is that a prepausal sequence of High tones is realized as Low after a floating Low tone, even if they are from different prosodic words. Here is one of his examples:

(27) *Kipare* (Bantu, Tanzania)

vá↓ná vékíjílá nkhúku ndórí nkhúndú jángú
'while the children eat those little red chickens of mine'

The change that applies to this string is that *all* of the High tones following the downstep marker (↓) become Low, despite the fact that this expanse crosses many word boundaries.

A related observation comes from Hausa (Inkelas and Leben 1990), where emphatic intonation raises High tone to the extra-High register. Once the extra-High register is triggered, all remaining High tones in the phrase are pronounced extra-High, even on words not emphasized. In (28), the leftmost two syllables have a Low tone, marked with a grave accent, and the remaining ones are lexically High-toned. /neemoo/ 'look for', is emphasized in (28b) but not in (28a). In (28b), when the register shifts up for /neemoo/, it stays up for the remaining High-toned words, even though they are not emphasized.

(28) *Hausa* (Chadic, Nigeria)

a. *Normal*
Bà mù | neemoo awarwaron Maanii ba
'We didn't look for Maanii's bracelet.'

b. *Emphatic*
Bà mù | *neemoo* awarwaron Maanii ba
'We didn't *look for* Maanii's bracelet.' (*neemoo* is emphasized)

[10] Similar cases arise in African languages: see example (28) below and the surrounding discussion.
[11] But see Ladd's (2008) review, cautioning against generalizing across languages or even across varieties of one language based on ToBI analyses.

Odden's other seemingly inescapable result is briefly that in some languages the sequence HH can contrast with a single H. He analyzes Shambaa (Bantu, Tanzania) as having two types of High-toned sequence: one type that surfaces as level High, as in (29a), and another surfacing with a downstepping pattern, as in (29b):

(29) *Shambaa* (Odden 1986, 1995)

```
a. /njoka/  'snake'        b. /ngoto/  'sheep'
      \/                          |  |
      H                           H  H

   Surface patterns
      njóká                      ngó↓tó
     [   ¯  ¯ ]                 [  ¯  – ]
```

Odden argues that the OCP ought to reduce H and HH to the same melody, yet clearly H in (29a) contrasts with HH. Here then is a lexical tonal contrast that contradicts the OCP.

However, it should not be missed that Shambaa also exemplifies the principle underlying the OCP – avoidance of sequences of identical tones – since wherever two High tones come together, they are broken up by a downstep. The same happens in some other Bantu languages, e.g. Namwanga (Bickmore 2000: 302), and in some non-Bantu languages, e.g. the Chadic language Miya (Schuh 1998). Cassimjee and Kisseberth (2001: 329–331) offer examples of closely related languages that differ in the conditions for inserting downstep between adjacent High tones, if downstep is inserted at all.

Subsequent work, beginning with Myers (1997), has gotten considerable mileage from regarding the OCP as a violable constraint – reasonably enough, since most constraints can be violated. Again, see Myers (1997). We return briefly to Shambaa in §10.1 in connection with the tonal foot.

As with Yip's (1988) proposal for edge-in mapping, it is still an open question whether the OCP for tone is the same constraint as the OCP applied to other features. McCarthy's (1986) suggestion to apply the OCP to antigemination in general met with a prompt retort from Odden (1988), who wrote on anti-antigemination. The matter is still unresolved.

10 Prosodic groupings of autosegments

Two areas of current exploration concerning autosegments are groupings of autosegments into "foot-like" structures (van Oostendorp and Revithiadou 2005) and the possibility of deriving the results of autosegmental theories of harmony from connectionist accounts.

10.1 Tonal feet

From the beginning, some of the basic units of prosodic phonology – notably prosodic words in the case of harmony processes and phonological phrases in the case of intonation – were found to form natural domains for sequences of autosegments. The pioneering autosegmental accounts of intonation grouped tones into tunes, while the enhanced autosegmental strings pioneered by Pierrehumbert (1980) and Pierrehumbert and Beckman (1988) went with phrase-like prosodic units.

Groupings of lexical tones into prosodic constituents came next. Rice (1987) and Goldsmith (1987) discussed cases in which High tone was attracted to a metrically strong position. Another early instance was Kenstowicz and Kisseberth's (1990) demonstration that metrical feet in Zigula (Bantu, Tanzania) determine where High tones shift to – in general, to the penult, the head of a disyllabic trochee. While Zigula lacks word stress phonetically, the TBU designated as the head typically attracts lexical High tones as well as intonational tones, and this has come to be regarded as a characteristic of feet in tonal languages.

For example, the Zigula 3rd person subject prefix /a-/ contributes a High tone to the verb. This High tone is realized on the penultimate syllable of the verb in the example below, where the other morphemes in the verb are toneless. The situation becomes more complicated when other morphemes contribute a High tone.

(30) *Zigula* (Kenstowicz and Kisseberth 1990: 167)

a- hugusahug(usa) 'he/she shells repeatedly'

H

Duanmu (1999) portrays a more complex interrelationship between metrical structure and tone in his account of the differences between Mandarin and Shanghai, illustrated here in the way the two languages deal with tone for the foreign name *Chicago*:

(31) a. *Mandarin* (transcribed in Pinyin)
H-H-H → H-H-H
zhi-jia-ge 'Chicago'

b. *Shanghai* (transcribed in phonetic symbols)
H-H-H → H-L-L
tsz-ka-ku 'Chicago'

In Mandarin, which contrasts full and light syllables, (31a) has three full syllables, and by foot constraints which Duanmu motivates independently, each syllable is stressed. This in turn guarantees, again by an independent constraint, that each syllable retains its High tone. Shanghai, on the other hand, does not contrast full and light syllables underlyingly (Duanmu 1993), and so the three syllables of (31b) group into a single foot, preserving only the tone of the head, the first syllable. The two remaining syllables are realized as Low by default.

Zec (1999) shows that in the neo-Štokavian dialect of Serbo-Croatian, stress and tone are both critical factors in the definition of the *foot*. This can be seen from the following foot inventory.

(32) *Serbo-Croatian* (Zec 1999)

without tone

a. $[\ \sigma_{\mu\mu}\ ]_{Ft}$ b. $[\ \sigma_{\mu}\sigma_{\mu}\ ]_{Ft}$

with tone

c. $[\ \sigma_{\mu\mu}\ ]_{Ft}$ d. $[\ \sigma_{\mu}\ ]_{Ft}$
H H

If a High tone (Zec's interpretation of a pitch accent; see CHAPTER 42: PITCH ACCENT SYSTEMS) is present, then a foot can be either one or two moras in length. In the absence of a High tone, the foot must be bimoraic – either a long syllable or two short ones.

Foot structure has also figured in recent accounts of African tone languages. For example, in Kera (Pearce 2006) the metrical foot, grounded phonetically, is the key to explaining the distribution and behavior of tone. She presents evidence (2006: 260) that when a word consists of exactly one foot, each syllable can carry a tone, but in words with two feet, each foot can only have one tone.

Leben (2002) offers a tonal foot reinterpretation of Odden's (1986) Shambaa data from (29) above. Odden proposes the underlying tonal contrast H *vs.* HH for the respective phonological representations in (29a) and (29b). But the only evidence for the underlying sequence HH is its surface realization [H$^{\downarrow}$H], the pattern that obtains in a phrase when a word ending in High directly precedes another High tone. This evidence is hardly conclusive. Suppose that some language inserts an epenthetic [t] between words ending in [n] and words beginning in [s]. This would not necessarily mean that word-internal [nts] has to be analyzed phonologically as /ns/. Leben (2002), agreeing that there is an important contrast in (29), suggests that (29a) has one foot, while (29b) has two:

(33) a. /njoka/ b. /ngoto/
(H) (H)(H)

He suggests that this contrast is analogous to the familiar one in stress languages, between one foot (as in English *attic*) and two (as in English *Aztec*).

Pearce (2006: 260–261) offers a summary, with a dozen references, of ways in which tones interact with metrical feet:

(34) *Types of interaction between tones and metrical feet* (Pearce 2006)

a. In the association of tones to heads.
b. In the deletion of tones on non-heads.
c. In the spreading of tones within the foot.
d. In the preference for certain tones on heads and non-heads.
e. In associating certain melodies with certain foot types.

Taken together, Pearce's references offer consistent cross-linguistic evidence for the tone-bearing unit as either the entire metrical foot or its head.

10.2 Tonal complexes

Akinlabi and Liberman (2001) argue for a similar entity. Their tonal complex is a grouping of lexical tones into a structure broadly described as the "tonal analog of the syllable." The basic claim is that constraints involving tonal complexes "can motivate deletion, epenthesis, spreading or reordering of tonal features, just as constraints on syllable or foot structure may motivate such processes in well-known cases of segmental phonology."

Tonal complexes help Akinlabi and Liberman explain why tone spreading in a schematic example like (35a) typically results in (35b) rather than (35c). (See Hyman 2007 for this observation.)

(38) a. *spreading* b. *shifting*

Cassimjee and Kisseberth's insight is that spreading and shifting are similar in expanding the H domain. The difference is that on the surface H comes to cover the entire domain, while shifting results in a single link at a domain edge.

The example below shows how closely spreading and shifting are related in three Nguni languages:[12]

(39) *Nguni languages* (Cassimjee and Kisseberth 2001)

	Swati	*Durban Zulu*	*Phuthi*
a.	[ba̲-já]naːtsa	[ú̲-kú]-ʧu-la	[kú̲-límí]siːsa
	'[cl.2] are drinking'	'to sing'	'to cultivate intensively'
b.	[ba̲-ja-khilikí]dzeːla	[bá̲-já-límá]liːma	[bá̲-já-límé]laːma
	'[cl.2] are stumbling'	'[cl.2] are cultivating (REP)'	'[cl.2] cultivate for one another'
c.	si-ja-[sɛ̲bé:]nta	si-ja-[la̲lé:]la	si-ja-[bɔ̲́ní:]sa
	'we are working'	'we listen'	'we show'

The H domains in these three languages are all constructed according to the same principles (see Cassimjee and Kisseberth 2001: 333–337). The surface tonal differences in the bracketed domains result from whether the underlying High tone (the underlined mora at the left edge of the bracketed expression) shifts or spreads or both. Cassimjee and Kisseberth informally represent a surface TBU that arises from underlying H with an underline – [ba̲] in the left column of (39a) – and surface H with an acute accent – [já] in the left column of (39a) – but this is purely a matter of notation, not a formal difference from the standard autosegmental representation. The differences among the three languages are summarized in (40).

(40) a. In Swati, input H *shifts* to the final syllable of the H domain.
b. In Durban Zulu, input H *spreads* from prefixes to the final syllable of the H domain and *shifts* from input stems to the final syllable of the H domain.
c. In Phuthi, input H *spreads* to the final syllable of the H domain.

Thus Swati involves shifting, Phuthi involves spreading, and Durban Zulu involves both. The ODT analysis makes this difference – which leads to tonal strings that sound very different among the three languages – a highly natural, yet rather superficial one. What recommends their representation of High domains is that it captures a basic fact about these three related languages, namely that the initial and final locations of Nguni High tones are all determined by the same principles, despite the superficial differences.

[12] The Durban Zulu forms are from Laura Downing, as cited by Cassimjee and Kisseberth (2001); the Phuthi forms are from Donnelly (1997).

(47) m a w a s a

\|/ V

[+nas][−nas]

McCarthy opts for the format in (47) for the convenience of using just one line per candidate in an OT tableau. Autosegments are not being abandoned here but are acquiring more structure – namely, heads.[13]

McCarthy's approach is extended to tone with revisions by Key (2006). Volk (2008) observes that ODT differs from Headed Span Theory in a least one important respect: the former but not the latter sanctions representations in which a TBU is linked with several non-contiguous surface tones. This is also a key difference between ODT and standard autosegmental practice.

Another question that requires further study, attributed to Lee Bickmore (McCarthy 2004: 5, fn. 5), is floating features. The version of FAITHHEADSPAN(●F) in (46) does not apply to floating elements in the input, but the fix may be simple, as McCarthy suggests: a reformulation along the lines suggested by Zoll (1996) and Lombardi (1998).

However, for output floating elements the solution is not so easy: a revision of any analyses that appear to require them. But any revisions would need to capture generalizations that floating tones famously have made so transparent. For example, tone assignment in Tiv normally places no more than one tone on any given TBU (Pulleyblank 1986). But in phrase-final position, a falling tone can appear on a single TBU, as shown by the two variants in (48), from Arnott (1964) (the arrow denotes downstep):

(48) *Tiv* (Arnott 1964)

mbá$^{\downarrow}$ kásév 'there are women'

[‾ – –]

kásév mbâ 'there are women'

[– – \ ]

A floating Low tone at the end of /mba/ accounts naturally for its downstepping effect on the following word and is motivated independently by the high fall on /mba/ before pause.

A more complex case for surface floating tones comes from Asongwed and Hyman's (1976) analysis of Ngamambo (Cameroon, Grassfields Bantu). They show that the five surface tone levels of Ngamambo, shown in (49), correspond to two surface tones, High and Low, which can be either docked on TBUs or floating in output forms. The hyphens at the beginning and end of the forms indicate whether they are a prefix or a root.

(49) *Ngamambo's five tone levels* (Asongwed and Hyman 1976)

High	mə́-
Mid	ʌ̄-
Lower Mid	ʌ-
Unreleased Low	-bàp˺
Low	ʌ̀-

[13] Finlay (2008) notes a similar approach, the Headed Feature Domains Theory of Smolensky (2006), which "makes slightly different restrictions on the representations of feature spans."

The Mid tone above is the alternant of High tone that appears on vowel-initial prefixes. The unusual number of floating tones in outputs helps to explain the complex morphotonemics of this language. For example, there are two types of High-toned noun. In isolation, both have the same tonal realization, but they differ systematically when a noun possessor follows:

(50) a. mə́ɲíp 'water' mə́ɲíp Àbò 'water of slave'
b. mə́fɔ́m 'fat' mə́fɔ̂m Àbò 'fat of slave'

Asongwed and Hyman account for this by positing a High tone sequence for roots like (50a), while the High tone sequence in (50b) is followed by a floating Low.[14]

Larry Hyman (personal communication) adds that floating tones are also a problem in ODT, again due to faithfulness requirements. Clearly, floating tones have a special role to play in this discussion. Their existence is practically predicted by autosegmental theory in assigning tone to an independent tier. And we have seen clear parallels between standard autosegmental phonology and the newer domain-based alternatives. The problem seems to be in formulating a version of faithfulness that will allow tone either to be associated with a position or to float. See also CHAPTER 45: THE REPRESENTATION OF TONE.

ACKNOWLEDGMENTS

Thanks to the editors and to two anonymous reviewers for essential help in revising and reshaping this chapter from the ground up. Special mention to Larry Hyman for dozens of specific suggestions and corrections based on an early draft of this chapter and to Chuck Kisseberth for corrections under severe time pressure to the account of Optimal Domains Theory.

REFERENCES

Akinlabi, Akinbiyi & Mark Liberman. 2001. Tonal complexes and tonal alignment. *Papers from the Annual Meeting of the North East Linguistic Society* 31. 1–20.
Anderson, Stephen R. 1969. West Scandinavian vowel systems and the ordering of phonological rules. Ph.D. dissertation, MIT.
Anderson, Stephen R. 1974. *The organization of phonology*. New York: Academic Press.
Anderson, Stephen R. 1976. Nasal consonants and the internal structure of segments. *Language* 52. 326–344.
Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

[14] Example (50) simplifies the facts, and a fuller account makes an even stronger case for output floating tones. The floating tone in (50b) appears to show up directly on the corresponding syllable in the possessed–possessor forms, but in fact the situation is more complex. In between the two nouns is an associative marker which itself may be a floating tone. That tone in turn is determined by what precedes, and notably by the final tone of the preceding noun, whether that tone is floating or not. See the charts in Asongwed and Hyman (1976: 48, 50) for examples of the dozens of possible combinations of disyllabic N_1, associative marker, and disyllabic N_2. Without floating tones in the output, it might be necessary to posit a set of arbitrary declension-like classes to account for the possibilities.

Arnott, David W. 1964. Downstep in the Tiv verbal system. *African Language Studies* 5. 34–51.

Asongwed, Tah & Larry M. Hyman. 1976. Morphotonology of the Ngamambo noun. In Larry M. Hyman (ed.) *Studies in Bantu tonology*, 24–48. Los Angeles: Department of Linguistics, University of Southern California.

Beckman, Mary E. & Gayle Ayers Elam. 1994. Guidelines for ToBI labeling. Available (August 2010) at www.ling.ohio-state.edu/~tobi/.

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.

Bickmore, Lee. 2000. Downstep and fusion in Namwanga. *Phonology* 17. 297–331.

Bird, Steven & D. Robert Ladd. 1991. Presenting autosegmental phonology. *Journal of Linguistics* 27. 193–210.

Buckley, Eugene. 1990. Edge-in association and OCP "violation" in Tigrinya. *Proceedings of the West Coast Conference on Formal Linguistics* 9. 75–90.

Cassimjee, Farida & Charles W. Kisseberth. 1998. Optimal Domains Theory and Bantu tonology: A case study from Isixhosa and Shingazidja. In Larry M. Hyman & Charles W. Kisseberth (eds.) *Theoretical aspects of Bantu tone*, 33–132. Stanford: CSLI.

Cassimjee, Farida & Charles W. Kisseberth. 2001. Zulu tonology and its relationship to other Nguni languages. In Shigeki Kaji (ed.) *Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese accentology and other topics*, 327–359. Tokyo: ILCAA.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1977. The autosegmental treatment of vowel harmony. In Wolfgang U. Dressler & Oskar E. Pfeiffer (eds.) *Phonologica 1976*, 111–119. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. & Kevin C. Ford. 1979. Kikuyu tone shift and its synchronic consequences. *Linguistic Inquiry* 10. 179–210.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Cole, Jennifer & Charles W. Kisseberth. 1995. Nasal harmony in Optimal Domains Theory. *Cognitive Science Technical Report* UIUC-BI-CS-95-02, University of Illinois (ROA–49).

Donnelly, Simon. 1997. Aspects of tone and voice in Phuthi. Ph.D. dissertation, University of Illinois.

Duanmu, San. 1993. Rime length, stress, and association domains. *Journal of East Asian Linguistics* 2. 1–44.

Duanmu, San. 1999. Metrical structure and tone: Evidence from Mandarin and Shanghai. *Journal of East Asian Linguistics* 8. 1–38.

Edmondson, Thomas & John T. Bendor-Samuel. 1966. Tone patterns of Etung. *Journal of African Linguistics* 5. 1–6.

Finlay, Sara. 2008. Formal and cognitive restrictions on vowel harmony. Ph.D. dissertation, Johns Hopkins University.

Firth, J. R. 1948. Sounds and prosodies. *Transactions of the Philological Society*, 127–152. Reprinted in F. R. Palmer (ed.) 1970. *Prosodic analysis*, 1–26. London: Oxford University Press.

Goldsmith, John A. 1975. Tone melodies and the autosegment. *Ohio State Working Papers in Linguistics* 20. 135–147.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT. Published 1979, New York: Garland.

Goldsmith, John A. 1977. English as a tone language. *Communication and Cognition* 11. 453–476.

Key, Michael. 2006. Headed spans and Bantu tonology. In Leah Bateman, Michael O'Keefe, Ehren Reilly & Adam Werle (eds.) *Papers in Optimality Theory III*, 187–207. Amherst: GLSA.

Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10. 421–441.

Ladd, D. Robert. 2008. Review of Jun (2005). *Phonology* 25. 372–376.

Leben, William R. 1971. Suprasegmental and segmental representation of tone. *Studies in African Linguistics*. Supplement 2. 183–200.

Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT. Published 1980, New York: Garland.

Leben, William R. 1977. Length and syllable structure in Hausa. *Studies in African Linguistics*. Supplement 7. 137–143.

Leben, William R. 1980. A metrical analysis of length. *Linguistic Inquiry* 11. 497–509.

Leben, William R. 2002. Tonal feet. In Ulrike Gut & Dafydd Gibbon (eds.) *Typology of African prosodic systems*, 27–40. Bielefeld: University of Bielefeld.

Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.

Liberman, Mark. 1975. The intonational system of English. Ph.D. dissertation, MIT.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.

Liberman, Mark & Ivan A. Sag. 1974. Prosodic form and discourse function. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 10. 416–427.

Lombardi, Linda. 1998. Evidence for MaxFeature constraints from Japanese. *University of Maryland Working Papers in Linguistics* 7 (ROA-247).

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.

McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12. 373–418.

McCarthy, John J. 1982. Nonlinear phonology: An overview. *GLOW Newsletter* 8. 63–77.

McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17. 207–263.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John J. 2004. Headed spans and autosegmental spreading. Unpublished ms., University of Massachusetts, Amherst (ROA-685).

McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague & Paris: Mouton.

McCawley, James D. 1970. A note on tone in Tiv conjugation. *Studies in African Linguistics* 1. 123–129.

Myers, Scott. 1997. OCP effects in Optimality Theory. *Natural Language and Linguistic Theory* 15. 847–892.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Odden, David. 1986. On the role of the Obligatory Contour Principle in phonological theory. *Language* 62. 353–383.

Odden, David. 1988. Anti antigemination and the OCP. *Linguistic Inquiry* 19. 451–475.

Odden, David. 1995. Tone: African languages. In Goldsmith (1995), 444–475.

Oostendorp, Marc van & Anthi Revithiadou. 2005. Quasi-opacity and headed spans in Silly and Megisti Greek. Paper presented at the 13th Manchester Phonology Meeting.

Padgett, Jaye. 1995. *Stricture in feature geometry*. Stanford: CLSI.

Paulian, Christiane. 1975. *Le kukuya: Langue teke du Congo*. Paris: Societé d'Etudes Linguistiques et Anthropologiques de France.

Pearce, Mary. 2006. The interaction between metrical structure and tone in Kera. *Phonology* 23. 259–286.

Pierrehumbert, Janet B. 1980. The phonology and phonetics of Japanese intonation. Ph.D. dissertation, MIT.

15 Glides

Susannah V. Levi

1 Introduction

The division of sounds into two major classes – consonants and vowels – is challenged by a set of sounds that exhibits characteristics of both categories. Variably referred to as "glides," "semi-vowels," or "semi-consonants," these *intermediate* segments require an examination of those properties that differentiate vowels and consonants.[1] Glides have appeared problematic for phonological theories because of their (apparent) variable behavior. Glides can pattern with vowels – even alternating with them in some languages – and can also pattern with consonants. This variable behavior has fueled a debate over the existence of glides as distinct phonemes that contrast with vowels. This chapter will present a typology of vowel-glide systems in which glides can exist both as distinct phonemes and as allophonic variants of vowels (hereafter "phonemic glides" and "derived glides," respectively). Glides derived from vowels behave phonologically like vowels, while phonemic glides pattern with consonants to the exclusion of vowels (Levi 2004), eliminating the *variability* in behavior. Herman (1994) and Hume (1995) argue for a distinction between "consonantal" and "vocalic" glides based on these different behaviors. In this chapter, the distinction between glides will be grounded in their phonological status as allophones of vowels or as distinct phonemes. The major questions surrounding glides relate to their phonological status.

i. Can glides be distinct phonemes different from vowels or do they exist only as allophones of vowels, based on their phonological environment and/or position in the syllable?
ii. If glides can be phonemic, do they behave like vowels or consonants, or do they instead form a distinct third category of segments that has a unique behavior different from either vowels or consonants?
iii. If they can be phonemic, how should they be represented? Is this representation more similar to vowels or consonants? Is the representation universal?
iv. How does the phonological status (phonemic or derived) relate to the surface phonetic realization? What does it mean to be a glide on the surface? What does it mean to be a glide phonologically?

[1] In some frameworks, glides form part of the class of approximants, along with liquids (Ladefoged 2006). As this term refers to a larger class of sounds, it will not be used here.

Answering questions (i)–(iii) is crucial to all theories of phonology. Whether using a derivational *SPE*-style analysis (Chomsky and Halle 1968), a feature-geometric representation, a government phonology approach, or an Optimality Theory (OT) approach, it is necessary to establish the featural make-up of glides to determine whether phonological processes can or should apply or whether constraints – often formulated in relation to a particular phonological feature – are violated. Question (iv) will be discussed briefly in §4.

An example of predictable glide behavior will illustrate some of the issues surrounding the status of glides. Latin is a classic example of a language in which surface glides and vowels are found in predictable phonological environments. High vowels surface as glides in word-initial position when they occur before another vowel (1a) and intervocalically (1b) (Steriade 1984). High vowels surface as vowels after consonants (2).

(1) *Glide realization*

a.	/iecur/	[je.kur]	'liver'
	/uenio/	[we.ni.oː]	'I come'
b.	/ouis/	[o.wis]	'sheep'
	/auus/	[a.wus]	'grandfather'
	/iuuenis/	[ju.we.nis]	'young'

(2) *Vowel realization*

/mulier/	[mu.li.er]	'woman'
/dies/	[di.eːs]	'day'
/mutuus/	[mu.tu.us]	'mutual'

Languages such as Latin that show this type of predictable behavior have been used to argue that glides are merely positional variants of vowels (see §3.1 and §3.2). §3.4 presents a typology of vocoid systems that include both derived and phonemic glides. Before examining the debate and typology, we first provide an overview of how glides have been discussed in the phonetic and phonological literature (§2). Following the typology section (§3), we conclude in §4 with a brief discussion of the issues that remain.

2 A brief history of glides: Phonetics and phonology

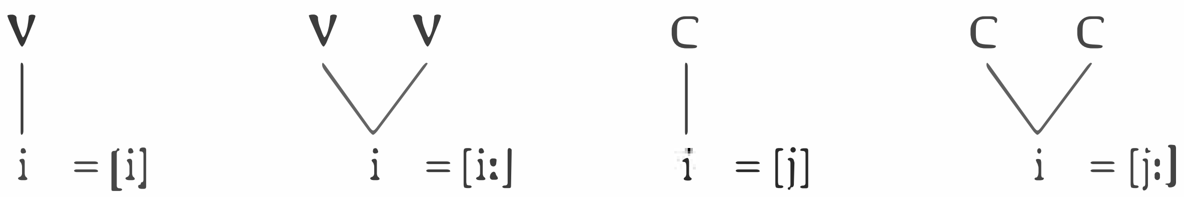
2.1 *Phonetics*

As with all issues involving glides, there is little consensus as to how they should be characterized phonetically. The most commonly discussed phonetic properties of glides – and how they compare to vowels – are degree of constriction, duration, amplitude, and even syllable position. Discussion of constriction as a distinguishing phonetic correlate of glides dates back over 100 years. Whitney (1865: 239) describes glides in English as "vowels of so close a position that they verge upon the consonants" (see also CHAPTER 21: VOWEL HEIGHT). Sweet (1877, 1907) targets constriction as the defining difference between consonantal and vocalic sounds which eventually provides the basis for the features [consonantal] and [vocalic]

With the reintroduction of the syllable in phonological theory came a push to eliminate the feature [syllabic]. Kaye and Lowenstamm (1984) argue that syllabification processes, not features such as [syllabic], determine syllable position. In their view, glides and vowels are featurally identical ([−consonantal, +vocalic]) and should *never* be distinct segments. They use the symbols /I/ and /U/ to highlight that syllabification determines whether they surface as a glide or a vowel.

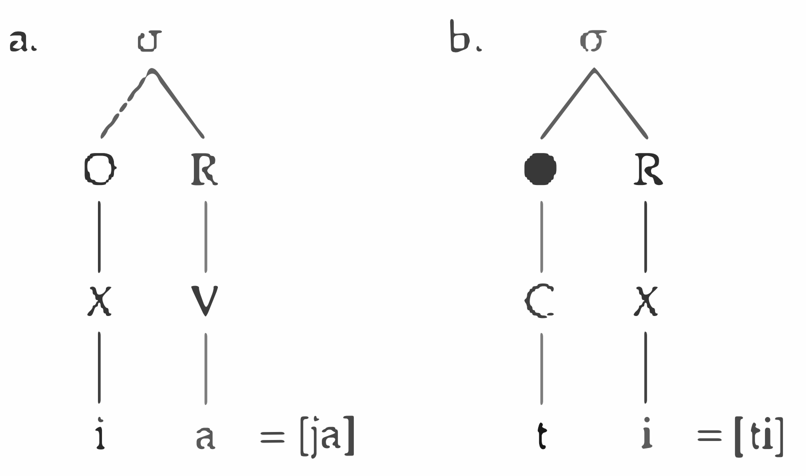
In theories that target the syllable, it is structure, not features, that distinguishes glides from vowels. Clements and Keyser (1983) introduce the CV tier between the segmental tier and the syllable tier, and eliminate the feature [syllabic]. The difference between vowels and glides is thus determined by whether the segment is dominated by a C or a V node, as in (4).

(4)



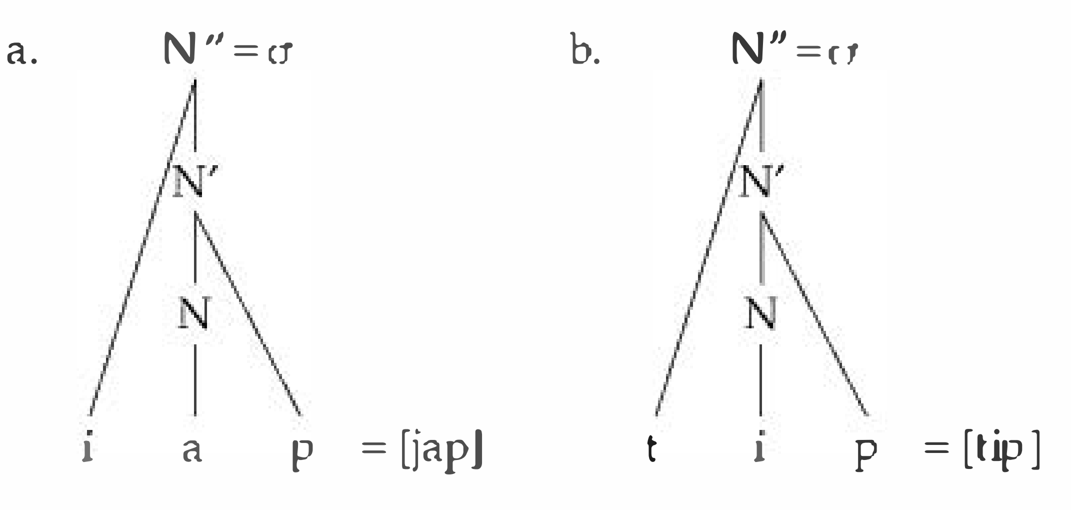
One criticism of CV theory was that it neglected to capture the empirical generalization that syllable position is largely predictable, making the CV tier redundant (Levin 1985). In a move to eliminate some of this redundancy, Steriade (1984) added X as a possible element to the CV tier to account for the predictability of glide/vowel alternations in Latin. This representation allows high vowels to alternate between peak and non-peak elements in the syllable without requiring rules that change skeletal slots (e.g. C → V). In Steriade's proposal, vowels and glides surface as in (5).

(5)

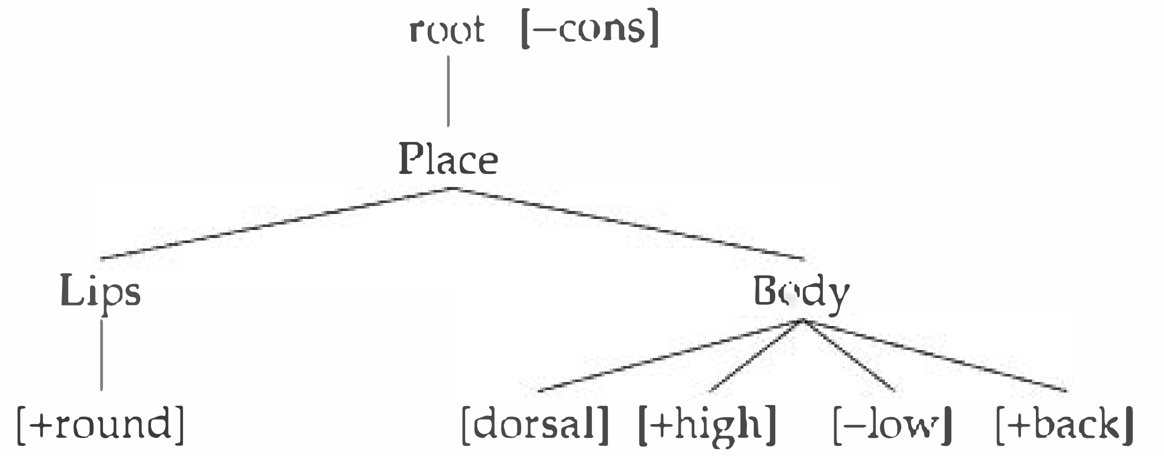


Noting the redundancy of the CV tier in general, Levin (1985) eliminated CV in favor of X for all skeletal positions. Non-peak elements are linked either to N″ (onsets) or to N′ (codas). The difference between vowels and glides thus remains on a higher structural level.

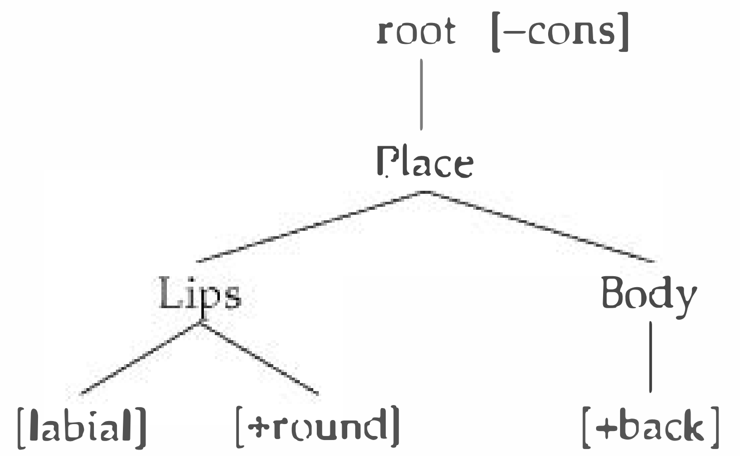
(6)



(10) a. /u/ (*and derived* [w])



b. /w/



3 The debate: Is there evidence for a phonemic distinction between vowels and glides?

Despite abundant work on glides, vowels and their alternations, there remains no consensus on the existence of glides as distinct phonemes (CHAPTER 11: THE PHONEME). There are two common ways glides are treated in the literature. The first is to assume that phonemic glides do not exist and that glides are only allophones of vowels. In this view, glides are treated as "high vowels in a non-syllabic disguise" (Durand 1986), and are only distinct on a syllabic level (e.g. Jakobson *et al.* 1952; Clements and Keyser 1983; Kaye and Lowenstamm 1984; Durand 1986; Deligiorgis 1988; Rosenthall 1994). The second treatment of glides has been to remain agnostic about their phonological status. Despite the rather prominent view that glides are only positional variants of vowels, many works include glides in consonant inventories.

As will be shown below, the confusion surrounding glides (i.e. their phonological status and the conflicting phonological evidence) can be attributed to the existence of two distinct types of glides. Specifically, glides can either be derived from vowels and thus be positional variants of vowels ("derived glides") or they can exist as distinct phonemes ("phonemic glides"). One reason glides have not been understood and have been viewed as behaving differently from other segments (e.g. forming the class of [0 consonantal] segments; see CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION) is that their *variable* behavior has been attributed to a single class of segments. Recognizing that glides can be derived from two distinct sources eliminates the variable behavior and the "duality" of glides.

3.1 Predictable alternations between glides and vowels

A common type of language is one in which glides surface in predictable environments, and are thus positional variants of underlying vowels. Latin (discussed in §1)

The large proportion of predictable cases compared to the small proportion of lexically nuclear vowels is expected in an analysis that treats the "special" vowels as marked (Harris and Kaisse 1999: 185).

Data from Korean also provide evidence for both predictable glide/vowel alternations and unpredictable vowels. Unmarked, underlying high vowels surface as vowels between consonants (e.g. [ʧi-ta] 'lose') and as glides intervocalically (e.g. [moj-ət-ta] 'gathered') (Yun 2003). In C __ V environments there is variability; some lexical items allow a glide while others require a vowel, generating hiatus as in (16). Forms that surface with a vowel in gliding environments (16b) are rare; only six out of 428 stems exhibit this pattern (Yun 2003). Here C' represents a fortis consonant (CHAPTER 111: LARYNGEAL CONTRAST IN KOREAN).

(16) a. *Gliding allowed*

/p^{h}i-/	[p^{h}i.ta]	[p^{h}jət.ta]	*[p^{h}i.ət.ta]	'straighten'
/ʧi-/	[ʧi.ta]	[ʧjət.ta]	*[ʧi.ɔt.ta]	'lose'
/ʧ'i-/	[ʧ'i.ta]	[ʧ'jət.ta]	*[ʧ'i.ət.ta]	'boil'

b. *Gliding disallowed*

/pi.-/	[pi.ta]	[pi.ət.ta]	*[pjət.ta]	'be empty'
/p'i.-/	[p'i.ta]	[p'i.ət.ta]	*[p'jət.ta]	'sprain'
/p^{h}i.-/	[p^{h}i.ta]	[p^{h}i.ət.ta]	*[p^{h}jət.ta]	'bloom'

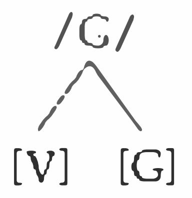
Analyses of these "vocalic" languages represent alternating vocoids like those from predictable languages in the previous section. Vowels that surface in *unexpected* environments are considered to be lexically marked as nuclear (e.g. Guerssel 1986; Roca 1997; Harris and Kaisse 1999). Levin (1985) argues that the property of being a syllable head could be specified underlyingly. In this chapter, this lexical marking is represented as /i=N/ or /u=N/, indicating that they must surface as syllabic. Because these segments are lexically marked, they are expected to be less numerous than those that exhibit the basic alternating pattern. Indeed, this marginal status is confirmed by Spanish and Korean.

Evidence suggesting that alternating segments and lexically marked vowels are featurally identical (though structurally distinct) comes from the Pasiego dialect of Spanish. In Pasiego, stressed high vowels cause harmonic raising of preceding mid vowels (Penny 1969a, 1969b; Hualde 1989), as in [be'ber] 'to drink' – [bi'bi:s] 'you (PL) drink (INDIC)'. Raising is also triggered by derived glides in stressed syllables, as in [amfi'θjon] 'infection' (cf. [amfes'tar] 'to infect') (Kaisse and Levi 2004). Finally, lexically nuclear vowels, also trigger harmony, as in [bi'bi.a] 'he was drinking (INDIC)' (Penny 1969b) and [kuxi'ri.a] 'I would take' (Hualde 1989). That derived glides, plain high vowels, and lexically marked high vowels trigger raising suggests a featural identity between these segments. Recognizing that surface glides are derived from vowels, and that the unusual vowels have marked structure but identical features unifies the analysis of vowel harmony to a single feature bundle associated with high vowels.

3.3.2 *"Glide" languages*

The previous sections provided evidence for languages with predictable realization of vowels and glides, as well as unpredictable distributions in which vowels occur in unexpected environments. There are also languages that exhibit glides in environments where vowels are expected. These languages with phonemic glides will be

(23) *Type IV*

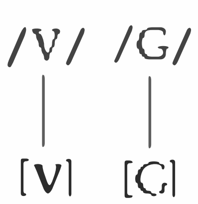


Thus far, no Type IV languages have been identified. As discussed in §3.4.2, languages that lack underlying high vowels are rare. The lack of Type IV (and also Type VII) suggests a possible implicational relationship: if phonemic glides alternate, there must also be underlying vowels that alternate. Determining whether a language is Type III or IV would require an examination of the phonology to consider whether other phonological processes could adjudicate between these two possible systems. For example, a Type III language with vowel harmony should show participation of both vowels and glides (as discussed for Pasiego), while a Type IV language should show transparency of vowels and glides. Other phonological phenomena that target featural, rather than syllabic distinctions, would also establish a difference between Type III and Type IV languages.

3.4.5 *Type V: Non-alternating underlying vowel and glide*

Type V languages contain both phonemic glides and vowels that do not alternate. Type V is the second type of language with phonemic glides. Yawelmani, Tahltan, and Turkish are languages of this type (Levi 2004).

(24) *Type V*



Yawelmani syllabification provides evidence for the absence of alternations between vowels and glides. Vowel hiatus is banned and repaired by [ʔ]-insertion and vowel deletion, as in (25). That gliding is not a legal repair for hiatus shows that high vowels cannot surface as glides (data from Newman 1944).

(25)	a.	[ʔ]-*insertion*		
		/ṣaṣa:-in/	[ṣaṣa:-ʔ-in]	'eye-ross'
		/pana:-it̪*'/	[pana:-ʔ-it̪*']	'one who is arriving'
	b.	*Vowel deletion*		
		/wili:l-d̪a:-ihni:/	[wilal-d̪-ihniʔ]	'one always preparing to depart'
		/paxju:-ila:-hin/	[paxj-ula:-hin]	'caused something to scatter'

Consonant clusters are also prohibited. Thus, in clusters involving glides, the only legal repair is vowel epenthesis as in (26), and vocalization is prohibited.

(26)	/ʔutj-t/	[ʔutj-u-t]	'fall-AORIST'	*[ʔuti-t, ʔutu-t]
	/logw-k'a/	[logiw-k'a]	'pulverize-IMP'	*[logu-k'a]

Two other types of evidence suggest an underlying contrast between vowels and glides in Yawelmani. First, all consonants and glides have plain and glottalized series (represented as [C']) (see (27)), whereas vowels do not.

(27) ṭ'ojix 'applying medicine' (Newman 1944: 19)
ṭ'oj'ix 'getting rusty' (Newman 1944: 19)
ṭ'aw' 'throw' (Archangeli 1984: 285)

A second type of evidence for a phonological contrast between vowels and phonemic glides comes from vowel harmony. In Yawelmani, glides are transparent to vowel rounding harmony (see CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS).

(28) *Glide as a non-trigger*

/boːwn-/	[bown-iṭ]	'trap-PASS AORIST'	(Archangeli 1984: 263)
	*[bown-uṭ]		
/joloːw-/	[joloːw-hin]	'assemble-AORIST'	(Kisseberth 1969: 24)
	*[joloːw-hun]		

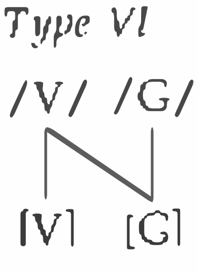
Evidence from hiatus and cluster resolution, sonorant series, and vowel harmony indicates the presence of contrasting segments (vowel and phonemic glide) which do not show surface alterations.

Several other languages exhibit Type V patterns. In Tahltan, glides participate in consonant harmony (Hardwick 1984; Shaw 1991) while vowels do not, suggesting that vowels and glides are phonologically distinct (see also CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS and CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). Additionally, there is no evidence from syllabification to suggest that these two segments should be treated as a single phoneme. In Turkish, vowels and consonants with secondary articulations participate in vowel harmony (CHAPTER 118: TURKISH VOWEL HARMONY), showing that vowel harmony is not based on syllable position (Levi 2001). In contrast, glides do not participate in vowel harmony, suggesting that they must be featurally distinct from vowels. Furthermore, phonemic glides in clusters force epenthesis (e.g. /mejl/ → [mejil] 'tendency'; Lewis 1967). Like Yawelmani and Tahltan, there is no evidence in Turkish suggesting alternations between vowels and glides.

3.4.6 *Type VI: Alternating underlying vowel and non-alternating underlying glide*

Type VI languages contain a vowel that alternates similar to Type III languages. In addition, Type VI languages contain a phonemic glide that only surfaces as a glide. Languages of this type include Karuk, Sundanese, Pashto, and Pulaar (Levi 2004).

(29) *Type VI*



In Karuk, the realization of an underlying vowel as a vowel or a glide is predictable based on syllabification parameters in the language. Tautosyllabic consonant clusters are banned, making the maximal syllable template CVC (Bright 1957). Thus, all medial CC sequences must be syllabified as VC.CV. The absence of sequences

The difference between phonemic and derived glides is also apparent in their behavior before vowel-initial suffixes. Phonemic /w/ patterns with /r/ and surfaces before vowel-initial suffixes as in (33a) and (33b). In contrast, morpheme-final derived glides pattern with other vocalic segments, exhibiting deletion and coalescence before vowel-initial suffixes, as in (33c).

(33)	a.	*Phonemic glide /w/*				
		/ikiiw/	→	[ikjiw]	'to fall'	
				[ikjíw-iʃ(rih)]	'to fall down'	
				[ikjiw-ûɪr]	'to fall for a long time'	
		/iw/	→	[ʔiw]	'to die'	
				[ʔíw-apuh]	'dead'	
			cf.	[ʔíːm-kara]	'to drown'	
	b.	*Other consonants*				
		/isʧur-ahi/	→	[iʃʧur-ahi]	'to be cracked'	
		/paθrih-a/	→	[paθrih-a]	'rain (N)'	
	c.	*Final* /u/ → [w]				
		/ikriu/	→	[ikriw] 'to sit'		
				/ikriu-is(rih)/ →	[ikrîːʃ(rih)]	'to sit down'
					*[ikriw-iʃ(rih)]	
				/ikriu-at/ →	[ikrêːt]	'lived'
					*[ikriw-at]	
		/ihiiu/	→	[ihjiw] 'to shout'		
				/ihiiu-unis/ →	[ihjûːniʃ]	'to shout at'
					*[ihjiw-uniʃ]	

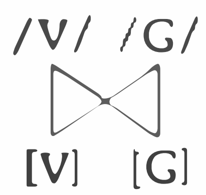
Other Type VI languages include Sundanese, Pashto, and Pulaar. In Sundanese, segments are divided into those that undergo nasal harmony and those that do not (Robins 1957; Cohn 1990, 1993). As expected, derived glides pattern with vowels in undergoing nasal harmony (e.g. /ɲiar/ → [ɲĩj̃ãr] 'seek-ACTIVE'). In contrast, phonemic glides block nasal harmony (e.g. /ŋajak/ → [ŋãjak] 'sift-ACTIVE'). An analysis of Sundanese that recognizes a distinction between two types of glides and the relationship between derived glides and their vowel source on the one hand and phonemic glides on the other simplifies the harmony system (Levi 2008a). In Pashto, glides surface in reverse sonority clusters (e.g. [wradz] 'day' (Penzl 1955); see also CHAPTER 49: SONORITY), where they would otherwise be unexpected if they were derived from vowels (Levi 2004). In other environments, glides and vowels surface predictably, suggesting that vowels can surface as either vowels or glides. In Pulaar, glides differ in how they behave in consonant alternations (Paradis 1992); one set of labial glides alternates with labial stops while another set alternates with velar stops, suggesting a distinction between two types of glides. Similarly, palatal glides are divided into two sets: those that alternate with velars and those that alternate with palatal stops. Using evidence from consonant gradation, vowel harmony, vowel epenthesis, and gemination, Levi (2004) argues that glides that alternate with labial or palatal stops are phonemic, while those that alternate with velar stops are derived.

French is another language that is often cited as including both derived and phonemic glides (Spence 1971; Valdman 1976; Walter 1977; Noske 1996; Hannahs (i.e. glides surfacing as vowels) as different from underlying vowels (CHAPTER ##: DERIVED ENVIRONMENT EFFECTS).

3.4.8 *Type VIII: Alternating underlying vowels and glides*

Type VIII languages show a balanced pattern where both phonemic glides and vowels can surface as vowels or glides. Type VIII languages fit the generalization mentioned in §3.4.4 that a phonemic glide can only alternate if the vowel also alternates. Imdlawn Tashlhiyt Berber is an example of this type of language (Levi 2004).

(36) *Type VIII*



Evidence for predictable realization of underlying vowels comes from syllabification, as in (37) (Elmedlaoui 1985: 27).

(37)	/i-kti/	[i-kti]	'he remembered'
	/i-ura/	[j-ura]	'he wrote'

Evidence from syllabification of other segments and from sonority plateaux suggests a systematic pattern (Dell and Elmedlaoui 1985), where the first member of a plateau surfaces as syllabic (38) (Elmedlaoui 1985: 39).

(38)	/i-sui/	[i-suj]	'he passed'
	/t-ikiu-t/	[t-ikiw-t]	'kind of plant'

Some words contradict this pattern. In these cases, the second of two vocoids surfaces as syllabic. Following Dell and Elmedlaoui, Levi (2004) analyses the first high vocoid of these sequences as a phonemic glide, making the sequence a sonority rise, rather than a plateau. Thus, the difference between the forms in (38) and (39) is that the former contain /CVV/, while the latter contain sequences of /CGV/ (Elmedlaoui 1985: 42).

(39)	/i-swi/	[i-swi]	*[isuj]	'excrement'
	/aˁ-kˁjˁuˁdˁ/	[aˁ-kˁjˁuˁdˁ]	*[aˁkˁiˁwˁdˁ]	'braid (hair)'

Imdlawn Tashlhiyt Berber is especially interesting because it allows any segment to surface as syllabic, including phonemic glides (Elmedlaoui 1985). Data in (39) and (40) show that phonemic glides can surface as glides or vowels.

(40)	/nw/	'be cooked'	[nwa]	PERFECTIVE
			[nwi]	NEGATIVE
			[nu]	AORIST

Taken together, these data provide evidence for a high vowel that surfaces as either a vowel or a derived glide and for a phonemic glide that surfaces as a glide or

Emeneau, M. B. & B. A. van Nooten. 1968. *Sanskrit sandhi and exercises.* Berkeley & Los Angeles: University of California Press.

Grammont, Maurice. 1933. *Traité de phonétique.* Paris: Delagrave.

Guerssel, Mohamed. 1986. Glides in Berber and syllabicity. *Linguistic Inquiry* 17. 1–12.

Halle, Morris. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26. 1–46.

Halle, Morris, Bert Vaux & Andrew Wolfe. 2000. On feature spreading and the representation of place of articulation. *Linguistic Inquiry* 31. 387–444.

Hannahs, S. J. 1998. Characterizing French glides. Paper presented at the 28th Linguistics Symposium on Romance Languages, Pennsylvania State University.

Hardwick, Margaret F. 1984. Tahltan morphology and phonology. M.A. thesis, University of Toronto.

Harris, James W. & Ellen M. Kaisse. 1999. Palatal vowels, glides and obstruents in Argentinian Spanish. *Phonology* 16. 117–190.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Herman, Rebecca. 1994. La double vie de [w] or the status of [w] in Karuk. *Studies in the Linguistic Sciences* 24. 233–244.

Hualde, José Ignacio. 1989. Autosegmental and metrical spreading in the vowel-harmony systems of northwestern Spain. *Linguistics* 27. 773–805.

Hume, Elizabeth. 1995. Representing the duality of glides. In Georges Tsoulas & Léa Nash (eds.) *Actes du congrès: Langues et grammaire.* Paris: Université Paris VIII.

Hume, Elizabeth & David Odden. 1996. Reconsidering [consonantal]. *Phonology* 13. 345–376.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge, MA: MIT Press.

Jones, Daniel. 1918. *An outline of English phonetics.* (8th edn., 1956.) Cambridge: Heffer.

Kaisse, Ellen M. & Susannah V. Levi. 2004. Vowel harmony: Nucleus to nucleus or vocalic node to vocalic node? Paper presented at the 78th Annual Meeting of the Linguistic Society of America, Boston.

Kaye, Jonathan & Jean Lowenstamm. 1984. De la syllabicité. In François Dell, Daniel Hirst & Jean-Roger Vergnaud (eds.) *Forme sonore du langage: Structure des representations en phonologie,* 123–159. Paris: Hermann.

Kisseberth, Charles W. 1969. Theoretical implications of Yawelmani phonology. Ph.D. dissertation, University of Illinois.

Ladefoged, Peter. 2006. *A course in phonetics.* 5th edn. Boston: Thomson Wadsworth.

Levi, Susannah V. 2001. Glides, laterals, and Turkish vowel harmony. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 37. 379–393.

Levi, Susannah V. 2004. The representation of underlying glides: A cross-linguistic study. Ph.D. dissertation, University of Washington.

Levi, Susannah V. 2008a. Reconsidering the variable status of glottals in nasal harmony. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 41. 299–312.

Levi, Susannah V. 2008b. Phonemic vs. derived glides. *Lingua* 118. 1956–1978.

Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.

Lewis, Geoffrey L. 1967. *Turkish grammar.* Oxford: Oxford University Press.

Liberman, Alvin, Pierre Delattre, Louis Gerstman & Franklin Cooper. 1956. Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology* 52. 127–137.

Maddieson, Ian. 1984. *Patterns of sounds.* Cambridge: Cambridge University Press.

Maddieson, Ian. 2008. Glides and gemination. *Lingua* 118. 1926–1936.

Maddieson, Ian & Karen Emmorey. 1985. Relationship between semivowels and vowels: Cross-linguistic investigations of acoustic difference and coarticulation. *Phonetica* 42. 163–174.

Mielke, Jeff. 2005. Ambivalence and ambiguity in laterals and nasals. *Phonology* 22. 169–203.

Montreuil, Jean-Pierre. 1994. On prosodization. In Chantal Lyche (ed.) *French generative phonology: Retrospective and perspectives*, 211–238. Salford: Association for French Language Studies.

Newman, Stanley. 1944. *Yokuts language of California*. New York: Viking Fund.

Noske, Roland. 1996. Is French optimal? In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, vol. 2, 485–507. Salford: ESRI.

O'Connor, J. D., Louis Gerstman, Alvin Liberman, Pierre Delattre & Franklin Cooper. 1957. Acoustic cues for the perception of initial /w, j, r, l/ in English. *Word* 13. 24–43.

Odden, David. 1996. *The phonology and morphology of Kimatuumbi*. Oxford: Clarendon Press.

Padgett, Jaye. 2008. Glides, vowels, and features. *Lingua* 118. 1937–1955.

Paradis, Carole. 1992. *Lexical Phonology and Morphology: The nominal classes in Fula*. New York: Garland.

Penny, Ralph J. 1969a. *El habla pasiega: Ensayo de dialectología montañesa*. London: Tamesis Books.

Penny, Ralph J. 1969b. Vowel harmony in the speech of Montes de Pas (Santander). *Orbis* 18. 148–166.

Penzl, Herbert. 1955. *A grammar of Pashto*. Washington: American Council of Learned Societies.

Pike, Kenneth L. 1943. *Phonetics: A critical analysis of phonetic theory and a technic for the practical description of sounds*. Ann Arbor: University of Michigan Press.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Robins, R. H. 1957. Vowel nasality in Sundanese: A phonological and grammatical study. *Studies in Linguistic Analysis*. 87–103. Oxford: Blackwell.

Roca, Iggy. 1997. There are no "glides," at least in Spanish: An optimality account. *Probus* 9. 233–265.

Rosenthall, Sam. 1994. Vowel/glide alternation in a theory of constraint interaction. Ph.D. dissertation, University of Massachusetts, Amherst.

Rosenthall, Sam. 1997. Metrical influences on syllable structure in Lenakel. *Linguistic Inquiry* 28. 187–193.

Schane, Sanford. 1968. *French phonology and morphology*. Cambridge, MA: MIT Press.

Schein, Barry & Donca Steriade. 1986. On geminates. *Linguistic Inquiry* 17. 691–744.

Shaw, Patricia A. 1991. Consonant harmony systems: The special status of coronal harmony. In Carole Paradis & Jean-François Prunet (eds.) *The special status of coronals: Internal and external evidence*, 125–157. San Diego: Academic Press.

Spence, N. C. W. 1971. The French semi-vowels: A fresh angle? *Lingua* 27. 198–215.

Steriade, Donca. 1984. Glides and vowels in Romanian. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 10. 47–64.

Stetson, Raymond H. 1928. *Motor phonetics*. Haarlem: Société Hollandaise de Science.

Straka, Georges. 1964. A propos de la question des semi-voyelles. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 17. 301–323.

Sweet, Henry. 1877. *A handbook of phonetics*. Oxford: Clarendon Press. Reprinted 1970, College Park, MD: McGrath Publishing.

Sweet, Henry. 1890. *A primer of phonetics*. (3rd edn., 1906.) Oxford: Clarendon Press.

Sweet, Henry. 1907. *The sounds of English: An introduction to phonetics*. 2nd edn., 1910. Oxford: Clarendon Press.

Taff, Alice. 1996. Acoustic correlates of Deg Xinag vowels. *University of Washington Working Papers in Linguistics* 13. 1–32.

Tranel, Bernard. 1987. *The sounds of French: An introduction*. Cambridge: Cambridge University Press.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.

16 Affricates

Yen-Hwei Lin

1 Introduction

An affricate is a single segment that has a complete closure with a fricative or delayed release. Like stops, affricates consist of a closure phase followed by a release phase, but unlike stops, the release of affricates contains additional supra-laryngeal properties. Phonological evidence based on phonotactics and sonorancy is often used to distinguish affricates from homorganic bisegmental stop + fricative sequences. For example, in English no stop + fricative sequences can occur in word-initial or syllable-onset position, but [ʧ] and [ʤ] can, and hence are treated as affricates. Kehrein (2002: 5–6) groups affricates into four types as in (1), but considers them to be stops phonologically.

(1) *The phonetic class of affricates*

   a. Orally centrally released strident affricates: e.g. [ts], [ʧ], [tɕ]
   b. Orally centrally released non-strident affricates: e.g. [pɸ], [tθ], [kx]
   c. Orally laterally released affricates: e.g. [tɬ]
   d. Nasally released affricates: e.g. [pm], [tn], [kŋ][1]

In phonological theory, the formal characterization of affricates has been a much-debated issue (Clements and Hume 1995: 256; see Kim 1997: 23–38 and Kehrein 2002: 11–15 for reviews of different proposals). What is at issue is how affricates should be represented to account for the way they pattern in phonological processes and segmental contrasts. More specifically, what phonological feature(s) should be adopted to represent affricates and how do affricates behave in phonological patterning? First, if affricates are represented as consisting of a stop component followed by a fricative component, one expects affricates to pattern with stops at their left edges and with fricatives at their right edges. Second, if the two components are unordered phonologically, affricates can pattern with stops or fricatives at either edge in phonological processes. Third, if affricates are phonologically just stops, they are expected to form a natural class with stops,

[1] Due to space limitation, I will not discuss nasally and laterally released affricates specifically, and for ease of reference, I will refer to affricates as segments with a fricative release.

but crucially not with fricatives. If affricates are phonological stops, how are they to be differentiated from simple stops? Under the first approach, affricates have the same representation in phonology and phonetics, whereas under the second and third approaches, the phonological and phonetic representations are different, which leads to the question of when the stop–fricative sequence becomes linearly ordered or the fricative release projected.

This chapter describes different positions with respect to these questions and issues through presentation of relevant data and arguments. In terms of the phonological representation of affricates, I classify the approaches to affricates into three major categories: (i) the *Stop Approach*, (ii) the *Affricate Approach*, and (iii) the *Complex Segment Approach*. The Stop Approach can be traced back to Jakobson *et al.* (1952: 24), where affricates are treated as strident stops, specified as [−continuant, +strident]. One problem with this proposal is that not all affricates are strident. Chomsky and Halle (1968: 329) therefore propose to represent all affricates as [−continuant, +delayed release]. This then accords affricates an independent phonological status, contrasting with stops, which are [−continuant, −delayed release], and fricatives, which are [+continuant, −delayed release]. This is the Affricate Approach (see Kehrein 2002: 11). The feature [delayed release] was later replaced by [+continuant] (Campbell 1974), leading to the Complex Segment Approach, in which affricates form a natural class with stops and with fricatives. Since [delayed release] has not been in common use in recent phonological literature, I focus in this chapter on proposals within the Complex Segment Approach and the Stop Approach since the early 1980s.

The proposals within the Complex Segment Approach differ in (i) how the stop and fricative components of affricates are organized and represented, and (ii) whether or not there is an asymmetrical relationship between the two components, and if there is, which component is the head. Within the Stop Approach, affricates are treated as stops with some additional feature(s) (e.g. [+strident]) manifested as part of the fricative release in articulation, and the proposals differ in what those additional features are in distinguishing affricates from simple stops, and whether the fricative release is projected in the phonological or phonetic component.

Some of the proposals also make different predictions regarding *edge effects*, i.e. cases where affricates pattern with stops with regard to rules sensitive to their left edges and with fricatives with regard to rules sensitive to their right edges. The general assumption is that if or when the stop and fricative components are ordered or when the fricative release is projected, affricates are expected to exhibit edge effects.

In what follows, §2 reviews the proposals and data supporting the Complex Segment Approach, §3 presents the evidence and proposals for the Stop Approach, and §4 addresses the question regarding the point in the derivation at which the stop and fricative components are ordered or the fricative release is projected. In §5, additional data relevant to the debate are discussed, and the final section concludes with brief remarks on remaining issues.

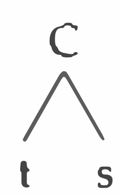
2 The Complex Segment Approach

What I refer to as a complex segment is defined as a single segment with multiple articulations in manner or place features, and hence covers both *contour segments*, such as affricates and prenasal stops, and *complex segments*, such as labial-velars, as proposed by Sagey (1986: §2.4). For Sagey, the multiple articulations in a contour segment are phonologically ordered, but those in a complex segment are not (see also CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION). For our purposes, I make a distinction between complex segments with internal ordering and those without.

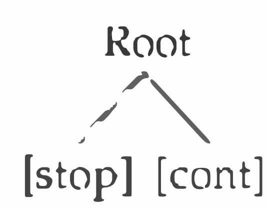
The major proposals under the Complex Segment Approach are summarized in (2) and (3). The *contour segment* analysis in (2a), in which the ordered stop and fricative are linked to one single skeletal slot, and the *contour feature* analysis in (2b), in which the ordered [stop] and [cont] features[2] are linked to a single segmental root node, are analogous to the analysis of contour tones in autosegmental phonology (Goldsmith 1976; CHAPTER 14: AUTOSEGMENTS; CHAPTER 45: THE REPRESENTATION OF TONE); e.g. a falling tone consists of a high tone at the left edge and a low tone at the right edge and is linked to one single tone-bearing unit. Evidence supporting this position comes from data showing edge effects for affricates.

(2) *Affricates as complex segments with internal ordering*

a. Clements and Keyser (1983)



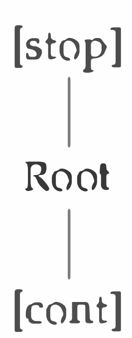
b. Sagey (1985)



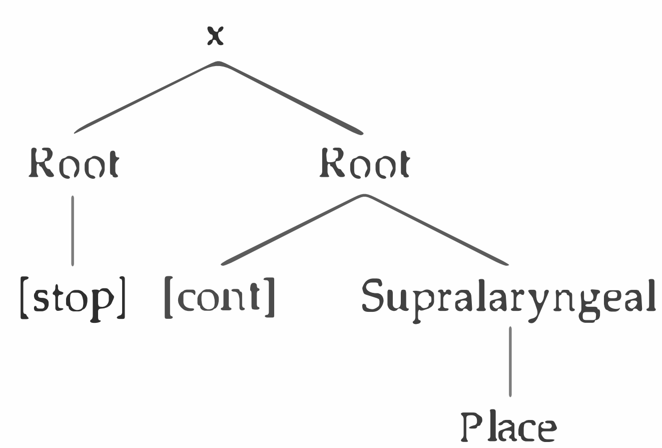
With [stop] and [cont] on separate tiers, the two stricture features are phonologically unordered in the *unordered feature* analyses in (3). Affricates thus do not exhibit edge effects, and can pattern with stops or fricatives on either edge.

(3) *Affricates as complex segments without internal ordering*

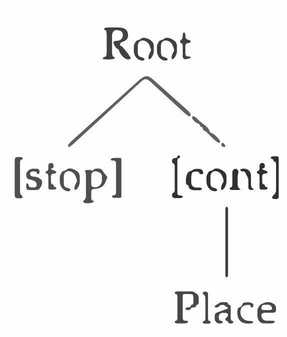
a. Lombardi (1990)



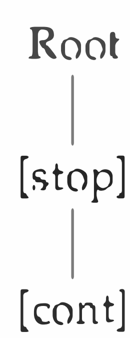
b. Hualde (1991)



c. van de Weijer (1992, 1996)



d. Schafer (1995)



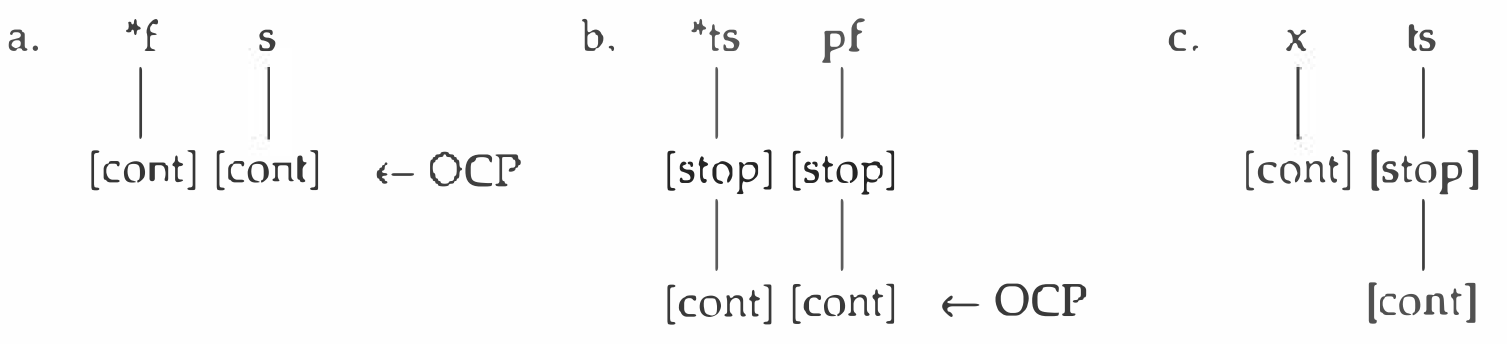
[2] For the rest of this chapter, I use [stop] for [−continuant] and [cont] for [+continuant] for ease of unified exposition in discussing different proposals. See also CHAPTER 13: THE STRICTURE FEATURES; CHAPTER 28: THE REPRESENTATION OF FRICATIVES.

phonemic inventory of Nisgha and the rule is blocked by Structure Preservation (Kiparsky 1985).

Strengthening also shows that affricates pattern with stops. For example, in Tswana, continuants are turned into stops or affricates after a specific nasal morpheme (Schafer 1995: 71): /N+βona/ → [mpona] 'see me', /N+sixa/ → [ntshixa] 'cut me', which is analyzed as spreading [stop] from the nasal to the continuants, resulting in stops or affricates.

Concluding that stops and affricates spirantize, and fricatives strengthen, Schafer argues that encoding the [stop] feature as the head stricture accounts for the patterning of stops and affricates in both strengthening and spirantization, whereas a principled explanation of such patterning is not available in a non-configurational representation.

The second piece of evidence comes from the restrictions on German morpheme-internal consonant clusters (Schafer 1995: 80–84). Fricative clusters are disallowed (e.g. *[f–s], *[s–v], *[ʃ–z]), as are affricate clusters (e.g. *[ts–pf]); however, fricative–affricate clusters are permissible (e.g. [f–ts] in *seufzen* [zoiftsən] 'sigh' and [x–ts] in *jauchzen* [jauxtsən] 'cheer'). As illustrated in (10a) and (10b), the OCP on [cont] can apply only to those [cont] features on the same tier. Since the [cont] feature of the affricate is secondary, it is not on the same tier as the [cont] feature of the fricative (10c). I will come back to the German data in §5.1.

(10)


Schafer demonstrates that affricates persistently pattern with stops phonologically. Assigning [stop] as the head stricture of the affricate means that the proposal actually shares some similarities with the Stop Approach. The main difference is that the [cont] feature is still retained in the representation of the affricate as a secondary stricture feature, and can be referred to by a rule or constraint.

2.3 *Summary and predictions of the Complex Segment Approach*

The different proposals under the Complex Segment Approach share the claim that both [stop] and [cont] are contained in the phonological representation of affricates, but they differ in the relationship between the two stricture features. The major prediction made by the Complex Segment Approach is that affricates can pattern with either stops or fricatives phonologically. If the two features are ordered (§2.1), affricates are expected to show edge effects. If the two features are unordered with equal status (§2.2.1–§2.2.2), affricates can pattern with stops or fricatives on either edge. Under the unordered approach, there are two proposals that claim an asymmetrical relationship between the two stricture features. The first maintains that [cont], but not [stop], dominates place features, which predicts that a rule that targets a [stop] segment with place features does not affect affricates (§2.2.3.1).

(15) *Polish strident assimilation*

a.	[−son, Cor, +ant] → [αPlace] / __ [+strid, αPlace]		
b.	*rozszerzyć*	[s ʃ] ~ [ʃ ʃ]	'broaden'
	okaz cierpliwości	[s tɕ] ~ [ɕ tɕ]	'example of patience'
	podczas działań	[z dʑ] ~ [ʑ dʑ]	'during the action'
c.	*chłopiec szukał*	[ts ʃ] ~ [ʧ ʃ]	'the boy was seeking'
	occie	[ts tɕ] ~ [tɕ tɕ]	'vinegar (LOC SG)'
d.	*twardszy*	[t ʃ] ~ [ʧ ʃ]	'harder'
	oddzielić	[d dʑ] ~ [dʑ dʑ]	'separate'

This can be viewed as an anti-edge effect, in which affricates and fricatives pattern together as triggers in a rule referring to the left-edge context, and hence as a counterexample to the ordered version of the Complex Segment Approach (cf. §2.2.1). This patterning of affricates with fricatives, however, can be attributed to shared [+strident] rather than [cont].

Additional examples illustrating the patterning of affricates with fricatives in terms of stridency include (i) strident dissimilation in Baztan Basque (Kim 1997: 49, 54–55), in which a strident affricate dissimilates to [−strident] after a strident fricative (e.g. /as+tsen/ → [asten] 'to raise'), (ii) vowel insertion between strident fricatives/affricates and -s under English plural formation (§2.2.2; Kehrein 2002: 55–56), and (iii) Hungarian strident coronal place assimilation (Kim 1997: 66–67), which assimilates coronal strident fricatives/affricates in place with the following coronal strident segment: e.g. /egeːs+ʃeːg/ → [ʃ+ʃ] 'health', /egeːs#ʧaːlaːd/ → [ʃ#ʧ] 'entire family', /nolts+ʃaːg/ → [ʧ+ʃ] 'eightness', /makaʧ#tsitsa/ → [ts#ts] 'obstinate kitten'.

3.2 *Re-analyses of cases where affricates are specified as* [cont]

One major obstacle for the Stop Approach is the presence of cases where affricates are claimed to be specified as [cont]. The first type of evidence for specifying affricates as [cont] is based on morpheme structure constraints and consonant harmony involving fricatives and affricates to the exclusion of stops. Recall that for Modern Yucatec Maya morpheme structure restrictions, Lombardi (1990: 389–390) proposes an OCP constraint on [cont] that affects both fricatives and affricates (§2.2.2), arguing that affricates must be specified as [cont]. Kehrein (2002: 46–51; cf. Rubach 1994: 141, fn. 24) suggests that the feature [strident] or Coronal (if coronal stops are unspecified for place) can be used instead to replace [cont] for the analysis, a solution also applicable to cases of sibilant/coronal harmony in languages like Tahltan (Shaw 1991; §3.4 below) and Basque (Hualde 1991; van de Weijer 1996; see §2.2.2).[9]

The second type of evidence comes from stop deletion/debuccalization (§2.2.1). For example, in Basque (5), a stop deletes before a stop, but an affricate loses the [stop] feature and becomes a fricative, and in Yucatec Maya, a stop becomes

[9] As shown in (10) in §2.2.3.2, Schafer (1995) also makes use of the OCP on [cont] to account for German consonant cluster restrictions. However, the feature [+strident] or Coronal cannot be used since the affricate [pf] has neither, and non-permissible affricate clusters such as [ts-pf] cannot be ruled out. I will come back to the German case in §5.1.

[h] and an affricate becomes a fricative before a homorganic stop. A unified account of stop deletion/debuccalization and affricate fricativization involves the deletion of the [stop] feature so that the stop becomes deleted or debuccalized, leaving the affricate with the [cont] feature. Kehrein (2002: 51–55; cf. Rubach 1994: 141, fn. 24) proposes to treat the situation with two independent rules: simple stops delete (or debuccalize), but strident stops (i.e. affricates) spirantize. He also points out that if [stop] deletion were the correct unified generalization, one could imagine a logically possible process of [cont] deletion in which fricatives delete and affricates become stops, and yet such a process is unattested.[10] In addition, the fact that in Yucatec Maya stops become [h] seems to suggest a general spirantization account: stops → fricative [h] and affricates → fricatives. Kim (1997: 46–59) analyzes stop deletion in Baztan Basque as a unified process of spirantization motivated by the OCP constraint on [stop]: both simple stops and strident stops become fricatives through dissimilation or spirantization, but the fricatives derived from simple stops, e.g. [x] from [k], [ð] from [d], and [θ] from [t], are deleted due to Structure Preservation (Kiparsky 1985; CHAPTER 76: STRUCTURE PRESERVATION: THE RESILIENCE OF DISTINCTIVE INFORMATION), i.e. the language does not have these fricatives.

The third type of potential problem concerns the affrication of stops before high vowels, which is typically analyzed as spreading [cont] from the vowel to the preceding stop, creating an affricate with both [stop] and [cont]: e.g. /tat+i+mas+u/ → [tatʃimasu] 'to stand (POLITE PRES)', /tat+u/ → [tatsu] 'to stand (PRES)' in Japanese (Clements 1999: 287; Kim 2001: 90). Under the Stop Approach, processes of affrication are argued to be feature insertion of [+strident] or addition of fricative noise between the stop and the vowel due to a phonology–phonetics mismatch (Kim 1997: 39, 2001: 102; Clements 1999: 289; see also CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). This re-analysis is based on the observation that the release transition from a simple stop to a high vowel has the turbulence and spectral properties similar to the noise of a strident fricative, and can be interpreted as the release of a strident affricate (Kim 1997: chs. 3–4, 2001; Clements 1999: 287–289 and references therein).

In sum, for all three types of data that seem to require the [cont] specification for affricates, proponents of the Stop Approach offer re-analyses claiming that affricates need not bear [cont].

3.3 *Phonological contrasts*

One common argument against treating affricates as strident stops is the fact that not all affricates are strident (Chomsky and Halle 1968: 329; Lombardi 1990: 409), e.g. [tθ] in Tahltan (Shaw 1991) and [pf] in German. Under the Stop Approach, place features such as [distributed], [laminal], or Labial are typically used to specify non-strident stops (e.g. Shaw 1991: 146; Kim 1997: 39; Kehrein 2002: 9–10). A labio-dental affricate like [pf] has sometimes been proposed to be a [+strident] stop, in contrast to a bilabial stop (Kim 1997: 106–108).[11]

[10] An anonymous reviewer comments that fricativization in Basque (Hualde 1991: 134f.; van de Weijer 1992: 142–143), where affricates become stops after fricatives/affricates, comes close to this. This process, however, does not delete fricatives after fricatives/affricates.

[11] This of course raises the question of how the feature [strident] should be defined, which I will not pursue here.

Examination of cross-linguistic phonological contrasts among affricates and stops leads Clements (1999) and Kehrein (2002) to argue that (i) affricates are phonological stops which bear no [cont] feature and are distinguished from non-affricated stops by place features or manner features such as [strident], [lateral], [nasal], and (ii) affrication or fricative release is best accounted for at the phonetic level when the phonological feature combinations are spelled out as sequential phonetic events, which Kehrein (2002: ch. 2) proposes to be a phonetic strategy for increasing the perceptibility of phonological contrasts.

The main finding is that strident, lateral, and nasal affricates contrast with stops, but non-strident affricates do not. Coronal affricates can contrast with stops at the same place of articulation in terms of stridency in languages, e.g. the contrast of simple stop, strident affricate, and strident fricative: /t ts s/ in Hamer and /t ʈʂ ʂ/ in Burushaski (Kehrein 2002: 17; cf. Clements 1999: 278–279), as can lateral and nasal affricates, e.g. the contrast of simple stop, lateral affricate, and lateral sonorant /t tɬ ɬ/ in Navajo and the contrast of simple stop, nasal affricate, and nasal sonorant /b bm m/ in Zing Mumuye (Kehrein 2002: 18–19).

On the other hand, non-strident affricates occur as allophones of stops, or as phonetic realizations of aspirated stops or of stops contrasting within a major place category (Kehrein 2002: 21). That is, languages do not contrast stops and non-strident affricates. Examples of allophonic variation include free variation of [k] and [kx] for some speakers of Thai, and allophonic derivation of bilabial affricates from bilabial stops before [u], e.g. /p/ → [pɸ] /__ /u/ in Lahu (Kehrein 2002: 22). In addition, the aspirated stops in those languages with laryngeal contrasts are likely to show affrication; e.g. in Navajo, the three-way /k k' k^h/ contrast is realized phonetically as [k k' kx^h]. Finally, apparent contrasts between stops and non-strident stops (e.g. /p/ *vs.* /pf/ in German, and /t/ *vs.* /tθ/ in Chipewyan) are contrasts at minor places of articulation, e.g. bilabial *vs.* labio-dental within the major category of Labial in German,[12] and apical [t] *vs.* laminal [tθ] in Chipewyan within the category of Coronal (Kehrein 2002: 23, 25; cf. Clements 1999: 275). Additional examples include palatal /cç/ *vs.* /k/ in Irish and /k/ *vs.* uvular /qx/ in Wolof within the Dorsal category (Kehrein 2002: 25; cf. Clements 1999: 283–285). In some cases, what were claimed to be palatal non-strident affricates in contrast with palatal stops are actually alveo-palatal strident affricates, e.g. /cç/ should be /t̠ɕ/ in Komi (Kehrein 2002: 28–29; Clements 1999: 281–282). Kehrein (2002: 29–30) thus concludes that (i) a stop and its homorganic non-strident affricate are never in contrast, (ii) non-strident affricates are phonetic realizations of their corresponding stops, and (iii) a stop is often realized as an affricate when it is in contrast with another stop within the same primary articulator with small place differences.

3.4 *Natural classes*

In all the analyses of affricates considered above, affricates have a stop component, and hence form a natural class with stops. Many of the examples we have seen earlier fall into this category. The Stop Approach differs from the Complex

[12] As pointed out by an anonymous reviewer, it is unclear which feature should be used to differentiate /p/ and /pf/. If the traditional feature [strident] is not used, an alternative must be considered (Clements 1999: 283).

suggesting an edge effect (cf. Piro §5.2 below). The solution offered by Schafer is that *affricate–affricate clusters are ruled out in the phonology when [stop] and [cont] are still unordered, but the lack of the *affricate–fricative clusters occurs in the phonetic component when the secondary [cont] stricture becomes ordered after [stop]. It is not specified, however, whether there is independent evidence in German suggesting that some restrictions must be phonological and some must be phonetic.

5.2 *Piro obstruent co-occurrence restrictions*

Following up on Steriade's (1989) initial analysis of consonant clusters in Piro (see also Kenstowicz 1994: 502–503), Lin (2005) conducts a comprehensive examination of Piro consonant co-occurrence restrictions, based on the data in Matteson (1965). Excluding a couple of minor restrictions and accidental gaps, the major generalizations held for obstruents are given in (17). The obstruents in Piro include /p t k ts tʃ tç s ʃ ç/, with /tç/ and /ç/ as non-strident palatal segments specified as both Coronal and Dorsal. The restrictions apply in underlying representation and/or during the derivation. When non-permissible clusters occur through morpheme concatenation and boundary vowel deletion, the first consonant of the cluster deletes, with compensatory lengthening of the preceding vowel (Matteson 1965: 33–34; Lin 1997).[16]

(17) *Piro obstruent co-occurrence restrictions*

a. *[p–p] *[t–t] *[k–k] *fricative–fricative *affricate–affricate
underlying morpheme and derivational restrictions (33)
e.g. nika 'he eats' – ka 'passive' → nikka → ni:ka 'he is eaten'
kose 'to pull' – çe 'always' – ta → kosçeta → ko:seta 'to always pull'
çitçi-tʃi → çitçtʃi → çi:tʃi 'foot'

b. *[t]–affricate
underlying morpheme and derivational restrictions (33)
e.g. hitsrukate-tʃi → hitsrukattʃi → hitsruka:tʃi 'chief'

c. affricate–[t] allowed
e.g. hajehitçtokota 'to hate secretly' (433)

d. fricative–affricate allowed
e.g. kaʃtʃa-ta 'to grab and eat' (350)
wanestsi 'approximation' (383)
stʃeha 'curl' (423)

e. *strident affricate–strident fricative
underlying morpheme restrictions

f. [tç]-fricative and [ç]–affricate allowed
e.g. haʃitçʃima 'a kind of fish' (256)
hitççetyawaka 'custom' (273)
hiçtçipaka 'to sway, reel' (393)

The interests and challenges of the data are: (i) *affricate–affricate clusters are ruled out as a class (17a), suggesting that these affricates may either be Coronal stops to be ruled out together with *[t–t] or have unordered stricture features similar to Schafer's (1995) analysis of German (§2.2.3.2); (ii) however, the edge

[16] The numbers after the examples in (17) indicate the page numbers in Matteson (1965).

effects exhibited by *[t]–affricate *vs.* affricate–[t] (17b) and (17c) and by fricative–affricate *vs.* *strident affricate–strident fricative clusters (17d) and (17e) suggest that [cont] must be present in the affricate and ordered at the right edge; (iii) affricate–fricative clusters are prohibited only when both are strident (17e) and (17f), which suggests the patterning of affricates and fricatives in term of stridency (Rubach 1994; Kim 1997), but this patterning involves edge effects, since strident fricative–strident affricate clusters are allowed (17d). As in Schafer's analysis of German consonant clusters (§5.1), one single affricate representation coupled with only one level of constraint evaluation seems impossible. If affricates are represented as ordered [stop–cont], the OCP on [cont] that rules out *fricative–fricative clusters would also rule out well-formed [tç]–fricative clusters. If the two features are unordered, the OCP on [cont] would rule out well-formed fricative–affricative and [tç]-fricative clusters. If affricates are just stops, the OCP on homorganic [stop] that rules out *[t–t] and *[t]–affricate, together with *[p–p] and *[k–k], would also incorrectly rule out well-formed affricate–[t] clusters.

Lin (2005) proposes an optimality-theoretic (Prince and Smolensky 1993) account in which (i) affricates are stops underlyingly and in the lexical phonology but have fricative release at the right edge post-lexically, and (ii) different constraints are active at these two different levels.[14] The independent evidence provided for the lexical–post-lexical division in Piro comes from cyclic boundary vowel deletion followed by cluster simplification, which shows that fricative clusters such as [s–ç] are simplified to [ç] cyclically in the lexical phonology but stop clusters such as [k–k] are allowed lexically but prohibited and simplified to [k] post-lexically (Lin 1997, 2005: 133–134).

The relevant constraints and lexical *vs.* post-lexical constraint rankings proposed by Lin (2005: 131–133) are given in (18).

(18) *Constraints and constraint rankings*

a. *Relevant constraints*
C-Faith
The corresponding consonants in the input and output are identical.
OCP-[cont]
Adjacent [cont] features are prohibited.
OCP-[stop] & OCP-Pl
Adjacent [stop] and identical major place features are prohibited.
OCP-aff
Adjacent [stop, cont] segments are prohibited.[15]
OCP-[cont] & OCP-[+strid]
Adjacent [cont] and [+strident] features are prohibited.

b. *Lexical constraint ranking*
OCP-[cont] >> C-Faith >> OCP-[stop] & OCP-Pl, OCP-aff, OCP-[cont] & OCP-[+strid]

c. *Post-lexical constraint ranking*
OCP-[stop] & OCP-Pl, OCP-aff, OCP-[cont] & OCP-[+strid] >> C-Faith >> OCP-[cont]

[14] See Lin (2005: 137–139) for an alternative analysis that makes use of Schafer's (1995) [stop]-as-head model, which, as in the German case (§5.1), also requires two levels of evaluation.

[15] This can be interpreted as an OCP effect against more than one marked segment within a local domain (Alderete 1997; Ito and Mester 2003).

(19) *Piro phonetic transitional vowel and syllabic consonant*

a.	stop–stop	[t^{ə}po]	'curve'
	stop–fricative	[k^{ə}su]	'. . .'s house'
	fricative–stop	[s̩kota]	'low abdomen'
b.	stop–affricate	[p^{ə}ʧowi]	'an edible root'
	fricative-affricate	[joçtçita]	'he changes course'
c.	affricate–stop	[ʧ^{ə}kotu]	'cebus monkey'
	affricate–fricative	[peʧiʧ^{ə}çeta]	'he always lies in wait'

This looks like an unexpected case of anti-edge effects, because the process refers to the right-edge context of the first consonant, and yet affricates pattern with stops rather than fricatives.

The process applies also to sonorants, which always show up as syllabic consonants when followed by any consonant, and a transition from a fricative to a sonorant leads to transitional vowel insertion (e.g. [smota] → [s^{ə}mota] 'blunt point'), unlike the syllabic fricatives in (19a) and (19b). Transitional vowel insertion always applies between a stop/affricate and any consonant. Lin (2005) gives a generalized sonority-based analysis in which (i) stops and affricates as a group are less sonorant than fricatives, which in turn are less sonorant than sonorants, (ii) a transitional vowel is inserted when the first consonant is a stop/affricate or it is less sonorant than the second consonant, and (iii) the first consonant becomes syllabic if it is a sonorant or it is more sonorous than the second consonant. It is concluded that if stricture properties determine relative sonority of segments, as is commonly assumed, the fact that affricates and stops belong to the same sonority class suggests that the stop part of the affricates is primary and the fricative part is secondary even at the phonetic level.

Alternatively, it is also likely that affricates, like stops but unlike fricatives and sonorants, have release positions (Steriade 1993; (12) in §3), so stops and affricates behave in the same way when the releases are not masked by the next consonant and thus perceptually interpreted as transitional vowels. Whatever the analysis and interpretation of the data, this phonetic patterning of affricates with stops in Piro may indicate the fundamental nature of affricates as being stops.

6 Concluding remarks

The Stop Approach has put forward cogent arguments and appears to have acquired a strong following in recent years. Some studies from child language acquisition, such as Menn (1973) on English and Cook (2006) on Chipewyan, seem to provide additional support to the affinity between stops and affricates by showing that stops and affricates are acquired earlier than fricatives and that fricatives are often replaced by stops or affricates in children's speech (cf. Jakobson's 1941 claim that stops and fricatives are acquired before affricates). Moreover, some studies exploring the interaction between phonetics and phonology indicate that the fricative component of the affricate is phonetic in nature. For example, as mentioned in §3.2, Clements (1999) and Kim (1997, 2001) provide a phonetically based explanation for high vocoid conditioned phonological assibilation (e.g. /t/ → [ʧ] / __ [i]): the affricate is not created by assimilating [cont] from the vowel

17 Distinctive Features

Jeff Mielke

1 Introduction

Distinctive feature theory is an effort to identify the phonetic dimensions that are important for lexical contrasts and phonological patterns in human languages. The set of features and its explanatory role have both expanded over the years, with features being used to define not only the contrasts but the groupings of sounds involved in rules and phonotactic restrictions, as well as the changes involved in rules. Distinctive features have been used to account for a wide range of phonological phenomena, and this chapter overviews the incremental steps by which the feature model has changed, along with some of the evidence for these steps. An important point is that many of the steps involve non-obvious connections, something that is harder to see in hindsight. Recognizing that these steps are not obvious is important in order to see the insights that have been made in the history of distinctive feature theory, and to see that these claims are associated with differing degrees of evidence, despite often being assumed to be correct.

The structure of the chapter is as follows. §2 describes the series of non-obvious claims that led to modern distinctive feature theory, and §3 briefly describes the particular features that have been proposed. §4 reconsiders some of the earlier claims, and §5 looks at some ongoing investigations.

2 Building a model of phonological behavior

Distinctive feature theory began largely as a model for reducing the number of phonological contrasts in a language, and the feature system that was developed for this purpose was gradually embellished in order to provide an account for more and more facts about sound patterns and typology. Trubetzkoy (1939: 66–89) developed the study of oppositions between speech sounds, which grew into the study of distinctive features. After Trubetzkoy's death in 1938, the early years of distinctive feature theory were associated most with Roman Jakobson.

2.1 Reducing contrasts with (mostly) binary features

Jakobson (1942: 235) hypothesized that differences that on their own are not meaningful, such as the differences between phonemes, are very demanding on perception and memory, and he concluded that the number of "primordial and unmotivated values" should be minimized. For example, the eight-vowel inventory of Turkish (/i ɨ y u e ɑ ø o/) involves 28 binary relations, as illustrated in Figure 17.1. Jakobson observed that many of these relations are essentially the same, e.g. the difference between /i/ and /y/ is basically parallel to the difference between /e/ and /ø/. Both relate an unrounded vowel and a rounded vowel which are otherwise largely the same. In the figure, solid lines with black circles at the end are used to represent independent contrasts. The contrasts between the eight vowels can be reduced to three orthogonal dimensions (height, backness, and rounding), and any of the 28 binary relations can be represented as differences along one or more of these three dimensions (see CHAPTER 2: CONTRAST). In this way, 28 binary oppositions can be captured by three binary distinctive features. The vowel inventory of Turkish is a particularly clean example, because it happens to be "cubic," fully exploiting three features, but any inventory containing pairs of segments that differ from each other along similar phonetic dimensions is reducible in this way to some extent.

The assumption of binarity for these features since the early days of feature theory can be attributed in part to the influence of Information Theory (Shannon and Weaver 1949). Jakobson *et al.* (1952) argued that Information Theory provides a sequence of binary selections as the most reasonable way to analyze communication, and that language is not merely amenable to such an analysis, but is inherently structured in this way. Some pairs of words, such as *bill/pill* and *bill/dill*, involve a difference of one feature, while pairs like *bill/fell* can be treated as a difference of more than one binary feature (since the difference between /b/ and /f/ can be reduced to two independently motivated dimensions, and the difference between /ɪ/ and /ɛ/ can be reduced to one). Jakobson *et al.* (1952) noted that each of the phonetic dimensions related to a distinctive feature is continuous, but features consistently pick out two polar points. Since the dichotomous scale was believed to be the optimal code, they saw no reason why language would be organized according to a more complicated system. Some features, such as [compact/diffuse] (see next section) were considered to be equipollent, having three values (compact, diffuse, or neither), based on the fact that there seem to be three degrees of vowel height along the same dimension (e.g. /i/–/e/–/æ/).

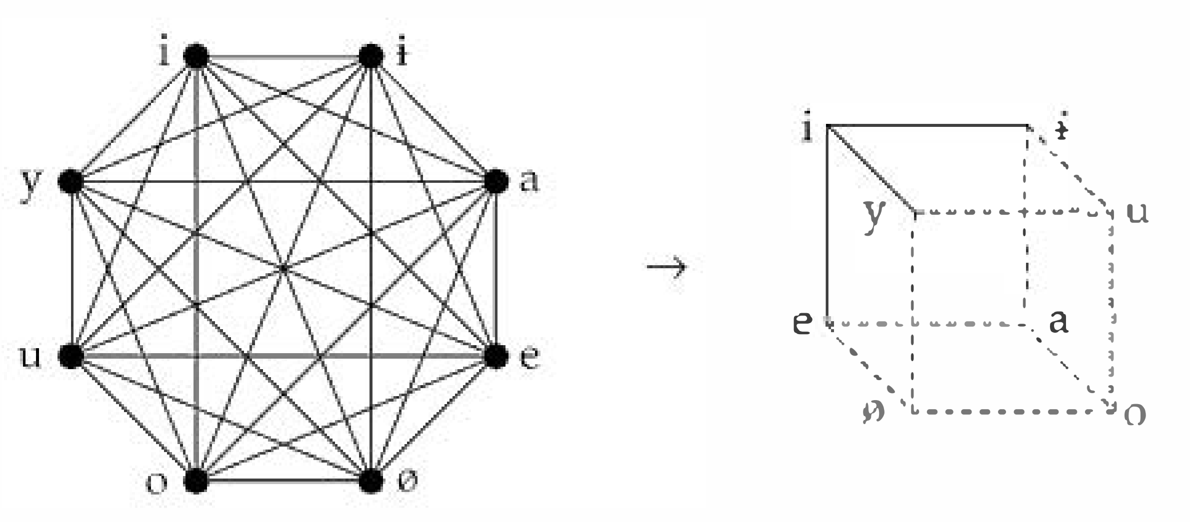


Figure 17.1 Reducing 28 binary relations to 3

The advances up to this point are summarized in (1).

(1) a. Claim: Segmental contrasts can be reduced to a smaller number of featural contrasts.
 b. Claim: Features are (mostly) binary.

Miller and Nicely (1955) provided some phonetic evidence in support of the idea that contrast between segments is divided into discrete channels. They showed that the distinctions between different consonants occupy different parts of the speech signal, and can be interfered with through different techniques for signal degradation. Miller and Nicely adopted voicing, nasality, affrication, duration (which could also be called "stridency") and place as features to distinguish /p t k f θ s ʃ b d g v ð z ʒ m n/. The oppositions associated with these features are affected differently by random masking noise, high-pass filtering, and low-pass filtering. The idea is that the information content of a message is distributed among different acoustic channels, and can consequently be selectively and independently degraded. Random masking noise degrades affrication, duration, and place distinctions more than it degrades voicing and nasality. Duration, on the other hand, is more resistant to high-pass filtering. These results are relevant to the design of telephone communication systems, but also supportive of the idea that different phonological contrasts are manifested in different parts of the speech signal. This provides evidence consistent with consolidating phonetic differences as in the Turkish example above in Figure 17.1. Although the dimensions investigated by Miller and Nicely are not the same dimension that distinguish the Turkish vowels, oppositions that are treated as parallel in phonological analyses were shown to be degradable in transmission, independent of other oppositions.

2.2 *Abstractness, universality, and innateness*

The reduction of segmental contrasts to a smaller number of distinctive features can be a purely language-particular endeavor. Jakobson (1942: 241) asserted that "[t]he description of a system of values and the classification of its elements can be made only from that system's own perspective, that is, from the perspective of the tasks that the system fulfills." He observed, though, that languages often make use of the same phonetic dimensions for contrasts (Jakobson 1942: 239). If two languages use the same phonetic dimension, then a distinctive feature in one language is fundamentally similar to the "same" feature used in the other. Jakobson contrasted this with the difference between similar-sounding phonemes in different languages. Since phonemes are treated as bundles of distinctive feature values, the phonological content of a phoneme depends on the segmental oppositions it is involved in. It was the *features*, then, that could be treated as basic building blocks, comparable across languages.

Jakobson *et al.* (1952: 40) reported that they detected only 12 distinctive features (Table 17.1) in the languages of the world "which underlie their entire lexical and morphological stock," but left open the possibility of adding more. These features were defined primarily in acoustic terms.

The universality of this kind of feature set can be interpreted as a fact about the phonetic dimensions that are available to the human vocal tract and the human auditory system rather than a claim about the features themselves. That is, if humans

Table 17.1 *Preliminaries* (Jakobson *et al.* 1952): Acoustically defined features

Fundamental source features:		
Vocalic (periodic voice source with non-abrupt onset)	*vs.*	Non-vocalic
Consonantal (acoustic zeros across the spectrum)	*vs.*	Non-consonantal
Envelope features:[a]		
Interrupted (abrupt onset)	*vs.*	Continuant
Checked (abrupt decay)	*vs.*	Unchecked
Strident (irregular waveform)	*vs.*	Mellow
Voiced	*vs.*	Voiceless
Resonance features:		
Compact (predominance of one formant region)	*vs.*	Diffuse
Grave (low end of spectrum dominates)	*vs.*	Acute (high end dominates)
Flat (formants shifted down)	*vs.*	Plain
Sharp (F2 and other formants shifted up)	*vs.*	Plain
Tense (longer and more energetic)	*vs.*	Lax
Nasal	*vs.*	Oral

[a] The term *envelope* in the feature categories refers to the "temporal envelope of sound intensity" (Jakobson *et al.* 1952: 21), i.e. how abruptly the sound starts or ends, and how smooth the intensity remains in between.

are physically limited to 12 phonetic dimensions for distinguishing sounds, then language is similarly limited in the range of distinctive features it can involve. The limiting factors are physiology and acoustics, not features. Jakobson *et al.* (1952: 31) argued that the feature [flat] (defined acoustically as downwardly shifted formants) could apply to both pharyngealization and labialization, since the two articulatory gestures have similar acoustic effects, since they appear never to be used contrastively in the same language, and since Bantu languages and Uzbek substitute labialized consonants for Arabic pharyngealized consonants in loanwords. This can be treated as an observation about acoustic and perceptual similarity or as the effects of a universal abstract feature. In later approaches, the features themselves were taken to be the basic limiting factor. While this practice of drawing explanations from (potentially innate) features themselves has been attributed to the influence of Chomsky's approach to syntax (e.g. Chomsky 1957, 1965), Halle (1983) reports that it was present all along:

> Considerations [that languages apparently do not make use of acoustic or articulatory correlates of features alone] were in our minds thirty years ago when Jakobson, Fant and I were working on *Preliminaries to Speech Analysis*, and it was these considerations that led us to draw a sharp distinction between distinctive features, which were abstract phonological entities, and their concrete articulatory and acoustic implementations. Thus, in *Preliminaries* we spoke not of "articulatory features" or of "acoustic features," but of "articulatory" and/or "acoustic correlates" of particular distinctive features. The model we had in mind was, therefore, of the type . . . where the abstract distinctive features constitute the link between specific articulatory and acoustic properties of speech sounds.

Quantal Theory (Stevens 1972, 1989) attributes the phonological oppositions used by languages to the non-linear relationships between articulatory and acoustic

2.3 *Features for alternations*

While Jakobson *et al.* (1952) focused on identifying the distinctive features used for contrasting sounds, these features have since been put to other uses in describing phonological phenomena. For example, it was discovered that the features used to define lexical contrasts could also be used to formulate phonological rules. Halle (1959: 65) describes the assimilation pattern in Russian in (3) in terms of features as follows: "Before acute compact (palatal) consonants, strident acute noncompact (dental) consonants become compact (palatal)" (see CHAPTER 71: PALATALIZATION and CHAPTER 121: SLAVIC PALATALIZATION).

(3) *Russian sibilant assimilation* (Halle 1959: 65)

/bez-ˈʧestʲij-o/	bʲiʃˈʧʲesʲtʲijə	'dishonor'
/bez-ˈʒalostʲn-o/	bʲiʒˈʒaəsnə	'pitiless'
/s-ˈʃum-om/	ˈʃʃuməm	'with noise'

This rule can be formulated in the *SPE* rule format (using *Preliminaries* features) as in (4).

(4) $$\begin{bmatrix} \text{consonantal} \\ \text{strident} \\ \text{acute} \\ \text{non-compact} \end{bmatrix} \rightarrow [\text{compact}] \;/\; \_\_ \begin{bmatrix} \text{consonantal} \\ \text{acute} \\ \text{compact} \end{bmatrix}$$

Here the feature [compact] is being used not simply to distinguish sounds from each other, but for two additional purposes. First, along with other features, it is being used to define the classes of sounds involved in a sound pattern, namely the class of sounds that become palatal in the context of the rule, and the class of sounds that trigger this change. Second, the feature [compact] is being used to define the change described by the rule. As a member of the set of distinctive features, [compact] can define contrasts, classes, and changes. These multiple functions mean that multiple types of evidence can be used to motivate a feature. Since there are often multiple ways to define the same class or the same segmental contrast, spreading of a single property is taken as particularly strong evidence for the existence of a feature.

The set of possible changes is identified with the set of possible contrasts, and the set of classes that are predicted to be involved in rules (or later, constraints) are defined as those classes of segments that share one or more feature values, to the exclusion of all other sounds in the inventory. Using the same features that define contrasts to describe classes of sounds involved in rules or to describe changes involved in rules is not a necessary step, and it interacts with the earlier claim that there is a particular feature set available to all languages.

(5) a. Claim: The distinctive features that define segmental contrasts also define changes in alternations.
    b. Claim: The same features also define classes of sounds that may be involved in alternations.

Halle (1972: 62) summarizes the goals of feature theory as follows:

> I take it that a study of the speech sounds of a given language must account for, among others, the following three sets of facts: it must yield insights into the articulatory aspects of the sounds; it must concern itself with the acoustic and psychoacoustic character of the sounds, and, finally, it must allow us to make sense of various regularities that can be observed in the behavior of different speech sounds and sets of speech sounds, regularities that have traditionally been referred to as phonological or morphophonological. The task of the student of speech sounds then is to discover a theory that will do justice to these different aspects of speech.

A further development of feature theory that was motivated in part by the account of alterations involve the number of values possessed by features. For example, Sagey (1986) argued that place features are privative (have only one value). This claim helps account for the observation that the negative values of these features do not seem to be involved in sound patterns. Dependency Phonology and Element Theory (Harris 1994; Harris and Lindsey 1995) take this further. See also CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION.

2.4 *Characterizing functional subgroupings of features*

A further development of the feature model was the hierarchical organization of features, in a way that is consistent both with the interdependence of their phonetic correlates and with the way they are grouped together in phonological patterns. In motivating constituency among distinctive features, Clements (1985: 229) observed that at least four articulatory parameters show considerable independence from each other: (a) laryngeal configuration, (b) degree of nasal cavity stricture, (c) degree and type of oral cavity stricture, (d) pairing of an active and a passive articulator:

> For example, one can maintain a certain oral tract configuration constant, say the one appropriate for producing the vowel [a], while varying the type of laryngeal configuration, or the position of the velum. Or one can hold the laryngeal configuration constant while varying the internal geometry of the oral tract.

However, within each category, it is diffiicult or impossible to vary one gesture while maintaining another. Clements observes that with the exception of laryngeal, which seems to be completely independent, there is limited mutual dependence between these parameters. Clements emphasizes that the justification for a model of feature organization must come from observed sound patterns rather than physiology, even if there is an apparent physiological basis to the organization. This move to features organized hierarchically was accompanied by a move to autosegmental rules, in which assimilation is treated as an association between a set of features forming a constituent in the hierarchy and the representation of a segment undergoing assimilation. Since there are assimilatory patterns that spread only laryngeal features, these features are grouped together as a constituent in the hierarchy, and such patterns are formalized as spreading of the laryngeal node, which dominates these features. Spreading of place features independently of other features is handled similarly. After the addition of feature organization and [vocalic], changing the definitions from acoustic to articulatory without changing much in the way of segments they apply to. In both cases, phonetic definitions were chosen to match the traditional understandings of consonants and vowels. Clements and Hume (1995) later advocated rebranding [consonantal] as its opposite [vocoid]. Chomsky and Halle (1968: 302) also added the feature [sonorant], and later entertained the idea of dropping [vocalic] in favor of [syllabic] (1968: 353–355). The difference between these two is that [+vocalic] includes both syllabic and non-syllabic liquids and excludes nasals, while [+syllabic] includes syllabic liquids and nasals, but excludes their non-syllabic counterparts. The motivation for making this change is accounting for truncation phenomena occurring before foreign words in French, whereby vowels truncate before vowels, while consonants truncate before consonants (including liquids and glides), so there is no elision in *le yogi* and no liaison in *les yogis*, as in (8). When the following word is part of the native vocabulary, glides pattern with vowels.

(8) *Elision in French native and foreign words* (Chomsky and Halle 1968: 353)

a. Native vocabulary: Glides pattern with vowels
le garçon
l'oiseau
l'enfant

b. Foreign vocabulary: Glides pattern with consonants
le yogi
**l'yogi*

The feature [syllabic] is needed to capture the distinction between the classes involved in truncation before foreign words (Chomsky and Halle 1968: 355):

> This example is, thus, of the greatest importance for our feature framework. If, as it now appears to us, [(9)] is indeed the correct formulation of the phonetic facts just discussed, and if, moreover, this example is shown to be more than an isolated instance, the feature framework will have to be revised along the lines [of replacing [vocalic] with [syllabic]].

(9) $$\begin{bmatrix} -\alpha\text{syll} \\ \alpha\text{cons} \end{bmatrix} \rightarrow \emptyset \ / \ \_\_\# \begin{bmatrix} -\alpha\text{syll} \\ +\text{foreign} \end{bmatrix}$$

History has shown that this example is indeed more than an isolated instance, and that the distinction between vowels and consonants is referred to by rules. [syllabic] is not typically necessary to contrast the segments in inventories, because [vocalic] does basically the same work, but the issue is capturing classes of sounds that pattern together. More recently, the feature [syllabic] has been mostly abandoned as a segmental feature, because its work (distinguishing segments that are or are not syllable nuclei) is done by prosodic structures (see e.g. Anderson 1981). But there are two important points here. First, the distinction made by the feature [syllabic] turned out to be supported by cross-linguistic evidence, showing that the initial proposal was on the right track. Second, though, the feature [vocalic] is automatically jettisoned as a consequence. The problem is that subsequent work in phonology has shown a need for this distinction too, but a one allomorph is treated as basic and the other derived, there is no way to characterize a derived class in terms of traditional distinctive features, or to describe it in terms of shared phonetic properties. The most straightforward reason for the unnaturalness of this class is the fact that the dental nasal patterns with the retroflex stops and not with the dental stops.

(12) *Kolami plural* /-(u)l/ *allomorphy* (Emeneau 1961: 46–50)

a. [-l] after /ʈ ɖ n̪ i i: e e: a a:/

singular	*plural*	
d̪uʈ	d̪uʈl	'hips'
eɖ	eɖl	'bullock'
to:ren̪	to:ren̪l	'younger brothers'
sir	sid̪l	'female buffalo'
kaje	kajel	'fish'
bi:-am	bi:l	'rice'
kala	kalal	'dreams'

b. [-ul] after /p t̪ k d̪ g s v z m ŋ j/

singular	*plural*	
ro:p	ro:pul	'plant'
ket̪	ket̪ul	'winnowing fans'
ma:k	ma:kul	'tree'
mood̪	mood̪ul	'particular man . . .'
d̪eg	d̪egul	'heaps, masses'
kis	kisul	'fires'
a:v	a:vul	'fathoms'
ga:z	ga:zul	'bangle'
d̪em	d̪emul	'draws on a pipe'
nenjeŋ	nenjeŋul	'meat'
poj	pojul	'hearth'

An alternative to a universal feature set as an explanatory force in phonology is that inventories and sound patterns are limited by phonetic and historical factors (factors often invoked to explain "crazy" phenomena), and the appearance of a single feature set responsible for disparate phonological phenomena may be due to the role of phonetics in all of these phenomena. This is consistent with Blevins's (2004) approach to sound patterns in general and Mielke's (2008) approach to feature effects. A move away from a small feature set is also seen in some approaches to Optimality Theory, such as Flemming (1995), Kirchner (1997), and Boersma (1998).

4.1 Segmental contrasts

The idea that features play a limiting role in determining possible contrasts in inventories has for the most part not been accompanied by a null hypothesis about what segmental contrasts would be like in the absence of features. Mielke (2008) argues that the record of sound patterns does not support the literal interpretation of a restricted universal set of features. The main criticism is that the similarity between sound patterns in unrelated languages can be accounted for in terms of shared historical and phonetic factors.

An example of this type of argument is final devoicing, which is documented in a wide range of unrelated languages (see CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION). Blevins (2006: 136–140) describes the sources of these final devoicing patterns in detail. To summarize briefly, the phonetic sources of phrase-final devoicing include the use of laryngeal spreading and closing gestures to mark phrase boundaries in many languages, phrase-final lengthening, and difficulty perceiving phrase-final consonant release. Further, since children are typically exposed to a high number of single-word utterances, there is a tendency for them to generalize phrase-final phenomena to the word-final context. Since final voicing does not have the same phonetic sources, it is not widely observed. Mielke (forthcoming) shows that uses of the feature [voice] to characterize sound patterns are dominated by a small number of phonetically motivated patterns that are mostly assimilatory. As long as there are phonetic and historical explanations for the recurrence of certain sound patterns, feature theory can be interpreted as a model of these historical and phonetic effects (and this is how it is interpreted by many linguists), but the nature of sound patterns does not support features as a primary source of explanation or the restrictive nature of a small universal feature set.

If features are not universal as a part of Universal Grammar, then there is no need to explain how they could have emerged in the human genome over a short period of time, or why signed and spoken languages appear to have completely different features (or how signed language features could have developed in the human genome at all). See CHAPTER 9: HANDSHAPE IN SIGN LANGUAGE PHONOLOGY, CHAPTER 10: THE OTHER HAND IN SIGN LANGUAGE PHONOLOGY and CHAPTER 24: THE PHONOLOGY OF MOVEMENT IN SIGN LANGUAGE on features in sign language phonology. There is, however, a need to show how features are learned, and this is a topic of ongoing research (Lin 2005; Lin and Mielke 2008; Mielke 2010).

Also, if features are not innate, then there are consequences for other arguably universal primitives, such as constraints in Optimality Theory (Prince and Smolensky 1993). Many of these constraints refer to features, so if the features are not innate, the constraints referring to them cannot be innate either. Some approaches to Optimality Theory, such as Hayes (1999), involve universal constraints that are induced from phonetics. Features referred to by these constraints would need to be either innate or similarly induced from available data. See CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS and CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS.

Crucial evidence for features as the basis of segmental contrasts would be contrasts that appear to be ruled out by features but not by phonetics, such as (perhaps) the unattested apical retroflex *vs.* sub-apical retroflex contrast. However, it remains to be shown whether the cross-linguistic frequency distribution of both types of retroflex is such that they would be expected to co-occur in the same language anyway.

On the other hand, proposed features predict a wide range of unattested contrasts, and overdetermine the representation options for many attested ones. As discussed above, it has been possible to maintain the claim that there are only three vowel heights by making use of auxiliary "height" features like [ATR]. This is possible in part because it is difficult to falsify without articulatory imaging. While there is X-ray evidence for English to support the use of [ATR] to distinguish, e.g.

The last definition requires reference to a feature theory:

(15) *Natural class: Feature theory-dependent definition*

A group of sounds in an inventory that share one or more distinctive features within a particular feature theory, to the exclusion of all other sounds in the inventory.

The reason for breaking the definition down in this way is that natural classes and phonologically active classes are implicitly treated as the same thing in many approaches to phonology. It is worthwhile to seriously ask whether these definitions all refer to the same thing. The identification of natural classes with phonologically active classes is an explicit working assumption in some approaches, e.g. Halle (2002), who argues that the position that phonologically active classes are featurally natural has led to improvements in analyses of phonological phenomena. Mielke (2008) collected a database of phonologically active classes in the world's languages based on grammars of 549 languages and conducted feature analyses on the 6,077 resulting phonologically active classes involved in alternations, using *Preliminaries* (Jakobson *et al.* 1952), *SPE* (Chomsky and Halle 1968), and Unified Feature Theory (Clements and Hume 1995) features. Chomsky and Halle (1968) does the best, but still leaves a residue of about 29 percent of the unnatural classes, which is not easily handled using feature proposals subsequent to *SPE* (e.g. the Kolami example above and hundreds of others).

The way in which natural classes are traditionally defined (conjunction of distinctive feature values, yielding the intersection of the classes defined by each feature value individually) is also non-obvious. This technique has allowed many phonologically active classes that are not defined by a single feature to be represented, providing support for the idea that features used for contrast are also accounting for phonologically active classes, but it also massively overpredicts the range of phonologically active classes. It is not obvious that phonologically active classes defined by more than one feature are less atomic than classes defined by one feature. For instance, the class defined by [+voice, −sonorant] (voiced obstruents) is more frequently active than the class defined by [+voice] alone, and the class of glides (which requires reference to syllabicity and something like [−consonantal] or [+vocoid]) is more frequently active than the class defined by [−consonantal]/[+vocoid] alone. The fortunate fact that these active classes can be defined in terms of features needed to define contrasts has made it unnecessary to propose features for them, but their reasons for being involved in sound patterns may have more to do with the phonetic properties of the sounds involved in a way that is not directly related to the features used to describe them.

If there is a mismatch between phonologically active classes and the features used for lexical contrast, an alternative is a separate model predicting classes that are likely to be involved in sound patterns, perhaps in terms of common sound changes affecting more than one segment (Blevins 2004; Mielke 2008, forthcoming). The phonetic parameters defining these classes would not necessarily correspond directly to the parameters needed to contrast segments from one another. The class of voiced obstruents can be defined atomically in terms of non-spontaneous voicing – consistent with feature analyses by Rice and Avery (1989) and Rice (1992) that use the feature [sonorant voice], distinct from [voice] – which is relevant for many of the sound patterns this class is active in;

moreover, the class of glides is frequently involved in sound changes resulting in vocalization/devocalization alternations.

Flemming (2005) argues for another type of phonologically active class that is a consequence of constraint interaction. In this view, classes of sounds that pattern together can do so as a result of separate constraints, each targeting a different featurally natural class. The resulting (apparent) phonologically active class is not necessarily featurally natural, since it is determined by constraint interaction, not directly by features. See also Yip (2004, 2005) for a constraint-based approach to the cross-linguistic variability of laterals.

Mielke (2005) argues that examining a large sample of phonologically active classes shows that they range quite continuously from phonetically natural to phonetically unnatural, and that there does not appear to be evidence for a distinct cut-off between natural and unnatural classes. In cases where phonetic correlates are less clear, phonological patterning is also less clear, indicating that it is the phonetic properties in which features are grounded that are important, and features themselves do not make more specific predictions about sound patterns. For example, the feature [continuant] appears to be categorical for segments such as stops and fricatives, which strongly display its phonetic correlates, but laterals – whose specification has been controversial and whose phonetic cues are ambiguous – pattern with continuants about as often as with non-continuants. See CHAPTER 31: LATERAL CONSONANTS for more on this issue.

4.3 *Defining alternations*

If the features needed for defining the change in alternations do not match the features needed for other purposes, they could be accounted for by a model of likely alternations. For example, it is known that not all features appear to spread. This observation was addressed within feature geometry by placing non-spreading features such as [consonantal] and [sonorant] in the root node. If assimilation is the result of phonologized co-articulation (Baudouin de Courtenay 1972 [1895]; Ohala 1993; Blevins 2004), then an account of co-articulation and its phonologization could possibly account more directly for sound patterns interpreted as feature spreading. Mielke (forthcoming) surveys the behavior of classes defined by various features, and concludes that the phonological behavior of particular features can be attributed to the phonologization of phonetic effects. Features that are frequently spread in assimilatory patterns are features whose phonetic correlates are believed to be easily co-articulated. Features that rarely or never spread seem to be those that are not easily involved in co-articulation without involving the correlates of other features. Feature values that are frequently involved in dissimilation ([−voice] and [−nasal]) are the opposite values of the feature values that assimilate most frequently, consistent with Ohala's (1981) claim that dissimilation is the result of mistakenly undone assimilation (see CHAPTER 60: DISSIMILATION). As such, dissimilation is dependent on the phonologized co-articulation of an opposite value.

5 New types of experimental evidence

Regardless of whether features are needed to account for the typology of contrasts and sound patterns, behavioral studies such as Studdert-Kennedy *et al.*

(1972) indicated that features are involved in representations used in language processing. More recently, advances in brain imaging technology have opened new lines of research into the mental organization of phonology, building on earlier behavioral research (see CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY). Two big questions here are whether brain activity provides evidence for abstract features in adults, and the separate issue of whether this organization is a consequence of language exposure or a consequence of innate aspects of the language faculty.

In a magneto-encephalography (MEG) study, Phillips *et al.* (2000) report evidence for the feature [voice] in the left-hemisphere auditory cortex of adult English-speaking participants. This is similar to Studdert-Kennedy *et al.*'s (1972) dichotic listening experiment. Since acoustic similarity is similarly controlled for, an abstract feature can be motivated over acoustic similarity. Acoustically distinct but featurally identical stimuli were treated as identical in the auditory cortex. Other MEG studies report evidence of abstract vowel features (Obleser *et al.* 2004) and featural underspecification in the mental lexicon (Eulitz and Lahiri 2004). Dehaene-Lambertz and Pena (2001) report electrophysiological evidence that newborns distinguish [pa] and [ta] in a way that they do not distinguish repetition of the same syllable produced by different speakers. Studies with infants have the potential to address more directly the questions about whether these abstract representations are rooted in innate features or in exposure to language data.

These are a few examples of types of evidence that were not available in the early years of distinctive feature theory. New techniques for studying phonetics and phonology – such as brain and vocal tract imaging, new behavioral methodologies and computer modeling, and electronic databases of inventories, sound patterns, and sound changes – stand to improve our understanding of sound patterns and our understanding of how features are involved in accounting for them.

REFERENCES

Anderson, Stephen R. 1981. Why phonology isn't "natural". *Linguistic Inquiry* 12. 493–539.

Aslin, Richard N. & David B. Pisoni. 1980. Some developmental processes in speech perception. In Grace H. Yeni-Komshian, James F. Kavanagh & Charles A. Ferguson (eds.) *Child phonology*, vol. 2: *Perception*, 67–96. New York: Academic Press.

Avery, Peter & William J. Idsardi. 2001. Laryngeal dimensions, completion and enhancement. In T. A. Hall (ed.) *Distinctive feature theory*, 41–70. Berlin & New York: Mouton de Gruyter.

Bach, Emmon & Robert T. Harms. 1972. How do languages get crazy rules? In Robert P. Stockwell & Ronald K. S. Macaulay (eds.) *Linguistic change and generative theory*, 1–21. Bloomington: Indiana University Press.

Baudouin de Courtenay, Jan. 1895. An attempt at a theory of phonetic alternations. In Edward Stankiewicz (ed. and trans.) 1972. *A Baudouin de Courtenay anthology*, 144–212. Bloomington: Indiana University Press.

Best, Catherine T., Gerald W. McRoberts & Nomathemba M. Sithole. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14. 345–360.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette. 2005. Some problems with feature economy: Labial fricatives, lone fricatives, and lost labials. In *Seoul Linguistics Forum 2005*, 211–230.

Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32. 117–166.

Boer, Bart de. 2001. *The origins of vowel systems*. Oxford: Oxford University Press.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Bradshaw, Mary M. 1999. A crosslinguistic study of consonant–tone interaction. Ph.D. dissertation, Ohio State University.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. 1992. Phonological primes: Features or gestures? *Phonetica* 49. 181–193.

Clements, G. N. 2003. Feature economy in sound systems. *Phonology* 20. 287–333.

Clements, G. N. 2009. The role of features in speech sound inventories. In Eric Raimy & Charles Cairns (eds.) *Contemporary views on architecture and representations in phonology*, 19–68. Cambridge, MA: MIT Press.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 245–306. Cambridge, MA & Oxford: Blackwell.

Dehaene-Lambertz, Ghislaine & Marcela Pena. 2001. Electrophysiological evidence for automatic phonetic processing in neonates. *NeuroReport* 12. 3155–3158.

Eimas, Peter D., Einar R. Siqueland, Peter Jusczyk & James Vigorito. 1971. Speech perception in infants. *Science* 171. 303–306.

Emeneau, M. B. 1961. *Kolami: A Dravidian language*. Annamalainagar: Annamalai University.

Eulitz, Carsten & Aditi Lahiri. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience* 16. 577–83.

Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles. Published 2002, London & New York: Routledge.

Flemming, Edward. 2005. Deriving natural classes in phonology. *Lingua* 115. 287–309.

Frisch, Stefan A. & Richard Wright. 2002. The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 30. 139–162.

Fromkin, Victoria A. (ed.) 1973. *Speech errors as linguistic evidence*. The Hague: Mouton.

Fromkin, Victoria A. 1988. Grammatical aspects of speech errors. In Frederick J. Newmeyer (ed.) *Linguistics: The Cambridge survey*, vol. 2: *Linguistic theory: Extensions and implications*, 117–138. Cambridge: Cambridge University Press.

Gick, Bryan. 2002. The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association* 32. 113–121.

Gick, Bryan, Douglas Pulleyblank, Fiona Campbell & Ngessimo Mutaka. 2006. Low vowels and transparency in Kinande vowel harmony. *Phonology* 23. 1–20.

Goldstein, Louis. 1977. Categorical features in speech perception and production. Prepared for the Working Group on Slips of the Tongue, 12th International Congress of Linguists, Vienna.

Goldstein, Louis, Marianne Pouplier, Larissa Chen, Elliot L. Saltzman & Dani Byrd. 2007. Dynamic action units slip in speech production errors. *Cognition* 103. 386–412.

Ladefoged, Peter & Daniel Everett. 1996. The status of phonetic rarities. *Language* 72. 794–800.
Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford & Malden, MA: Blackwell.
Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48. 839–862.
Lin, Ying. 2005. Learning features and segments from waveforms: A statistical model of early phonological acquisition. Ph.D. dissertation, University of California, Los Angeles.
Lin, Ying & Jeff Mielke. 2008. Discovering place and manner features: What can be learned from acoustic and articulatory data? *Penn Working Papers in Linguistics* 14. 241–254.
Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John J. Ohala & Jeri J. Jaeger (eds.) *Experimental phonology*, 13–44. Orlando: Academic Press.
Mackie, Scott. 2007. The limits of feature economy. M.A. mémoire, University of Ottawa.
Mackie, Scott & Jeff Mielke. Forthcoming. Feature economy in natural, random, and synthetic inventories. In G. N. Clements & Rachid Ridouane (eds.) *Where do phonological contrasts come from? Cognitive, physical and developmental bases of phonological features*. Amsterdam & Philadelphia: John Benjamins.
Maddieson, Ian & Kristin Precoda. 1990. Updating UPSID. *UCLA Working Papers in Phonetics* 74. 104–111.
Martinet, André. 1955. *Économie des changements phonétiques*. Berne: Francke.
Martinet, André. 1968. Phonetics and linguistic evolution. In Bertil Malmberg (ed.) *Manual of phonetics*, 464–487. Completely revised and extended edition. Amsterdam: North-Holland.
McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.
McCarthy, John J. 1991. Semitic gutturals and distinctive feature theory. In Bernard Comrie & Mushira Eid (eds.) *Perspectives on Arabic linguistics III*, 63–91. Amsterdam & Philadelphia: John Benjamins.
McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.) *Phonological structure and phonetic form: Papers in laboratory phonology III*, 191–233. Cambridge: Cambridge University Press.
Mielke, Jeff. 2005. Ambivalence and ambiguity in laterals and nasals. *Phonology* 22. 169–203.
Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.
Mielke, Jeff. 2010. Getting the features you want. Unpublished paper presented at the Conference on Computational Modeling of Phonological Acquisition, University of Alberta.
Mielke, Jeff. Forthcoming. Phonologization and the typology of feature behavior. In Alan C. L. Yu (ed.) *Origin of sound patterns: Approaches to phonologization*. Oxford: Oxford University Press.
Miller, George A. & Patricia Nicely. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27. 338–352.
Mowrey, Richard A. & Ian R. A. MacKay. 1990. Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America* 88. 1299–1312.
Obleser, Jonas, Aditi Lahiri & Carsten Eulitz. 2004. Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience* 16. 31–39.
Odden, David. 2005. *Introducing phonology*. Cambridge: Cambridge University Press.
Ohala, John J. 1981. The listener as a source of sound change. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 17. 178–203.
Ohala, John J. 1993. Coarticulation and phonology. *Language and Speech* 36. 155–170.
Padgett, Jaye. 2002. Feature classes in phonology. *Language* 78. 81–110.

18 The Representation of Clicks

Amanda Miller

1 Introduction

Click consonants are a type of complex segment. Complex segments are defined as single segments that have two oral constrictions that are nearly simultaneous (Sagey 1990). Click consonants have an anterior constriction, which is either labial or coronal, and a posterior constriction, which is uvular in those Khoesan languages in which posterior place of clicks has been investigated with articulatory methods. The posterior place of clicks in Zulu starts out as velar in the closure, and releases at a uvular place of articulation. Clicks are unique in that they are produced with an ingressive lingual (also known as velaric) airstream mechanism, which is produced by trapping air between a lingual or linguo-labial cavity formed between the two oral constrictions. The tongue moves, in different ways for different clicks, to expand this oral cavity and thus to rarefy, or decompress, the air within it. When the anterior constriction is released, air rushes in to make the characteristic popping sound. The lingual ingressive airstream differentiates clicks from pulmonic stop consonants, which are produced on an outward flow of air from the lungs, and from other complex segments, such as labial-velars and labial-coronals, which are produced using a pulmonic or glottalic airstream.

Clicks have played an important role in phonological theory because of the phonological complexity that they exhibit, and the large number of click contrasts found in Khoesan language inventories. Clicks exhibit at least three major areas of complexity that are not found in most other consonants: (1) the double place of articulation features; (2) the overlap of the two constrictions for the length of the segments; and (3) the non-pulmonic airstream mechanism. In early representations of clicks, the suction used to form the airstream was recognized. However, later proposals capture the two places of articulation features in clicks, and assume that the non-pulmonic airstream is a redundant feature that is predictable from having two simultaneous oral places of articulation. Recent representations incorporate features capturing the unique airstreams involved in clicks.

In order to capture the phonological representation of clicks, there are five major dimensions that must be accounted for. First, there are the four major dimensions used for all stop consonants: place of articulation, manner of articulation (including lateral contrasts), laryngeal setting, and nasality (see CHAPTER 22: CONSONANTAL

Table 18.1 Recognized and unrecognized click types

Click	Symbol	Source
bilabial	[ʘ]	IPA (2006)
dental	[ǀ]	IPA (2006)
alveolar (Nǀuu), postalveolar (Juǀ'hoansi)	[!]	IPA (2006)
lateral alveolar	[ǁ]	IPA (2006)
palatal	[ǂ]	IPA (2006)
retroflex	[‼]	Doke (1925), Snyman (1997), Miller *et al.* (2009c)

PLACE OF ARTICULATION, CHAPTER 13: THE STRICTURE FEATURES, CHAPTER 31: LATERAL CONSONANTS, and CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION for more discussion). In addition, clicks involve a non-pulmonic airstream, referred to as velaric (Beach 1938; Ladefoged and Traill 1984, 1994; Ladefoged and Maddieson 1996) or lingual (Miller *et al.* 2007; Miller *et al.* 2009a), which is unique to these consonants. Click languages exhibit a range of inventory sizes, and the languages exhibit a range of complexity on each of these five dimensions.

Place of articulation features for clicks are complicated, because both the anterior and posterior places of articulation must be accounted for, as well as the relative timing of these features. Historically, researchers thought that all clicks had a velar posterior place of articulation (Doke 1923; Beach 1938; Traill 1985; Ladefoged and Maddieson 1996). Thus it was thought that posterior place was not contrastive. Click types are therefore named according to the anterior place of articulation. Recent phonetic and phonological studies have provided evidence that the posterior constrictions of clicks are post-velar, and that some of the clicks involve tongue root retraction.

Six distinct click types have been attested in Khoesan languages. These are provided in Table 18.1. The bilabial click occurs only in the related Tuu languages !Xóõ (Traill 1985) and Nǀuu (Miller *et al.* 2009a), and in the Ju-ǂHoan language ǂHoan (Bell and Collins 2001).[1] Other languages with a bilabial click, such as ǀXam, are now extinct. The retroflex click has been described phonetically by Doke (1925) and Miller *et al.* (2009c) in Grootfontein !Xung. Snyman (1997) provides a survey of !Xung languages, and transcribes the retroflex click in several !Xung-speaking areas surrounding Grootfontein. Heine and König (2010) describe a different click in Ekoka !Xun as retroflex, but phonetic description of this click is lacking. Sands (2007) suggests that it may be a retracted lateral, based on preliminary acoustic analysis of a few tokens. The click corresponds historically to the palatal click in other Ju-ǂHoan languages, and behaves synchronically similarly to the palatal click with respect to the Back Vowel Constraint, described in §4.1.

Language inventories display a wide range of variation in the number and types of clicks. Table 18.2 lists the languages that contain one-, three-, four-, and five-click inventories. The Cushitic language Dahalo and the Bantu languages Rumanyo

[1] Ju-ǂHoan is the name of a language family, Juǀ'hoan or Juǀ'hoansi the name of a language.

Table 18.2 Number of click types in consonant inventories of languages

One-click inventories	Dahalo, Rumanyo, Southern Sotho, Mbukushu
Two-click inventories	—
Three-click inventories	Hadza, Xhosa, Zulu, Sandawe
Four-click inventories (all coronal clicks)	Ekoka !Xung, G|ui, Ju|'hoansi, Khoekhoe, Khwe, Mangetti Dune !Xung, Naro, Yeyi
Five-click inventories (including the labial click)	ǂHoan, Nluu, !Xóõ
Five-click inventories (all coronal clicks)	Grootfontein !Xung

Mbukushu has only one click type – the dental click.[2] Southern Sotho contains only the alveolar click type. There are no languages exhibiting two click types. The Bantu languages Xhosa and Zulu, and the isolates Hadza and Sandawe, each have three-click inventories, containing the dental, central alveolar, and lateral alveolar click types. Many languages have four-click inventories, containing the four coronal click types recognized by the IPA. These languages contain all languages of the Khoe-Kwadi group, as well as most of the members of the Ju branch of the Ju-ǂHoan family. The Ju-ǂHoan language Ekoka !Xung has four click types, but with the retracted lateral click type in place of the central palatal click type [ǂ].

Most of the languages that have five-click inventories are Tuu languages, but ǂHoan also has five click types. These languages include the four coronal click types recognized by the IPA, as well as the labial click type. However, Grootfontein !Xung has the four coronal click types recognized by the IPA, as well as a retroflex click type, [‼], yielding five coronal click types in all (Miller *et al*. 2009c).

Clicks also occur paralinguistically in many languages throughout the world. For example, a lateral click is used in English to encourage a horse to trot (Ladefoged 1982), and a dental click is used to express disapproval, commonly represented as *tsk-tsk*. Clicks are also used as discourse markers in English (Wright 2007), and have been reported to occur in an auxiliary language of Lardil, Damin, which has bilabial, dental, apico-alveolar, and apico-domal nasal clicks (Dixon 1980; Hale and Nash 1997). Clicks are also found in Chinese nursery rhymes (Nathan 2001). Extremely weak clicks have also been shown to occur in German, when alveolar–velar stop sequences overlap at word boundaries (Fuchs *et al*. 2007).

The phonetic characteristics of the retroflex click and the laterally released palato-alveolar click are still being explored, but Miller *et al*. (2009c) describe the retroflex click as a postalveolar click in Grootfontein !Xung. One of the speakers produced this click with a sub-apical contact, but there was both inter-speaker and intra-speaker variation. Grootfontein !Xung is a centrally located lect. Ju|'hoansi, spoken to the east, has an alveolar click in cognate words, and Mangetti Dune !Xung and other northern lects have a lateral alveolar click in the same words (Miller-Ockhuizen and Sands 1999).

[2] Both Rumanyo and Mbukushu clicks exhibit a lot of variation, and can be realized as dental, alveolar, or lateral alveolar click types.

The manner of articulation of clicks is relatively complex. All clicks involve two complete constrictions, and are thus non-continuants. Clicks contrast in the manner of articulation of their release properties. They can have either a release with a complete constriction or a fricated release, as in pulmonic affricates. Thus clicks can be either stops or affricates. The stop and fricative portions of the releases are mostly uvular, though there are additional release properties that may be epiglottal (Miller 2007). The affricates are also airstream contour segments, and will be discussed with the airstream contrasts in §6. There is clear phonological evidence from positional distribution that all click consonants, including nasal clicks, behave phonologically as obstruents. Thus clicks only occur in initial position of roots in the Ju-ǂHoan, Khoe-Kwadi, and Tuu language families.

Click languages display a large range of variation in terms of the complexity of laryngeal contrasts. Table 18.3 lists the languages according to the number of voice onset time (VOT) contrasts contained in the oral click inventories. Languages of the Ju branch of the Ju-ǂHoan language family, and the Tuu language !Xóõ, display the most complexity on the laryngeal dimension, with four-way VOT contrasts. The voiced aspirated clicks are parallel to the voiced aspirated pulmonic stops in Ju|'hoansi in exhibiting a voiceless interval, which Miller-Ockhuizen (2003) attributes to a larger glottal opening gesture than is found in Hindi voiced aspirates. The Bantu language Mbukushu has a simple two-way VOT contrast. The laryngeal settings are independent from the click mechanism. Eleven out of the 19 languages included in this survey contain a three-way VOT contrast.

The fourth dimension that must be accounted for in clicks is the airstream dimension. Miller *et al.* (2007) and Miller *et al.* (2009a) introduce the term "lingual airstream" to replace the term "velaric airstream." They note that the term "velaric" suggests that this airstream is somehow initiated by the velum or that it involves a velar stop, and provide evidence that the posterior release in clicks is uvular not velar. The term "lingual" reflects the anatomical source of air. The tongue is used to create a low-pressure cavity, the anterior release of which initiates the ingressive flow of air. Sagey (1990) argues that airstream is a matter of phonetic implementation, and does not need to be specified in the phonological representation of clicks. I provide evidence that N|uu contains sounds that differ solely in terms of airstream. There are stops that are produced fully with a lingual airstream mechanism, with the shift from ingressive lingual airstream to pulmonic egressive airstream occurring at the CV boundary between a click and a following vowel. These stops contrast solely with a class of stops that involve a shift from lingual ingressive airstream to a

Table 18.3 A subset of VOT contrasts found in click languages[a]

No voicing contrast	Dahalo, Hadza,[b] Khoekhoe
Two-way VOT contrast	Mbukushu
Three-way VOT contrast	Grootfontein !Xung, G|ui, ǂHoan, Xhosa,[c] Zulu, Khwe, Mangetti Dune !Xung, Naro, N|uu, Sandawe, Yeyi
Four-way VOT contrast	Ekoka !Xung, Ju|'hoansi, !Xóõ

[a] Rumanyo is not included in this table.
[b] Hadza has a three-way VOT contrast in pulmonic consonants, but no voicing or aspiration contrasts in the click inventory.
[c] Ladefoged and Traill (1994) refer to the Xhosa voiced click as murmured.

2 Background

Clicks occur primarily in southern African and east African languages in seven language families. The east African click languages, Hadza and Sandawe, are both spoken in Tanzania, and there has thus far been little diachronic evidence to link them genealogically to southern African Khoesan languages, or to each other (Sands 1998), though see Güldemann and Elderkin (2010) for evidence that Sandawe is related to the Khoe-Kwadi language family. Hadza is currently thought to be a language isolate. Dahalo is a Cushitic language with clicks (Maddieson *et al.* 1993) which is spoken in Kenya. Dahalo has thus far not been shown to be related to the geographically closest click languages, Hadza or Sandawe. However, Ten Raa (1969) has suggested that there might be a common substratum between Dahalo and Sandawe.

The majority of click languages are spoken in southern Africa. The non-Bantu southern African click languages were described by Greenberg (1966) as belonging to one language family called Khoisan. They were grouped together as the Northern, Southern, and Eastern Southern African Khoisan branch of the Khoisan family, along with Hadza and Sandawe. The southern African Khoesan languages are currently spoken mainly in Namibia, Botswana, and South Africa, but related languages were historically present in Zimbabwe, Zambia, and Angola, and there are still small pockets of speakers in these countries today (Brenzinger 2001).

Güldemann (2006) argues that the non-Bantu southern African click languages belong to at least three distinct families: the Khoe-Kwadi family, the Ju-ǂHoan family, and the Tuu family. He has demonstrated that similarities in the number of clicks in the inventories of Tuu and Khoe-Kwadi languages can be attributed to a substrate, rather than a genealogical relationship between these languages. I follow Güldemann in using the term "Khoesan" as a neutral way of referring to click languages that do not belong to the well-established Bantu or Cushitic families, with no implied genealogical relationship. I use "Khoesan" rather than"Khoisan" because the spelling matches the orthographies of the languages involved.

In southern Africa, clicks also occur in a number of Bantu languages, most notably southern Bantu languages of the Nguni group, including Zulu and Xhosa. These languages have been shown to have adopted clicks through the process of Hlonipa and regular borrowing (Herbert 1990), but they are now a fully functional part of the language inventories. Clicks also occur in the Namibian Bantu languages Rumanyo (Gciriku), Mbukushu, Mbalan'we, and Fwe (Baumbach 1997) as well as Yeyi, spoken in both Namibia and Botswana (Fulop *et al.* 2003), and Tumbuka, spoken in Malawi.

In this chapter, I focus mainly on phonological patterns found in Khoesan languages, though I also refer to evidence from Zulu. Many of the Khoesan languages are underdocumented and underdescribed. The patterns discussed in this chapter all come from a subset of languages that contain sufficient description of phonological patterns. The languages discussed, and their sources, are listed in Table 18.6.

3 Unit *vs.* cluster analyses of clicks

The presence of clicks in a consonant inventory increases the size of the inventory. However, there is a large range in inventory sizes among the different languages

Table 18.6 Click languages discussed in this chapter and their sources

Language	*Family*	*Source(s) of phonological description*
Dahalo	Cushitic	Maddieson *et al.* (1993)
Ekoka !Xung	Ju-ǂHoan	Heine & König (2010)
Grootfontein !Xung	Ju-ǂHoan	Doke (1925)
G|ui	Khoe-Kwadi	Nakagawa (2006)
Hadza	Isolate	Sands *et al.* (1993)
ǂHoan	Ju-ǂHoan	Bell & Collins (2001)
Ju|'hoansi	Ju-ǂHoan	Snyman (1970, 1975); Miller-Ockhuizen (2003)
Khoekhoe	Khoe-Kwadi	Beach (1938); Brugman (2009)
Khwe	Khoe-Kwadi	Kilian-Hatz (2003)
Mangetti Dune !Xung	Ju-ǂHoan	Miller *et al.* (2008)
Naro	Khoe-Kwadi	Visser (2001)
N|uu	Tuu	Miller *et al.* (2009a); Miller (2010a)
Rumanyo (Gciriku)	Bantu	Mölig & Shiyaka-Mbereme (2005)
Sandawe	Isolate	Wright *et al.* (1995); Hunziker *et al.* (2008)
!Xóõ	Tuu	Traill (1985, 1994)
Xhosa	Bantu	Ladefoged & Traill (1994)
Yeyi	Bantu	Fulop *et al.* (2003)
Zulu	Bantu	Doke (1926); Thomas-Vilakati (2010)

containing clicks. Güldemann (2006) provides inventory size, as well as the proportion of non-clicks to clicks in the inventories of Tuu and Khoe languages, showing a range of inventory sizes within each of these families. The Bantu language Zulu has 45 segments, 12 of which are clicks (Thomas-Vilakati 2010). The Tuu language !Xóõ has the largest number of segments that has been documented so far, with a total of 119 contrastive segments, 83 of which are clicks. N|uu has 73 segments, 45 of which are clicks. G|ui is described by Nakagawa (2006: 259) as having 53 segments and a number of clusters involving clicks, but he notes that under a unit analysis it would have 89 segments. Ju|'hoansi has 89 contrastive segments, 47 of which are clicks. ǂHoan has 55 clicks, and Khwe 32.

Traill (1993), Güldemann (2001), and Nakagawa (2006) argue that a click containing a pulmonic release should be represented as a consonant cluster comprised of a sequence of a lingual segment and a pulmonic segment. Güldemann's (2001) and Nakagawa's (2006) arguments for a cluster interpretation of clicks are that the interpretation decreases the size of the consonant inventory, and that under a cluster interpretation all of the second members of the cluster exist as independent segments in the inventory. Further, Güldemann (2001) argues that the simple stop types that are contained in the clusters are more frequent than the cluster types.

Ladefoged and Traill (1984: 11) note that cluster analysis of clicks for !Xóõ would result in clusters of segments that differ in voicing, which was a language type not known to exist at the time.

Miller *et al.* (2009a) argue that not all click release types in N|uu occur as single segments in that language. This is problematic for cluster analyses of clicks, because segments that occur in clusters almost always occur as single segments as well. Further, Miller *et al.* (2009a) argue that if airstream contours are represented as consonant clusters with clicks as their first members, then languages such as G|ui would be the only languages in the world with obstruent–obstruent clusters but not obstruent–sonorant clusters. A survey of cluster types by Kreitman (2008) found no languages with obstruent–obstruent clusters that did not also have obstruent–sonorant clusters. Kreitman's study did find languages with clusters of mixed voicing – most notably Modern Hebrew. I argue that large inventories arise in languages with clicks, because these languages make full use of airstream as a contrastive dimension. Miller-Ockhuizen (2003) shows that pulmonic stop–sonorant clusters that occur in loanwords from English and Afrikaans into Ju|'hoansi are broken up by epenthesis, providing phonological evidence that Ju|'hoansi does not allow stop–sonorant clusters.

(1) lists the inventory of G|ui consonants, including the interpretation of all of the clicks as single units, in the spirit of Miller *et al.* (2009a), and Nakagawa's (2006) interpretation of some of the clicks as clusters. Nakagawa's analysis transcribes a distinction between "velar" clicks and "uvular" clicks, and transcribes all clicks with either a "k" or a "q," following Ladefoged and Maddieson (1996). Miller *et al.* (2009a) and Miller (2010a) have shown, using ultrasound imaging of the tongue during click production in N|uu, that the posterior release of [ǂ] and [ǂ͡q], on the one hand, and [!] and [!͡q], on the other, occurs at the same location, and that in all four clicks the location is uvular. In [ǂ] and [ǂ͡q], the posterior release location is back uvular and does not involve tongue-root retraction, while in [!] and [!͡q], the posterior release location is front uvular and involves a ballistic tongue-root retraction movement. Miller *et al.* (2009a) argue that the glottal stop in N|uu is allophonic in vowel-initial words, and thus does not occur as a single segment in the language. Furthermore, the glottal fricative in N|uu and most Khoesan languages is voiced, and thus does not correspond to the voiceless aspiration found in the voiceless nasal aspirated click.

(1)

a. *Unit analysis*	*Cluster analysis*	b. *Unit analysis*	*Cluster analysis*
!	/k!/	$!^{χ}$	/k!/ + /χ/
$^{g}!$	/g!/	$!^{χ'}$ or $!^{H}$	/k!/ + /qχ'/[4]
$^{ŋ}!$	/ŋ!/	!q	/k!/ + /q/
$!^{h}$	/k!ʰ/	ɢ!	/k!/ + /ɢ/
$!^{ʔ}$	/k!'/	$!q^{h}$	/k!/ + /qʰ/
		!q'	/k!/ + /q'/
		$^{ŋ}!^{ʔ}$	/k!/ + /ʔ/
		$^{ŋ}!^{h}$	/k!/ + /h/

[4] Nakagawa notes that this click and the similar pulmonic consonant [qχ'] are both realized with a lateral release. However, he notes that this click does not spread the lateral feature to a medial consonant the way that lateral clicks do. He thus claims that the lateral release is phonetic detail.

4.1 *The Back Vowel Constraint*

The Back Vowel Constraint is a CV co-occurrence restriction that blocks the co-occurrence of front vowels with a class of consonants, including labial and alveolar clicks (see also CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). In this section, I discuss the BVC patterns in G|ui (Nakagawa 2006), Ju|'hoansi (Miller-Ockhuizen 2003), N|uu (Miller 2010a), and !Xóõ (Traill 1985; Sagey 1990), and provide different analyses of the BVC. Most authors (Traill 1985; Sagey 1990; Clements and Hume 1995; Miller 2010a) analyze the BVC as assimilation between the posterior place of articulation of clicks and following front vowels. Earlier analyses viewed the posterior constrictions in clicks as velar, based on early phonetic studies. Miller (2010a) views the posterior constrictions of clicks as uvular, based on her ultrasound investigations of the posterior constrictions in clicks (Miller *et al.* 2007; Miller *et al.* 2009a; Miller 2010a), and she states the BVC as assimilation to an [RTR] feature of clicks. Nakagawa (2006) analyzes the BVC as assimilation between the anterior tongue shape of clicks and following front vowels. Miller's (2010a) and Nakagawa's (2006) analyses capture different patterns of linguo-pulmonic contour segments (Miller *et al.* 2009a) in N|uu and the phonetically similar clicks that are analyzed as clusters of clicks followed by uvular pulmonic segments in G|ui. The two analyses also make different predictions about the patterns found with non-click consonants, particularly labials.

Previous phonetic literature has described the posterior constriction in all clicks as velar (Doke 1923; Beach 1938; Ladefoged and Traill 1984, 1994; Traill 1985; Ladefoged and Maddieson 1996), and phonological representations of clicks have captured the posterior constriction with the phonological feature [+back] (Chomsky and Halle 1968; Sagey 1990) or the feature [dorsal] (Traill 1993; Clements and Hume 1995). The initial description of the BVC (Traill 1985) in !Xóõ treats all clicks, as well as velar stops, as being subject to the constraint. The BVC, stated in (4), changes a front vowel to a [+back] vowel by assimilation to the [+back] feature of the click.

(4) *Back Vowel Constraint* (Traill 1985)

if:	C_1	V_1
	<+back>	
then:	C_1	V_1
	<+back>	<+back>

A second process, Dental Assimilation, changes back vowels following dental [|] and palatal [ǂ] clicks to front vowels, in effect undoing the BVC. Chomsky and Halle (1968) and Sagey (1990) use the feature [+anterior] to classify the dental and palatal clicks separately from the central and lateral alveolar clicks. The Dental Assimilation rule in Sagey (1990) crucially requires palatal clicks to be [+anterior], in keeping with the fact that the palatal click has a long laminal constriction stretching from the dental place of articulation to the post-alveolar place of articulation (see Sands *et al.* 2007 and references therein).

Miller-Ockhuizen (2003) argues that there are two classes of clicks: those that are subject to the BVC, and those that are not. She proposes that the central alveolar [!] and lateral alveolar [||] clicks that are subject to the BVC have a uvular posterior constriction, which is specified for a [pharyngeal] feature. She proposes the BVC in (5), using Clements and Hume's (1995) feature theory (see CHAPTER 27: THE ORGANIZATION OF FEATURES):

(5) *Back Vowel Constraint* (Miller-Ockhuizen 2003)

*{[pharyngeal]$_{\text{Vplace}}$ [coronal]$_{\text{Vplace}}$ }$_{\sigma}$

[pharyngeal] and [coronal] cannot be specified on the same or different V-place within a syllable.

She argues that front vowels are underlyingly present in Ju|'hoansi following dental and palatal click types, and dental and palatal clicks are unspecified for the feature [pharyngeal].

Traill (1997) proposes the acoustic feature [+grave] to account for BVC patterns in !Xóõ, which also contains the labial click. His analysis captures the fact that labial and alveolar clicks pattern together with respect to the BVC. However, Miller-Ockhuizen (2000) shows that pulmonic labial consonants and labial clicks pattern differently in !Xóõ, and thus this feature is problematic. Traill (1985) considered initial labial consonants to be only found in loanwords in !Xóõ. Labial consonants are low frequency in root-initial position, but extremely frequent in root-medial position in most Khoesan languages (see Miller-Ockhuizen 2003: 131 for Ju|'hoansi and Miller (2010a) for N|uu).

Miller (2010a) provides the results of a database study over the N|uu lexicon. The phonological patterning of N|uu consonants with respect to the BVC is summarized in Table 18.7. Miller (2010a) analyzes the BVC as assimilation of a front vowel to an [RTR] feature on the consonant.

Table 18.7 Patterning of N|uu consonants with respect to the BVC

	Occur with front and back vowels	*Occur with back vowels and the [əi] and [æ] allophones of /i/ and /e/*
Pulmonic stops	Labial [b β] Alveolar [s z ɾ] Palatal [c ɟ ʃ ʒ] Velar [k g]	Uvular [q χ kχ']
Clicks	Dental click [ǀ] Palatal click [ǂ]	Labial click [ʘ] Central alveolar click [!] Lateral alveolar click [ǁ]
Clicks with airstream contours (stops)	Dental clicks [ǀ͡q] Palatal clicks [ǂ͡q]	Labial click [ʘ͡q] Central alveolar click [!͡q] Lateral alveolar click [ǁ͡q]
Clicks with airstream contours (affricates)		Uvularized labial click [ʘ͡χ] Uvularized dental click [ǀ͡χ] Uvularized central alveolar click [!͡χ] Uvularized lateral alveolar click [ǁ͡χ] Uvularized palatal click [ǂ͡χ]

The patterning of dental and palatal linguo-pulmonic stops, [ǀq͡] and [ǂq͡], in Nǀuu provides evidence that these clicks are unspecified for [RTR] (see CHAPTER 25: PHARYNGEALS). Herzallah (1990), Elorietta (1991), McCarthy (1994), and Rose (1996) have shown that both uvulars and pharyngeals cause retraction of front vowels in Semitic languages. Uvular and pharyngeal sounds are thus generally characterized as all being specified for the feature [RTR]. The Nǀuu click patterns suggest that there are uvular constrictions that don't involve the feature [RTR]. Phonetically, the posterior place of articulation in Nǀuu palatal clicks is back uvular, at the point where the oral tract and pharynx meet, and it does not involve tongue-root retraction (Miller *et al.* 2009a; Miller 2010a). The posterior constriction in dental and palatal clicks is more like the production of [u] (Miller *et al.* 2009b), which involves tongue-root raising (Esling 2005), not tongue-root retraction. Miller (2010a) notes that these differences in articulation are consistent with their differences in featural specification. The alveolar click is specified for the feature [RTR], while the palatal click is unspecified for the feature.

Nakagawa (2006) describes the patterning of Gǀui consonants with respect to the BVC. The patterns seen in Gǀui largely mirror those found in Nǀuu, except that the click clusters with uvular offsets, [ǀq !q ǁq ǂq] (Miller *et al.*'s 2009a linguo-pulmonic stops), are all subject to the BVC in this language. Nakagawa accounts for the BVC in Gǀui using the two constraints in (6). One constraint accounts for the click patterns, and the other for the non-click patterns. Nakagawa notes that Cq represents a non-click uvular consonant, i.e. /q ɢ q^h q' qχ' χ/, whether it occurs as an independent consonant or as a cluster offset, and "!/ǁ" stands for an apical click. The alveolar clicks would thus be captured with his feature [+apical], and the uvulars would be captured with the feature [+grave], though this would predict that velar and labial pulmonic consonants are also subject to the constraint. However, Nakagawa (2006: 229) notes that there are 17 words in Gǀui that have a velar consonant followed by a [−back] vowel. Nakagawa attributes the presence of front vowels following some velars to the phonetic ambiguity of velars.

(6) a. *Cq V[−back]
    b. *!/ǁ V[−back]

Miller-Ockhuizen (2003) notes that the velar stops [k] and the velar fricative "x" (Snyman 1970, 1975; Dickens 1994) pattern differently in Juǀ'hoansi, and she re-analyzes the fricative as uvular [χ] and the velarized clicks [ǀx !x ǁx ǂx] as uvularized clicks [$ǀ^χ$ $!^χ$ $ǁ^χ$ $ǂ^χ$], based on the raised F1 values in vowels following these sounds. The place difference accounts for the different BVC patterns, as the uvular fricative and the uvularized clicks are [RTR]. This same disparity in [k] and "x" is found in Gǀui (Nakagawa 2006: 231), although Nakagawa maintains the traditional description of the dorsal fricative as "velar" and the [ǀx !x ǁx ǂx] clicks as velarized. As a result, he describes velars as patterning ambiguously. The velar stop [k] in !Xóõ is phonologically ambiguous (Traill 1985). There are both forms with the [əi] allophone of /i/, as in *kəi* 'grow', and the [e] allophone of [e] as in the word *kèhn* 'being' in Traill's (1994) dictionary. Further, Traill (1985) notes that the concordially determined forms *ki* (class 1) and *ke* (class 3) are often realized in speech as [ti] and [te], due to the BVC, and Afrikaans loanwords containing [k] are realized with [t] in !Xóõ. The dorsal fricative "x" occurs only with

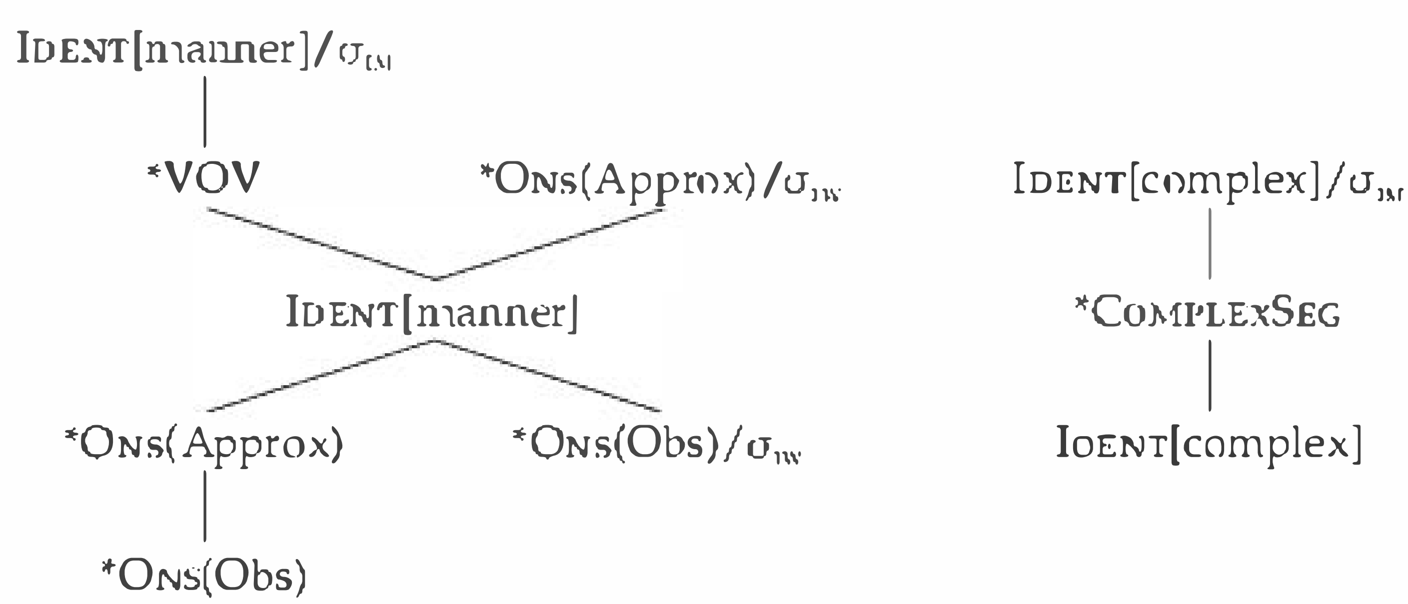
Table 18.8 Distribution of segments in Khoekhoe (Brugman 2009)

Segment	*Root-initial*	*Clitic-initial*	*Root-medial*
[ǀ ǀx ᵑǀ ᵑǀʰ ᵑǀʼ] [ǃ ǃx ᵑǃ ᵑǃʰ ᵑǃʼ] [ǁ ǁx ᵑǁ ᵑǁʰ ᵑǁʼ] [ǂ ǂx ᵑǂ ᵑǂʰ ᵑǂʼ]	✓		
[p t k ʔ] [ts kx] [s x h]	✓	✓	(p)
[m n]	✓	✓	✓
[ɾ]		✓	✓
[β]			✓

(9) *ONS(Approximant) >> *ONS(Nasal) >> *ONS(Obstruent)

These constraints are ranked with IDENT[manner] constraints, and the constraints are relativized to different positions. In order to capture the unique positioning of clicks, she uses the constraint *COMPLEXSEG. The complete ranking of constraints that accounts for the Khoekhoe patterns is provided in (10), where "IW" is "initial" and "IM" is "medial."

(10) *Constraint ranking for Khoekhoe segment distribution* (Brugman 2009)



Additional evidence for *COMPLEXSEG comes from the distribution of clicks in three different forms of pronouns found in N|uu. Table 18.9 provides a list of three forms of all of the pronouns found in N|uu, which Collins and Namaseb (2007) describe as the simple form, the click form, and the "a" form.

The distribution of these forms of pronouns is syntactically determined (Collins and Namaseb 2007). The simple forms occur in subject position (for non-questions), object position, and possessor position. Collins and Namaseb note that they are not syntactically weak in the traditional sense, since they can appear in coordinate structures, can be modified by numerals and "only," and can be clefted. The click forms of the pronouns appear obligatorily as the subject in yes–no questions

Table 18.10 !Xóõ onsets in different morpho-prosodic positions (Brugman 2009)

Segment	*Root-initial*	*Clitic-initial*	*Root-medial*
clicks	✓		
[t' tˣ dˣ tˣ' dˣ'] [ts' dz' tsˣ dzˣ tsˣ' dzˣ']	✓		
[ˀm ʔn]	✓		
[tʰ tsʰ d dʰ dz dzʰ]	✓		
[q ʔ ts ʃ x h]	✓		
[p t k s]	✓	✓	
[m n]	✓	(n)	✓
[ɾ]		✓	✓
[b j ɲ]			✓

Table 18.11 Featural specification of guttural consonants and vowels

Value for [pharyngeal]	[pharyngeal]		*Unmarked for* [pharyngeal]	
[spectral slope] *value*	[high]	[low]	[high]	[low]
aspirated consonants/breathy vowels			✓	
glottalized consonants/glottalized vowels				✓
uvularized consonants	✓			
epiglottalized consonants/epiglottalized vowels		✓		

A constraint, ALIGN([spectral slope]$_V$, L; μ_1, L), is proposed by Miller-Ockhuizen (2003: 138) to account for the distribution of guttural vowels within the first syllable of a root. Given the recent advances in positional augmentation, we can capture the distribution of both guttural consonants and vowels in the first syllable of a root by ranking the constraint *GUTTURAL below IDENT[Guttural]/σ_{1W}, and above IDENT[Guttural].

G|ui limits the medial consonants to the set of /b w r j m n/ (Nakagawa 2006: 114), and allows the full set of 89 consonants (or 53 segments plus all possible clusters) initially. Thus, G|ui seems also to have a constraint licensing obstruents in initial position, and sonorants in medial position, just like the other Khoesan languages. However, Nakagawa (2006) reports monomoraic non-root words that contain clicks in G|ui. It is possible that these words may have the same prosodic status as roots, similar to the post-positions and complementizers in Khoekhoe (Brugman 2009: 202). It is also possible that the constraints are ranked differently in G|ui. Clicks are limited to root-initial position in all of the other documented Ju-ǂHoan, Khoe-Kwadi, and Tuu languages, though their distribution in non-roots is unexplored.

a re-analysis in either of these languages. The diachronic evidence from click change provided in §3 suggests that, at least historically, these segments may have also differed in posterior place phonologically. Without further phonetic studies, we do not know if these clicks may also differ, or have differed in the past, in speakers' productions or in listeners' perceptions. Any theory that accounts for clicks needs to account for airstream contour segments in N|uu.

7 Conclusion

I have provided an overview of the types of evidence that are available for place of articulation, manner of articulation, and airstream of clicks. I have given an overview of unit analysis and cluster analysis of click consonants. I have noted that the unit analysis of clicks results in larger click inventories. Cluster analyses of clicks, on the other hand, result in smaller segment inventories. This comes at a cost, as it results in Khoesan languages being the only languages in the world that allow obstruent–obstruent clusters without also allowing obstruent–sonorant clusters.

There are three main sources of evidence for the phonology of clicks. The Back Vowel Constraint and nasal assimilation provide evidence as to the place of articulation in clicks. The positional distribution of clicks provides evidence that clicks are obstruents, and that their status as complex segments makes them more marked. The existence of airstream contour segments in Nluu provides evidence that airstream is an independent dimension of sound structure. All of these types of evidence must be accounted for by any phonological theory that accounts for clicks.

REFERENCES

Baumbach, Erdmann. 1997. Bantu languages of the eastern Caprivi. In Wilfred Haacke & Edward Elderkin (eds.) *Namibian languages: Reports and papers*, 307–451. Cologne: Rüdiger Köppe Verlag.

Beach, D. M. 1938. *The phonetics of the Hottentot language*. Cambridge: William Heffer & Sons.

Bell, Arthur & Chris Collins. 2001. ǂHoan and the typology of click accompaniments in Khoisan. *Cornell Working Papers in Linguistics* 18. 126–153.

Bennett, William G. 2008. False nasality and the typology of click distribution. Unpublished ms., Rutgers University.

Brenzinger, Matthias. 2001. Angola and Zambia. In Steven Robins, Elias Madzudzo & Matthias Brenzinger (eds.) *An assessment of the status of the San in South Africa, Angola, Zambia and Zimbabwe*, 63–75. Windhoek: Legal Assistance Centre.

Brugman, Johanna. 2009. Segments, tones and distribution in Khoekhoe prosody. Ph.D. dissertation, Cornell University.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 245–306. Cambridge, MA & Oxford: Blackwell.

Collins, Chris & Levi Namaseb. 2007. Nluuki grammar. Unpublished ms., New York University.

Ladefoged, Peter. 2007. Articulatory features for describing lexical distinctions. *Language* 83. 161–180.
Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages.* Oxford & Malden, MA: Blackwell.
Ladefoged, Peter & Anthony Traill. 1984. Linguistic phonetic description of clicks. *Language* 60. 1–20.
Ladefoged, Peter & Anthony Traill. 1994. Clicks and their accompaniments. *Journal of Phonetics* 22. 33–64.
Maddieson, Ian & Peter Ladefoged. 1989. Multiply articulated segments and the feature hierarchy. *UCLA Working Papers in Phonetics* 72. 116–138.
Maddieson, Ian, Siniša Spajić, Bonny Sands & Peter Ladefoged. 1993. The phonetic structure of Dahalo. *UCLA Working Papers in Phonetics* 84. 25–65.
McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.) *Phonological structure and phonetic form: Papers in laboratory phonology III*, 191–233. Cambridge: Cambridge University Press.
McCarthy, John J. & Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology 1993*. 79–153.
Miller, Amanda. 2007. Guttural vowels and guttural co-articulation in Ju|'hoansi. *Journal of Phonetics* 35. 56–84.
Miller, Amanda. 2010a. Tongue body and tongue root shape differences in N|uu clicks correlate with phonotactic patterns. In Susanne Fuchs, Martine Toda & Marzena Żygis (eds.) *Turbulent sounds: An interdisciplinary guide*, 245–280. Berlin & New York: Mouton de Gruyter.
Miller, Amanda. 2010b. A prosodic account of Ju|'hoansi consonant distributional asymmetries. In Matthias Brenzinger & Christa König (eds.) *Proceedings of the 2nd International Symposium on Khoisan Linguistics*, 53–84. Cologne: Rüdiger Köppe Verlag.
Miller, Amanda, Johanna Brugman, Bonny Sands, Levi Namaseb, Mats Exter & Chris Collins. 2009a. Differences in airstream and posterior place of articulation among N|uu clicks. *Journal of the International Phonetic Association* 39. 129–161.
Miller, Amanda, Levi Namaseb & Khalil Iskarous. 2007. Posterior tongue body constriction locations in clicks. In Jennifer Cole & José Hualde (eds.) *Laboratory phonology* 9. 643–656. Berlin & New York: Mouton de Gruyter.
Miller, Amanda, Bonny Sands & Sheena Shah. 2008. Field notes on Mangetti Dune !Xung. Unpublished ms.
Miller, Amanda, Abigail Scott, Bonny Sands & Sheena Shah. 2009b. Rarefaction gestures and coarticulation in Mangetti Dune !Xung clicks. In Maria Uther, Roger Moore & Stephen Cox (eds.) *Proceedings of the 10th Annual Conference of the International Speech Communication Association*, 2279–2282. Rundle Mall, Australia: Causal Productions.
Miller, Amanda, Sheena Shah & Bonny Sands. 2009c. Five coronal click types in Grootfontein !Xung. Paper presented at the 83rd Annual Meeting of the Linguistic Society of America, San Francisco.
Miller-Ockhuizen, Amanda. 2000. C-V coarticulation and complex consonants: Evidence for ordering in click place gestures. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.) *Item order in language and speech*, 301–330. Prague: Karolinum Press.
Miller-Ockhuizen, Amanda. 2003. *The phonetics and phonology of gutturals: A case study from Ju|'hoansi.* New York: Routledge.
Miller-Ockhuizen, Amanda & Bonny Sands. 1999. !Kung as a linguistic construct. *Language and Communication* 19. 401–413.
Möhlig, Wilhelm & Karl Peter Shiyaka-Mbereme. 2005. *A dictionary of the Rumanyo language.* Cologne: Rüdiger Köppe Verlag.
Nakagawa, Hiroshi. 2006. Aspects of the phonetic and phonological structure of the G|ui language. Ph.D. dissertation, University of Witwatersrand.

Nathan, Geoffrey S. 2001. Clicks in a Chinese nursery rhyme. *Journal of the International Phonetic Association* 31. 223–228.

Rose, Sharon. 1996. Variable laryngeals and vowel lowering. *Phonology* 13. 73–117.

Sagey, Elizabeth. 1990. *The representation of features in nonlinear phonology: The articulator node hierarchy*. New York: Garland.

Sands, Bonny. 1998. *Eastern and southern African Khoisan: Evaluating claims of distant linguistic relationship*. Cologne: Rüdiger Köppe Verlag.

Sands, Bonny. 2007. The contribution of language documentation to historical phonology. In Peter K. Austin, Oliver Bond & David Nathan (eds.) *Proceedings of the Conference on Language Documentation and Linguistic Theory*, 209–219. London: School of Oriental and African Studies.

Sands, Bonny, Johanna Brugman, Mats Exter, Levi Namaseb & Amanda Miller. 2007. Articulatory characteristics of anterior click closures in N|uu. In Trouvain & Barry (2007), 401–404.

Sands, Bonny, Ian Maddieson & Peter Ladefoged. 1993. The phonetic structures of Hadza. *UCLA Working Papers in Phonetics* 84. 67–87.

Smith, Jennifer L. 2000. Prominence, augmentation, and neutralization in phonology. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 26. 247–257.

Smith, Jennifer L. 2002. Phonological augmentation in prominent positions. Ph.D. dissertation, University of Massachusetts, Amherst.

Snyman, Jan W. 1970. *An introduction to the !Xũ language*. Cape Town: A. A. Balkema.

Snyman, Jan W. 1975. *Žu|'hõasi fonologie en woordeboek*. Cape Town: A. A. Balkema.

Snyman, Jan W. 1997. A preliminary classification of the !Xũũ and Žu|'hõasi dialects. In Wilfred Haacke & Edward Elderkin (eds.) *Namibian Languages: Reports and Papers*, 21–106. Cologne: Rüdiger Köppe Verlag.

Ten Raa, Eric. 1969. Sanye and Sandawe: A common substratum? *African Language Review* 8. 148–155.

Thomas-Vilakati, Kimberley D. 2010. *Coproduction and coarticulation in IsiZulu clicks*. Los Angeles: University of California Press.

Traill, Anthony. 1979. The !Xóõ strength hierarchy. In Jan W. Snyman (ed.) *Bushman and Hottentot linguistic studies*, 1–38. Pretoria: University of South Africa.

Traill, Anthony. 1985. *Phonetic and phonological studies of !Xóõ Bushman*. Hamburg: Buske.

Traill, Anthony. 1993. The feature geometry of clicks. In Paul M. S. von Staden (ed.) *Linguistica: Festschrift E. B. van Wyk 'n Huldeblyk*, 134–140. Pretoria: J. L. van Schaik.

Traill, Anthony. 1994. *A !Xóõ dictionary*. Cologne: Rüdiger Köppe Verlag.

Traill, Anthony. 1997. Linguistic phonetic features for clicks: Articulatory, acoustic and perceptual evidence. In Robert K. Herbert (ed.) *African linguistics at the crossroads: Papers from Kwaluseni*, 99–117. Cologne: Rüdiger Köppe Verlag.

Trouvain, Jürgen & William J. Barry (eds.) 2007. *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken: Saarland University.

Visser, Hessel. 2001. *Naro dictionary: Naro–English, English–Naro*. Gansi, Botswana: Naro Language Project.

Wright, Melissa. 2007. Clicks as markers of new sequences in English conversation. In Trouvain & Barry (2007), 1069–1072.

Wright, Richard, Ian Maddieson, Peter Ladefoged & Bonny Sands. 1995. A phonetic study of Sandawe clicks. *UCLA Working Papers in Phonetics* 91. 1–24.

19 Vowel Place

Bruce Morén-Duolljá

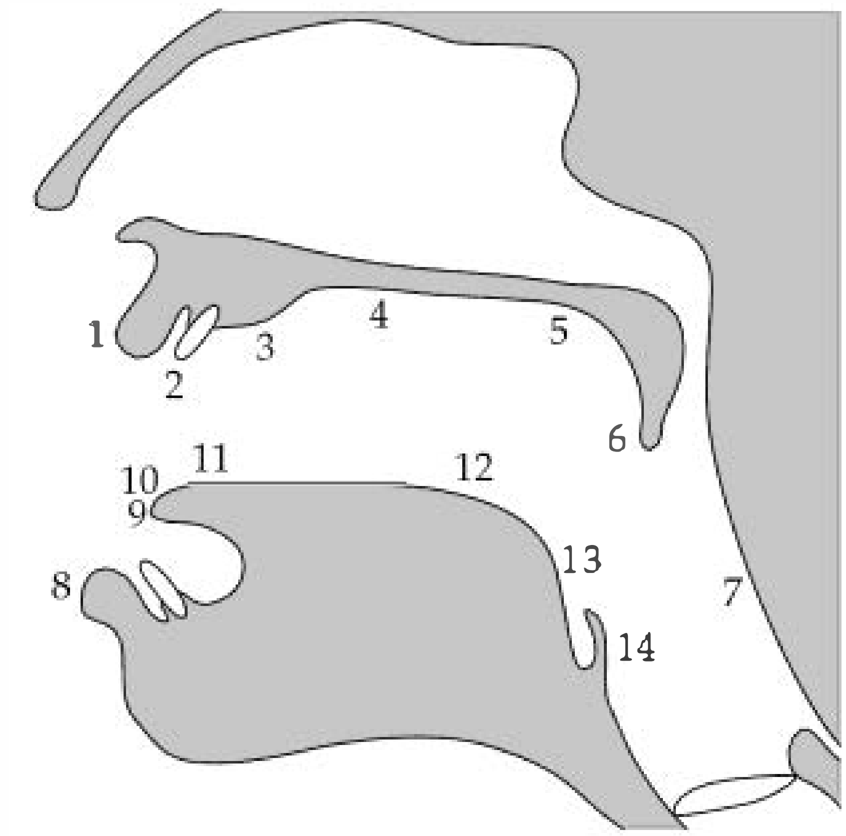
What is vowel place and what role does it play in the sound patterns of language? The term "vowel place" means different things to different people. There are two main reasons for this. First, the relationship between the articulation and the acoustics of "place of articulation" is complex and multifaceted. Second, phonological features have a range of model-specific definitions. The purpose of this chapter is to give a general introduction to the phonetic and phonological issues necessary to understand vowel place.

1 The phonetics of "place"

At its most basic, "the behavior of the vocal tract in speech can be described as an alternation of closing and opening" (Chomsky and Halle (*SPE*) 1968: 301). These closings and openings result in a complex set of aerodynamic and acoustic effects. Depending on the amount of the contraction or expansion, combined with language-particular phonotactics, a resulting acoustic signal is interpreted as a consonant and/or a vowel. The type and degree of the vocal tract openings and closings on the articulatory side of things corresponds roughly (but, as we will see, not directly) to what are commonly and descriptively referred to as phonological major class features, consonant manner features, and vowel height features (CHAPTER 21: VOWEL HEIGHT).[1] However, it is not enough to look at sound patterns only in terms of constriction. One must also attend to the precise location at which constrictions occur within the vocal tract.

We have known since the groundbreaking work of Chiba and Kajiyama (1942) that constrictions and expansions along the vocal tract have particular acoustic effects that can be explained in terms of tube acoustics. Depending on the proximity of a constriction or expansion to a node (i.e. point of minimal air displacement) or a loop (i.e. point of maximal air displacement) associated with a tube's resonance frequency, there will be different effects on that resonance frequency. A constriction at a loop causes a decrease in frequency, and a constriction at a node causes an increase. It is the location along the length of the vocal tract where articulatory gestures take place, combined with their acoustic correlates, that corresponds to

[1] I will not discuss openings and closings of the glottis.



Passive articulators
1 upper lip
2 upper teeth
3 alveolar ridge
4 hard palate
5 velum (soft palate)
6 uvula
7 pharyngeal wall

Active articulators
8 lower lip
9 tongue tip
10 tongue blade
11 tongue front
12 tongue dorsum
13 tongue root
14 epiglottis

Figure 19.1 Vocal tract diagram

a large extent (but not absolutely) to what is commonly known as phonological place of articulation.

From an articulatory perspective, place involves passive and active articulators (see Figure 19.1). The former are static landmarks along the upper surface of the vocal tract. The latter are mobile articulators along the lower surface of the vocal tract that move nearer to or further from the static landmarks during speech.

If one looks at the documented combinations of active and passive articulators found cross-linguistically, one sees evidence of all physically felicitous combinations. Further, neither active nor passive articulator seems to be privileged in having primary responsibility for defining place across contexts. For example, the passive uvula articulator can be used to single out the set of uvulars, and the active epiglottis articulator can be used to single out the set of epiglottals. However, the active tongue dorsum articulator is used for both dorsals and uvulars, whereas the passive pharyngeal wall articulator is used for both pharyngeals and epiglottals.

Combinations of active and passive articulators result in particular shapes of the vocal tract and cause the general acoustic effects described by Chiba and Kajiyama. To illustrate this in simplified terms, we can look at the schematized vocal tract in Figure 19.2. Here we see a tube closed at one end by the larynx and open at the other end by the lip orifice. The first three formants (F1, F2, and F3) are indicated by sine waves, and loops and nodes associated with these formants are indicated by solid and dashed vertical lines, respectively. The F1/F2/F3 loop labeled A corresponds to the labial region, the F3 node labeled B corresponds to the alveolar region, the F2 node C and F3 loop D correspond to the velar region, and E, F, and G correspond to the pharyngeal region.

The acoustic results of contractions at the various loops and nodes in the vocal tract are seen most clearly in F2 transition targets associated with stops. These correspond quite nicely to what are commonly described as the primary consonant places. This is illustrated in Figure 19.3 for labial, alveolar, and velar stops in particular vowel contexts. We see that labial constrictions have a low F2 target

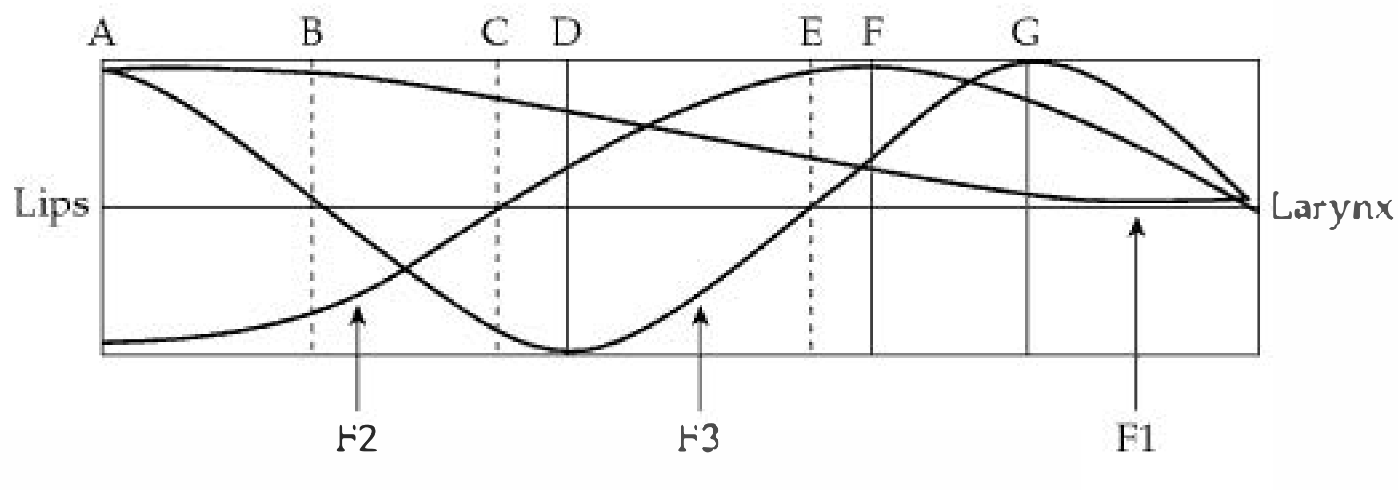


Figure 19.2 Illustration of vocal tract tube acoustics

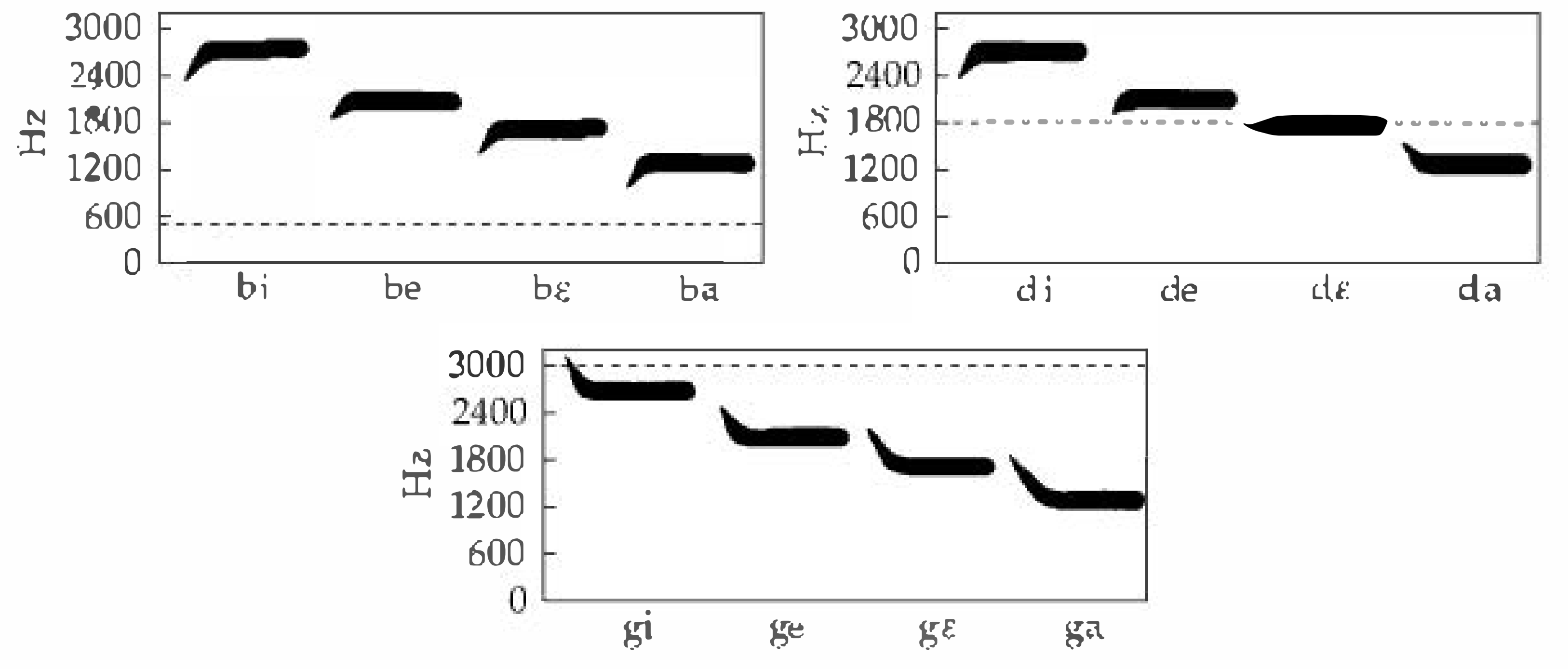


Figure 19.3 F2 schematics for several consonant–vowel combinations. Dashed lines indicate stop transition targets (modified from Delattre *et al.* 1955)

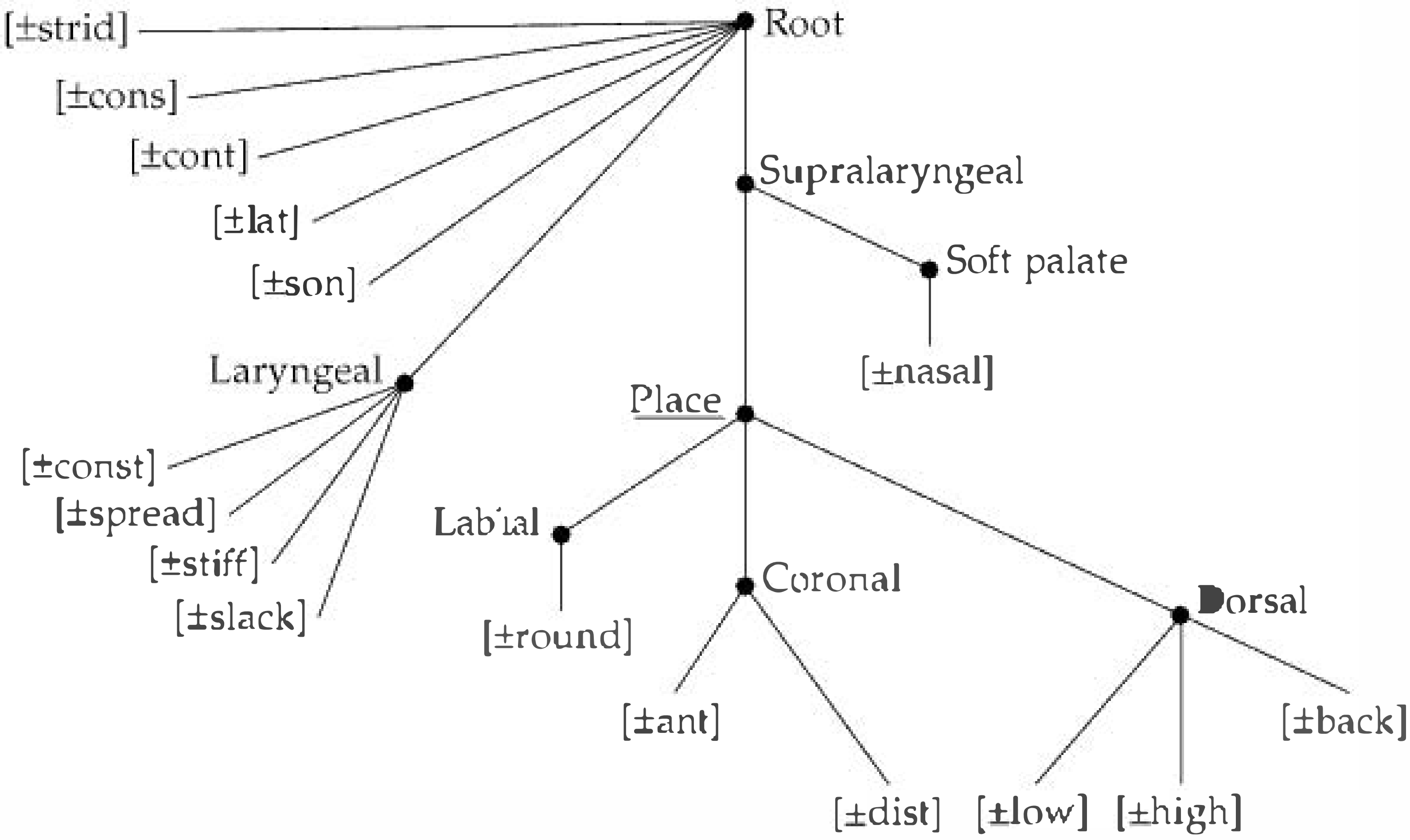
(below 600 Hz). This is because the labial region is a loop for F2. Alveolar constrictions have a mid-range F2 target (approximately 1800 Hz), because this region is neither a loop nor a node for F2. Velars have a high F2 target (above 2500 Hz), because this is a node for F2.[2]

Although there are complications to this story when one looks carefully at particular contexts and at combinations of articulations in running speech, there is general consensus in the literature with respect to the relationship among some consonant articulations, acoustics, and phonological place features. This is in no small part due to the rather extreme constrictions that consonants require for their production. These allow for a robust and precise articulatory and acoustic characterization of consonants. Further, these extreme articulations and their natural proximity to resonance nodes and/or loops also produce "quantal effects" (Stevens 1972, 1989), where relatively large movements along some regions of the vocal tract produce only small acoustic results, while small movements along other regions produce large acoustic results. Thus it is not surprising that certain consonant places are quite common across the world's languages. It is also not

[2] Although not shown in Figure 19.3, the close proximity of an F2 node and F3 loop in the velar region results in the well-known "velar pinch" typical of velar consonant formant transitions in some contexts – i.e. F2 is high and F3 is low (Stevens 1989).

[dorsal]) are dependent. This idea has been modified a number of times over the years, for example in (1).

(1) *Feature geometry of Sagey (1990)*



It is important to note that this type of geometry groups traditional place features and non-place features under "place" organizing structures. For example, Sagey uses place, labial, coronal, and dorsal organizing "place" nodes that have dependent [±round], [±anterior], and [±back] "place" features and dependent [±distributed], [±low], and [±high] "non-place" features.

Although there are language-particular phonetic and phonological facts that have led to this proposal, there are a number of unanswered questions that must be asked. For example, what is the formal difference between a place node and a place feature? Why can some structures, such as [coronal], be both a feature and an organizing node for dependent features, while other structures, such as C-place, V-place, and aperture, can only be an organizing node? Is there a formal reason for this? What is "place" in a model where both place features and non-place features are dependent on the same "place" nodes? It is beyond the scope of this chapter to answer these questions, but it is important to think about what "place" really means and how it might be modeled.

To summarize, we have to contend with at least three meanings of the term "vowel place" in the literature. It can be a descriptive term for related articulatory and/or acoustic characteristics. It can be a descriptive term for related phonological features. It can be a formal term for segment-internal organizing structure.

4 The phonology of vowel place

If we assume there is phonological vowel place from a featural and/or an organizational perspective, we have to ask ourselves what it is used for. There are seven major types of phenomena that are relevant here:

(2) a. Vowel contrasts
 b. Consonant contrasts
 c. Vowel interactions
 d. Consonant–vowel interactions
 e. Consonant–vowel alternations
 f. Vowel neutralizations
 g. Consonant interactions

The first five will be discussed below.

4.1 *Vowel place contrasts*

The most obvious use of vowel place features is to establish contrasts among vowels (see CHAPTER 2: CONTRAST). However, given the interactions among vowel articulations and acoustics discussed above, it is difficult to establish absolute and precise articulatory or acoustic characteristics for vowel place and translate these directly to phonological structures. It is thus common to describe vowel inventories via a combination of articulatory configuration and relative position along F1 and F2 continua in charts such as that in Figure 19.5. The vertical axis represents F1 and relative openness of the oral cavity (see CHAPTER 21: VOWEL HEIGHT). The horizontal axis represents F2 and relative location along the front–back dimension of the oral cavity, where there is a constriction or expansion.

What is important for our purposes is the number of differences along the horizontal axis if relative F1/openness/tongue height is controlled for. It is also important to relate those differences to articulator configurations and phonological patterns. As we see in Figure 19.5, there seem to be six major categories along the F2 acoustic dimension for most vowel heights[3] – [i y ɨ ʉ ɯ u], [e ø ɘ ɵ ɤ o], and [ɛ œ ɜ ɞ ʌ ɔ], and these divide articulatorily into three main areas based on tongue position (i.e. front, central, and back), with a further split based on lip configuration (i.e. rounded and unrounded).

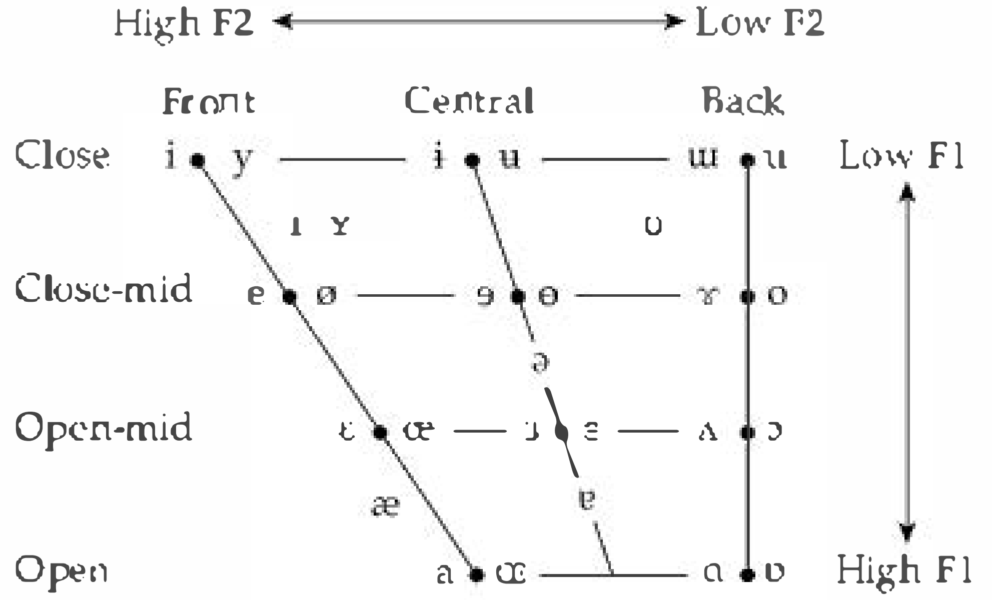


Figure 19.5 Standard vowel chart (IPA)

[3] I do not discuss here the important work of Lass (1984), which claims that there is evidence for ten place distinctions. This claim is not based on contrasts within the speech of individual speakers, but on what he calls "dialect distinguishability." Because there is no evidence that so many contrasts are needed within a given phonological grammar, I take this to indicate fine-grained phonetic distinguishability that does not translate into phonological feature specifications. This is also the conclusion of Rice (2002).

There are four interpretations for the fact that lip rounding can combine with tongue configuration to produce contrasts along the F2 continuum. First, vowel features are defined acoustically, rather than articulatorily. After all, one cannot claim that Figure 19.5 represents articulatorily defined vowel features having to do with tongue position (i.e. "front–back") and height, and then include labial characteristics. The second is to claim that [round] is not a vowel place feature, but some other type of feature. This leaves vowel place as defined solely via tongue position – this is the position taken by Chomsky and Halle (1968). The third is to suggest that vowel features are not defined purely on acoustic or articulatory characteristics, but different features are acoustically or articulatorily defined. This seems to be the position taken by most generative phonologists, even if they do not state so explicitly. Finally, one could claim that vowel place features are articulatorily defined via the location along the vocal tract where constrictions or expansions occur. This is essentially the position of Articulatory Phonology (Browman and Goldstein 1986, 1992; see also CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATIONS).

While it is fairly uncontroversial that the vowel space is divided roughly into six phonetic regions along the F2 dimension when one looks at cross-linguistic inventories, there is considerable debate about how these phonetic regions should be classified from a phonological perspective. Clements (1991) suggests that three distinct, unary, place features/nodes are needed to capture six contrastive vowel places – two features/nodes for tongue position along the front–back dimension (i.e. [coronal] and [dorsal])[4] and one feature/node for lip rounding (i.e. [labial]). This is a minor modification of *SPE*-based feature theory. There are, of course, a number of ways to implement this system, depending on one's assumptions about underspecification/full specification, feature co-occurrence markedness, etc. (CHAPTER 4: MARKEDNESS; CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION). (3a) and (3b) show two possible implementations for high vowels.[5]

(3)

a.	[labial]	[coronal]	[dorsal]	b.	[labial]	[coronal]	[dorsal]
[ɨ]				—			
[ʉ]	✓			—	✓		
[i]		✓		[i]		✓	
[ɯ]			✓	[ɯ]			✓
[y]	✓	✓		[y]	✓	✓	
[u]	✓		✓	[u]	✓		✓
—		✓	✓	[ɨ]		✓	✓
—	✓	✓	✓	[ʉ]	✓	✓	✓

As we see in (3a), the central unrounded vowel [ɨ] is placeless, three vowels ([ʉ i ɯ]) have one place feature specified, and two vowels ([u y]) have two places. Interestingly, factorial combination predicts two additional vowels that seem to

[4] These labels make use of two aspects of the phonetics. Descriptively, "coronal" refers to the passive articulator (i.e. the crown of the oral cavity) and "dorsal" refers to the active articulator (i.e. tongue dorsum). It is unclear what the advantage is of having this mixed set of labels.

[5] This is an illustration, and it is not meant as an endorsement of this over other approaches.

Thus the fact that Wolof and Nimi'ipuu non-RTR/RTR vowel pairs differ in both F1 and F2 does not automatically justify dismissing a pharyngeal place hypothesis. More evidence is needed. One might fruitfully compare the subtle F2 differences here with those between the labial and non-labial vowel pairs discussed in §4.1 (e.g. [i] and [y]). What makes labiality an obvious candidate for a phonological place distinction for vowels, but not pharyngeality?

Finally, there is evidence from some Arabic varieties that vowels can have contrastive pharyngeal place. For example, Cairene Arabic is usually described as having just three contrastive vowels, /i u a/; however, recent studies show that there are four contrastive vowels in this language – /i u a ɑ/ (Youssef 2006, 2007). The vowel [ɑ] is a contrastive "emphatic" vowel that triggers what is called emphasis spread and is also the result of spreading emphasis to /a/. If the descriptive term "emphasis" is interpreted as involving a phonological pharyngeal feature (as is commonly done), then this language has a contrastive pharyngeal and non-pharyngeal low vowel. As one might expect from looking at Wolof and Nimi'ipuu, the difference between pharyngeal and non-pharyngeal vowels in Cairene Arabic is realized acoustically by both F2 and F1 differences, where the pharyngeal vowel has a higher F1 and a lower F2.

To summarize, there are at least five vowel places from an articulatory perspective: labial, coronal, velar, pharyngeal, and retroflex. The first four primarily involve passive articulators, while the fifth primarily involves a particular active–passive articulator combination. Further, the acoustic correlates of the first three involve almost exclusively variation in F2, while the fourth involves both F1 and F2, and the fifth F3. There is currently disagreement in the literature on vowel place contrasts with respect to how many vowel place features are needed to establish cross-linguistically possible and language-particular phonological contrasts. This is in no small part due to the fact that different theories use different aspects of phonetics (and their relationship to phonology) to define phonological place features.

4.2 *Consonant place contrasts*

In addition to being used to account for place contrasts among vowels, vowel features have also been used to account for place contrasts among consonants (see also CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). This is traditionally done in two ways. Either some set of vowel features is used to determine what is called "secondary place of articulation," or they are used to determine primary place of articulation. We will begin with the former.

4.2.1 *Consonant secondary place*

It has long been observed that some languages make use of what is called secondary place (CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION; CHAPTER 71: PALATALIZATION). This may be described as a less extreme constriction of the vocal tract occurring roughly simultaneously with a more extreme constriction at a different point along the vocal tract. This less extreme closure is approximately that of a glide and is often attributed to the specification of a consonant with an extra vowel feature (e.g. Chomsky and Halle 1968; Clements 1985; Odden 1991). Examples of secondary articulation are labialization (e.g. [tʷ]), palatalization (e.g. [tʲ]), velarization (e.g. [tˠ]), and pharyngealization (e.g. [tˤ]). Some languages claimed to have single secondary place contrasts are listed in Table 19.2.

Table 19.2 Sample of languages described as having secondary place distinctions

	labialization	*palatalization*	*velarization*	*pharyngealization*
Chipewyan	✓			
Russian, Irish		✓		
Arabic varieties, Ponapean			✓	
Arabic varieties				✓

Table 19.3 Sample of languages described as having multiple secondary place distinctions

	labialization	*palatalization*	*velarization*	*pharyngealization*
Nambakaengo	✓	✓		
Nupe		✓	✓	
Marshallese	✓	✓	✓	
Salishan, Caucasian and Semitic varieties	✓			✓
Caucasian and Semitic varieties		✓		✓

It has also been observed that each secondary place can contrast with other secondary places within a single language (Table 19.3).

4.2.2 Consonant primary place

There is also a long tradition of using vowel features to define consonant primary place. For example, Chomsky and Halle (1968) defined all postalveolar consonant places via a combination of vowel place and height features. What is interesting regarding the SPE features [±high], [±low], and [±back] is that while the former two are used by Chomsky and Halle to express vowel height and the latter to express vowel place, the three join forces to express a variety of primary consonant places when combined with [−coronal, −anterior], as shown in (8) (Chomsky and Halle 1968: 305). This mix of different types of vowel features to define "consonant place" makes one wonder if it makes sense to discuss place as a phonologically relevant concept at all, or whether "place" is simply a convenient descriptive label that is ultimately epiphenomenal. Further, if place is a phonologically relevant concept, then what is its phonetic grounding, given that phonetically disparate characteristics are used to define place in different contexts?

(8)

	palatals	velars	uvulars	pharyngeals
[high]	+	+	−	−
[low]	−	−	−	+
[back]	−	+	+	+

4.4.1 Consonant effects on vowels

In some languages, consonant place has a direct effect on vowel place. In those cases where the relevant consonant and vowel place are phonetically similar, this may be interpreted as assimilation and a sharing of the same feature by two major classes of segments (CHAPTER 81: LOCAL ASSIMILATION).

The first case we will look at is "vowel fronting" in Serbian. This involves the fronting of back mid vowels in some suffixes when following postalveolar coronal consonants.

(12) [ɔ] → [ɛ] / [t͡ʃ̠ d͡ʒ̠ t͡ʃ̠ʷ d͡ʒ̠ʷ ʃ̠ ʒ ɲ ʎ j] + __

(13)	a.	*Neuter noun nominative accusative singular* [-ɔ] ~ [-ɛ]					
		sel-o	[s̪ɛlɔ]	'village' ~	*polj-e*	[pɔʎɛ]	'field'
	b.	*Masculine instrumental singular* [-ɔm] ~ [-ɛm]					
		grȃd-om	[graːd̪ɔm]	'city' ~	*mȗž-em*	[muːʒɛm]	'husband'
	c.	*Genitive singular* [-ɔg] ~ [-ɛg]					
		dobr-og	[d̪ɔbrɔg]	'good' ~	*loš-eg*	[lɔʃ̠ɛg]	'bad'
	d.	*Masculine plural* [-ɔv-i] ~ [-ɛv-i]					
		grad-ov-i	[grad̪ɔvi]	'city' ~	*múž-ev-i*	[muːʒɛvi]	'husband'

As pointed out by Morén (2006a), Serbian's quite rich set of coronal consonants combined with this assimilation process suggest that the triggers and targets of velar fronting share a vowel "coronal" place feature. The importance to the present discussion is that these data complement other facts of Serbian to show that palatal consonants in this language are specified with a "vowel" place feature and no "consonant" place feature.

Another case comes from well-known restrictions on vowel quality allowed between two labial consonants in Cantonese (Cheng 1991; Hume 1991). As Cheng shows, non-low vowels must be front (i.e. [i e y ø]) within a syllable containing a coronal onset and coda. One account for this is that the two consonants must share their consonantal coronal place within the syllable domain and intervening vowels must become coronal as a result of feature spread through them. Unlike Serbian, these facts suggest that the front vowels make use of a consonant place feature, not a vowel place feature.

(14)	[kʰut]	'bracket'	[kʰyt]	'decide'
	[ho]	'river'	[hø]	'boots'
	[tʰuk]	'bald head'	[kʰut]	'bracket'
	[tʰok]	'to support'	[kot]	'to cut'
	[tʰyt]	'to take off'	*[tʰut]	
	[tʰøn]	'a shield'	*[tʰon]	

Khoisan languages have a similar relationship between consonants and vowels in that only back vowels can surface following velar and uvular consonants (including clicks) (Traill 1985).

To summarize, the Serbian, Cantonese, and Khoisan patterns suggest that "vowel place" and "consonant place" are occasionally convenient as descriptive labels, but vowels seem to be able to have "consonant place" and consonants seem to be able to have "vowel place."

4.4.2 *Vowel effects on consonants*

In addition to consonant quality affecting vowels, vowel quality can also affect consonants (Clements 1991; Hume 1991). The most obvious case is that of vowels causing secondary place to appear on adjacent consonants. There are numerous examples of this in the literature, and they have the general pattern illustrated in (15).[9]

(15) a. /t+i/ → [tʲi] b. /p+i/ → [pʲi]
/t+o/ → [tʷo] /p+o/ → [pʷo]
/t+ɑ/ → [tˤɑ] /p+ɑ/ → [pˤɑ]

Another type of vowel-to-consonant place interaction is when vowel place affects primary consonant place. Serbian "velar palatalization" is an example of this. As discussed in Morén (2006a), velar palatalization is a process by which velar obstruents become apical postalveolar when followed by some suffixes beginning with a front vowel.

(16) [k x g] → [t̠͡ʃ̠ʷ ʃ̠ ʒ̠] / __ + [i ɛ]

(17) a.	*Masculine vocative singular* [-ɛ]					
	vojnīk	[vɔjniːk]	'soldier'	~	*vojnīč-e*	[vɔjniːt̠͡ʃ̠ʷɛ]
	bôg	[bɔːg]	'god'	~	*bož-e*	[bɔʒɛ]
	siromāh	[ʂirɔmaːx]	'poor'	~	*siromāš-e*	[ʂirɔmaːʃ̠ɛ]
b.	*Diminutive* [-it̠͡ʃ(-a)]					
	krûg	[kruːg]	'circle'	~	*kruž-ić*	[kruʒit̠͡ʃ]
	rûka	[ruːka]	'hand'	~	*ruč-ić-a*	[rut̠͡ʃ̠ʷit̠͡ʃa]
c.	*Denominal* [-it̠-i]					
	muka	[muka]	'pain'	~	*muč-it-i*	[mut̠͡ʃʷit̠i]
	sûh	[ʂuːx]	'dry'	~	*sûš-it-i*	[ʂuːʃ̠it̠i]

Unlike some other Slavic languages, Serbian consonants cannot have secondary coronal place, so the effect of "palatalization" of velars is one of primary place change, not secondary place addition. As was the case with Serbian mid vowel fronting discussed in §4.4.1, velar palatalization suggests that the trigger and the target share a phonological place feature. However, the difference is that mid vowel fronting involves a consonant feature affecting a vowel, while velar palatalization involves a vowel feature affecting a consonant.

A final example comes from consonant epenthesis. In some varieties of English, there is consonant resyllabification or epenthesis in certain environments (Kahn 1976). This is sometimes called glide formation (CHAPTER 15: GLIDES), epenthesis, or consonant "intrusion," depending on the context and quality of the consonant – [j w], [ʔ], or [ɹ], respectively. The details regarding the conditioning environments and subtle dialect differences are beyond the scope of this chapter, but what is important here is that this type of phenomenon is widely distributed across the English-speaking world (Kahn 1976; McCarthy 1993; Uffmann 2007; Krämer

[9] While palatalization and labialization are quite common cross-linguistically, pharyngealization in the context of a low back vowel is not. However, this is found in some Arabic varieties as a reflex of emphasis spread.

For example, Matumbi initial underlying vowels surface as vowels when followed by a consonant, but they surface as glides when followed by a vowel (Odden 1996).

(21)	/u+teliike/	[uteliike]	'you cooked'
	/i+taabua/	[itaabwa]	'books'
	/i+a+tuumbuka/	[jaatuumbuka]	'they will fall'
	/i+otu+i+k+e/	[jootwiike]	'they have holes'
	/u+a+teleke/	[waateleke]	'you should cook'
	/i+ula/	[juula]	'frogs'

In feature theories where vowels and glides are the same segment in different syllabic positions, nothing need be said about place feature specification. However, in feature theories where vowels and glides are featurally distinct, then something must be said about their surface place similarities – especially if those theories postulate different place features for consonants and vowels.

Less obvious vowel–consonant alternations involve liquids. It is common among the Slavic languages to have alternations between a lateral and a vowel. For example, Serbian syllable-final laterals surface as mid back round vowels (Morén 2006a), while in Slovenian, they surface as a high back round vowel/glide (Morén and Jurgec 2007). The former is shown in (22).

(22)	*Serbian*				
	pev-ao	/pɛval/	→	[ˈpɛ.vaɔ]	'sang'
	orao	/ɔral/	→	[ˈɔ.raɔ]	'eagle'
	posao	/pɔs̪al/	→	[ˈpɔ.s̪aɔ]	'work'
	mišao	/miːs̪al/	→	]ˈmiː.s̪aɔ]	'thought'
	čitao-ca	/t͡ʃʷitalt͡s̪a/	→	[ˈt͡ʃʷi.taɔ.t͡s̪a]	'reader'

Ignoring the issue of how to establish the correct vowel height during vocalization, the unrounded anterior coronal lateral is clearly related to a surface vowel that is both back and round. This place relationship is not straightforwardly accounted for in some theories.

While one might argue that the Slavic data show alternations of underlying consonants with surface vowels, Yoruba seems to show the opposite. According to Akinlabi (2007), Yoruba [i] surfaces as [ŋ] following a vowel.

(23) /àwɔ́ ı wɔ́ tã́/ → [àwɔ́ ŋ̀ wɔ tã́]
flowing NEG flow finish
'that which flows (of gown) without end'

He interprets this as fortition, in which the high front vowel becomes a syllabic nasal so as not to require (perhaps not license) an onset. What is relevant to us is that a front vowel alternates with a velar nasal. Given that most feature theories assume that coronal is the default/unmarked place, it is not a trivial matter to explain why a coronal vowel alternates with a velar consonant.

Finally, vowels can also alternate with consonant secondary articulations.[13] As noted by Clements (1991), Etsako high vowels appear as secondary place on

[13] Similar alternations occur in Luganda (Clements 1991) and Matumbi (Odden 1996).

because they involve significant non-F2 effects. Closer inspection of physical and phonological parallels between vowels and consonants suggests that the traditional view needs to be reconsidered.

From a phonological perspective, vowel place features are relevant to seven major types of phenomena. I have discussed five of these, and each both supports and challenges some of the more common approaches to vowel place found in the literature. Combined, they suggest that we have a fairly good understanding of relevant phenomena and formalisms, even if we sometimes disagree about the details of data interpretation or which feature theory is best.

Finally, this chapter began by posing the question: "What is vowel place and what role does it play in the sound patterns of language?" To begin answering this question, we explored the phonetics and phonology of "place" more generally, as well as examined why some scholars make different choices regarding the relationship between phonetics and phonology and between vowels and consonants. We saw that vowel place is both simpler and more complicated than one might expect, and that more work is needed to resolve some long-standing and fundamental issues.

REFERENCES

Akinlabi, Akinbiyi. 2007. Category change as vowel reduction. Paper presented at the 30th Generative Linguistics in the Old World Conference, University of Tromsø.

Anderson, John M. & Colin J. Ewen. 1987. *Principles of dependency phonology*. Cambridge: Cambridge University Press.

Aoki, Haruo. 1966. Nez Percé vowel harmony and proto-Sahaptian vowels. *Language* 42. 759–767.

Backley, Phillip. 1997. Monovalency and the status of RTR. *UCL Working Papers in Linguistics* 9. 1–26.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180.

Cheng, Lisa L. 1991. Feature geometry of vowels and co-occurrence restrictions in Cantonese. *Proceedings of the West Coast Conference on Formal Linguistics* 9. 107–124.

Chiba, Tsutomu & Masato Kajiyama. 1942. *The vowel: Its nature and structure*. Tokyo: Kaiseikan.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. 1991. Place of articulation in consonants and vowels: A unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5. 77–123.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

Davis, Stuart. 1995. Emphasis spread in Arabic and Grounded Phonology. *Linguistic Inquiry* 26. 465–498.

Delattre, Pierre, Mark Liberman & Franklin S. Cooper. 1955. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America* 27. 769–773.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.

Harris, John & Geoff Lindsey. 1994. Phonological interpretations of historical lengthening. In Wolfgang U. Dressler, Martin Prinzhorn & John R. Rennison (eds.) *Phonologica 1992*, 96–106. Turin: Rosenberg & Selier.

Harris, John & Geoff Lindsey. 1995. The elements of phonological representation. In Jacques Durand & Francis Katamba (eds.) *Frontiers of phonology: Atoms, structures, derivations*, 34–79. London & New York: Longman.

Heegård, Jan & Ida Elisabeth Mørch. 2004. Retroflex vowels and other peculiarities in Kalasha sound system. In Anju Saxena (ed.) *Himalayan languages: Past and present*, 57–76. Berlin & New York: Mouton de Gruyter.

Hulst, Harry van der & Jeroen van de Weijer. 1995. Vowel harmony. In Goldsmith (1995), 495–534.

Hume, Elizabeth. 1991. Vowel-to-consonant assimilation and its implications for feature theory. *Proceedings of the West Coast Conference on Formal Linguistics* 20. 138–151.

Hume, Elizabeth. 1994. *Front vowels, coronal consonants and their interaction in nonlinear phonology*. New York: Garland.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Kahn, Daniel. 1976. Syllable-based generalizations in English phonology. Ph.D. dissertation, MIT.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1985. The internal structure of phonological elements: A theory of charm and government. *Phonology Yearbook* 2. 305–328.

Krämer, Martin. 2008. Taking a free ride can cau[r]se severe hyperrhoticity. Paper presented at the 16th Manchester Phonology Meeting.

Lass, Roger. 1984. Vowel system universals and typology: Prologue to theory. *Phonology Yearbook* 1. 75–111.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Mahanta, Shakuntala. 2008. Local vs. non-local consonantal intervention in vowel harmony. In Sylvia Blaho, Camelia Constantinescu & Erik Schoorlemmer (eds.) *Proceedings of ConSOLE XV*, 165–188. Leiden: University of Leiden.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John J. 1993. A case of surface constraint violation. *Canadian Journal of Linguistics* 38. 169–195.

Morén, Bruce. 1999. Distinctiveness, coercion and sonority: A unified theory of weight. Ph.D. dissertation, University of Maryland at College Park.

Morén, Bruce. 2006a. Consonant–vowel interactions in Serbian: Features, representations and constraint interactions. *Lingua* 116. 1198–1244.

Morén, Bruce. 2006b. "ATR" harmony in Kàlòng and Nez Perce without [±ATR]: Abstraction, economy and primitives. Paper presented at the 3rd Old World Conference in Phonology (OCP3), Budapest.

Morén, Bruce & Peter Jurgec. 2007. Consonants and vowels in Slovenian and Serbian: Phonetic similarities, (morpho)phonological differences and vice versa. Paper presented at the 7th European Conference on the Formal Description of Slavic Languages (FDSL-7), Leipzig.

Odden, David. 1991. Vowel geometry. *Phonology* 8. 261–289.

Odden, David. 1996. *The phonology and morphology of Kimatuumbi*. Oxford: Clarendon Press.

Rice, Keren. 2002. Vowel place contrasts. In Mengistu Amberber & Peter Collins (eds.) *Language universals and variation*, 239–270. Westport, CT: Praeger.

Rosenthall, Sam. 1994. Vowel/glide alternation in a theory of constraint interaction. Ph.D. dissertation, University of Massachusetts, Amherst.

Sagey, Elizabeth. 1990. *The representation of features in nonlinear phonology: The articulator node hierarchy*. New York: Garland.

Schane, Sanford A. 1984. The fundamentals of particle phonology. *Phonology Yearbook* 1. 129–155.

Schein, Barry & Donca Steriade. 1986. On geminates. *Linguistic Inquiry* 17. 691–744.

20 The Representation of Vowel Length

David Odden

1 Introduction

Length is the phonological correlate of durational differences between sounds, tied to the phonological concept "quantity." The concept of length usually considers duration segmentally, attributing length to particular segments (the vowel or consonant is the locus of the property "long" or "short," and not the syllable, foot, or word). Length being a phonological attribute, distinctions are discrete mental categories, not physical measurement, and, like most phonological attributes in generative theories, length is traditionally treated binarily.

Establishing vowel length in a language is not always straightforward. English is said both to have and not have vowel length, so that *peat* may be analyzed as having a long vowel (Halle 1977) as contrasted to *pit*, thus [iː] *vs.* [i], or the distinction can be treated as a tense/lax opposition (Chomsky and Halle 1968), thus [i] *vs.* [ɪ]. Because of the influence of voicing on vowels – they have shorter duration before a voiceless obstruent – English has a cline of four durational patterns, illustrated in *bead, beat, bid, bit*. Phonetic duration does not automatically translate into phonological length, and durational differences, especially ternary oppositions, are not necessarily of a continuous type (1, 2, 3, 4 units), but may instead reflect multiple, intersecting phonological phenomena.

Some languages distinguish "full" and "reduced" vowels. Chuvash (Krueger 1961) has two reduced vowels [ă ĕ], which contrast with full vowels [i y ɨ u a e o]; the two Proto-Slavic "yers" [ĭ ŭ] (see CHAPTER 122: SLAVIC YERS) are widely claimed to be "reduced," and a number of Ethiopian Semitic languages have multiple reduced vowels (Pam 1973). Rather than introducing a new orthogonal vowel property to the inventory of features, one which seems to never combine with independent length, the Chuvash vowels could be treated as short /a e/ *vs.* long /iː yː ɨː uː aː eː oː/. This would have the typologically peculiar consequence that all high vowels are long. However, in support of a prosodic distinction between these vowels, stress assignment is sensitive to the full/reduced distinction whereby reduced vowels are skipped over in favor of stressing full vowels.

Segmental influences on duration, discussed in Lehiste (1970), include tongue height, place of articulation, and stop/fricative properties of the following consonant, as well as pitch (especially under contrastive pitch movements), stress, and the fact of a syllable being open or closed. It can be unclear whether a perceptible durational pattern should be attributed to category change, so should we claim that, in the phonology, /bit/ *beat* → [bit] and /bid/ *bead* → [biːd] (similarly /bɪt/ *bit* → [bɪt], /bɪd/ *bid* → [bɪːd])? Or should we leave pre-voiceless durational adjustments to phonetics and have the output of the phonology be [bit], [bid], [bɪt], and [bɪd]? Answering these questions involves foundational issues in the theory of phonological and phonetic grammars that are well beyond the scope of this chapter. Unfortunately, many issues pertaining to the phonology of vowel length require understanding this relationship. If we cannot clearly identify "long vowels" contrasted with "short vowels," it will be difficult to develop a theory of how long *vs.* short vowels are represented.

A theoretical puzzle confronting linear (*Sound pattern of English*-type) theories of representation was how to resolve numerous contradictions in phonological length. The paradox is that long vowels sometimes behave as though they are single segments distinguished from short vowels by a feature, and sometimes they behave as though they are two segments. The resolution of this contradiction played an important role in the development of autosegmental phonology. The goal of this chapter is to familiarize the reader with the basic facts of vowel length and to attempt to explain them. §2 considers the fundamental problems pertaining to the featural treatment of vowel length. §3 shows how autosegmental representations resolve those problems, but also raise new questions as we attempt to make more precise the nature of the higher-level representational atoms that express length. Facts such as compensatory lengthening which seem to elucidate vowel length are often cited as evidence to support an autosegmental representation of length, but many autosegmental treatments of compensatory lengthening gloss over important details, a fact which undermines the explanatory value of the autosegmental account. On balance, it seems that the autosegmental account of length does broadly resolve the fundamental one-segment/two-segment paradox, but important details remain to be ironed out. Finally, §4 considers the claim that some languages have a three-way vowel length contrast. Such (rare) languages are potentially important for a theory of length; unfortunately, for the vast majority of such languages, so little is known that it is difficult to say that length ternarity is a real phenomenon, rather than an epiphenomenon arising from binary length and some orthogonal dimension.[1]

[1] A reviewer notes a number of substantive asymmetries in vowel length phenomena, some of which have at least implicitly influenced theorizing on length. For example, length is often related to syllable structure, such that long vowels appear in open syllables and short vowels appear in closed syllables, and there are rules of open syllable lengthening and closed syllable shortening. Furthermore, closed syllable lengthening or open syllable shortening appear to be unattested. Vowels may lengthen before voiced consonants, but they do not shorten before voiced or lengthen before voiceless. The latter fact is not widely attested in languages and appears not to have been elevated to the status of fact that it is agreed that phonological theories must explain, whereas the former fact, especially the widely attested prohibition against syllables of the type [XVVC_0], has had an impact on theorizing about vowel length, in that we have sought to reduce the generalization *[XVVC_0] to a general representational limit. The purpose of this chapter is to investigate the representation of length, not all facts relevant to length. It is just outside the scope of this chapter to investigate why no language has been attested with closed syllable lengthening, yet such questions are certainly germane to a widespread concern of representational theorizing, namely saying why certain kinds of rules exist, yet others are not found.

2 Segmental theories of representation

The classical *SPE* feature-theoretic account of vowel length is that long vowels are [+long] and short ones are [−long], thus rules affecting vowel length manipulate a feature on a par with nasality, rounding, or backness – for which a segmental transcription such as [a:] or [ā] would be appropriate. It was also recognized in works such as McCawley (1968) that long vowels could be represented as sequences of identical vowels, thus [aa].

Lithuanian evidence for treating long vowels and diphthongs identically is discussed in Kenstowicz (1970), which supports a representation of long vowels as adjacent identical short vowels. One argument is that rising and falling tones only appear on long vowels and diphthongs or else in syllables ending with a liquid or nasal, i.e. bisegmental vowel plus sonorant sequences. There are words like /kâimas/ 'village', /váikas/ 'child', with diphthongs, /kârtis/ 'pole', /kártis/ 'bitterness', with V + resonant sequences, and /matî:ti/ 'to see', /matǐ:s/ 'he will see', with long vowels. There are no level-H toned syllables with a diphthong, vowel plus consonant, or long vowel. In contrast, only level H or L (unmarked) appear on short vowels, as in /mésti/ 'to throw', /dúri:s/ 'door'. This pattern is explained if the language allows just one H on one segment within the word, where coda sonorants can bear tone, and crucially, long vowels are really two-segment sequences, thus /matíiti/, /matiís/, /káimas/, /vaíkas/, /kártis/, /kar̃tis/. If long vowels simply have the feature [+long], explaining the distribution of H tone becomes very complex.

Another argument for equating long vowels with vowel sequences is an accent shift before endings with an underlying initial H, which is described very simply as deleting the first of two adjacent H tones. This description only works if long vowels are represented as VV, and becomes more complex if stated in terms of monosegmental long vowels and a primitive rising/falling/level distinction. An H underlyingly on the last tone-bearing unit of the stem (retained before the toneless accusative singular suffix *-a:*) seems to shift to the underlyingly H-initial instrumental singular suffix, as illustrated by the alternation [pieʃtúk-a: ~ pieʃtuk-ú] 'pencil (ACC SG ~ INSTR SG)', the latter from /pieʃtúk-úo/.

(1)	*acc sg*	*instr sg*			
	pieʃtúk-a:	pieʃtuk-ú	←	/pieʃtúk-úo/	'pencil'
	mígl-a:	migl-ú	←	/mígl-úo/	'mist'
	vaík-a:	vaik-ú	←	/vaík-úo/	'child'
	pǒ:n-a:	po:n-ú	←	/poón-úo/	'master'
	kâim-a:	kâim-u	←	/káim-úo/	'village'
	vê:j-a:	vê:j-u	←	/véej-úo/	'wind'

In the final two forms in (1), the suffix H deletes for reasons not of relevance here.

Retranscribing long vowels as sequences and decomposing contour tones into H on V_1 *vs.* V_2 makes the nature of the alternation clearer: H deletes before adjacent H.

There are shortening rules in Lithuanian which treat long vowels and diphthongs analogously. One rule shortens a word-final falling-toned long vowel or diphthong, which accounts for alternations in adjectives in their indefinite and definite forms.

(5)	*'you should V'*	*'you should go V'*	
	ulubé	ukalubé	'ask'
	upakatike	ukapakátike	'shake down'
	ukatʃabánike	ukakatʃábanike	'be confused'
	utʃaŋgaále	ukatʃaŋgáale	'wonder'
	upapaákije	ukapapáakije	'grope'
	ulaambáte	ukalaámbate	'lick'
	ukeeŋgéembe	ukakeéŋgeembe	'dig up'

Similar arguments show the equivalence of long vowels and vowel sequences in Kuria tone assignment (Odden 1987). In short, there is ample evidence that long vowels act like two vowels.

Work by Pyle (1971), Kenstowicz (1970), and Fidelholtz (1971) also noted evidence for treating vowel length with a feature. Southern Paiute (6a) and Tübatulabal (6b) were often cited as illustrating different treatments of length according to language. In Southern Paiute, devoicing only affects one half of a long vowel (arguing for a sequence), but in Tübatulabal, reduplication of the first V copies vowel length (requiring a featural representation).

(6) a. mantʃáạqàạ 'to hold out one's hand'
maróọqwàj'ɨqwạ '(I) stretch it'

b. *telic* *atelic*
toːjlaːn oː-doːjlaːn 'to teach him'
togoːjˀan o-togoːjˀan 'to decoy it for him'

Sometimes long vowels act like both a single segment and a sequence in one language. Kenstowicz motivates a rule in Lithuanian rounding /aː/ to [oː], affecting the output of a rule lengthening vowels in certain verbs in the non-present. When the vowel is underlying short /a/, the surface result is [oː].

(7) tupʲú tuúpti 'perch'
drebʲú dreébti 'splash'
vagʲú voógti 'steal'

The problem is that the concept "long *a*" is difficult to capture in the sequence theory, since that theory says that length is not a feature that can be referred to. One would have to refer to "[a] which is preceded or followed by another [a]." Since arguments for the sequence theory succeeded because of the resulting simplifications, these complications are important.

Pyle (1971) shows that in West Greenlandic, lowering of high vowels before uvulars affects long and short vowels equally, thus /aluq/ → [aloq] 'sole', /puːq/ → [poːq] 'bag'. If long vowels were vowel clusters, we predict /puuq/ → *[puoq] – note also that vowel lowering does not skip a vowel, thus /ukiuq/ → [ukioq] 'year', not *[ukeoq]. Contrarily, though, a metathesis rule breaking up final consonant clusters ([meːraq]: /meːraq-t/ → [meːrqa-t] 'child (SG, PL)') does not apply when the cluster is preceded by a vowel sequence ([umiaq]: /umiaq-t/ → [umiat] 'canoe (SG, PL)'), or by a long vowel ([ikaːq]: /ikaːq-t/ → [ikaːt] 'scaffold (SG, PL)'). Thus, like Lithuanian, West Greenlandic supports both modes of representation.

between singulars and plurals, including lengthening, shortening, diphthongization, tone changes, vowel quality changes, consonant changes, and suppletion.[3]

(10)	*sg*	*pl*		*sg*	*pl*	
	gɔ́l	gɔ́ɔ́l	'clan'	dwɛ̤́l	dwɛ̤̀l	'temporary house'
	bó̤ó̤k	bó̤k	'animal skin'	jic	jì̤ì̤t̤	'ear'
	d̤ɔ̤k	da̤k	'boy'	ɲáŋ	ŋɛ̤ŋ	'crocodile'
	kwir	kwer	'road'	cìn	cì̤ì̤n	'intestine'
	dit	diɛt	'bird'	kèw	kèèt̪	'gazelle'
	ruon	run	'year'	kow	kuoot̪	'thorn'
	lwet̪	lwét̪	'lie'	kat	kɛ̤́ɛ̤́t	'shelter'
	bél	bel	'sugar cane'	diir	dɪ̤́r	'cricket'
	cók	có̤k	'foot'	wə́ŋ	ɣɔ́k	'cow'
	kàl	ka̤l	'fence'	tik	djà̤r	'woman'

Ladd *et al.* (2009) argue that there is some predictability, but nevertheless there simply is no length-switching rule, thus no such argument for treating length as a feature.

Another counterexample to the function/representation hypothesis is Tübatulabal reduplication (6b), which is not a segment quality rule but rather a prosodic rule, yet one that copies the length of the first root vowel in the same way that it copies vowel quality features. Any rule of prevocalic vowel shortening (as exists in Kamba) which specifically only applies to long vowels and not vowel sequences is a problem for the rule-functional claim. The reason is that a rule that lengthens or shortens a vowel is by definition a prosodic rule, and prosodic rules are hypothesized to always treat long vowels in the same manner as vowel sequences. In Kamba, long vowels shorten before another vowel, as illustrated below with the remote future prefix /-kâ:-/ (/-káa-/).

(11)		
	to-kâ:-tálâ	'we will count'
	to-kâ:-konà	'we will hit'
	to-ká-o-à	'we will buy'
	to-ká-i-o-à	'we will buy them'
	to-ká-.a.à	'we will divide'

As discussed in §4 below, Kamba contrasts identical-vowel sequences with long vowels. Forms like [tokáioà] demonstrate that the prevocalic shortening seen in the mapping /tokáa-o-a/ → [tokáoà] is not the result of a general vowel-cluster reduction rule. Only long vowels undergo shortening, thus long vowels are, contrary to hypothesis, not treated the same as vowel sequences.

A severe problem for a two-segment theory is posed by Gokana, discussed in Hyman (1982). A rule of postvocalic shortening shortens the 3rd singular object suffix, /EE/, seen in [bāē dìv-ēē] 'they hit him', whereas /bāè sà-EE/ becomes [bāē sà-ɛ̄] 'they chose him'. This poses the same problem as Kamba prevocalic shortening. In addition, Gokana inserts [r] between long vowels after the logophoric suffix, /EE/, seen in [bāè dīv-èè] '(they$_i$ said) they$_i$ hit (it)', whereas

[3] These examples were provided to me by the late Keith Denning, based on his fieldwork.

/bāè sīī-EE/ becomes [bāè sīī-rèè] '(they$_i$ said) they$_i$ caught (it)', but /bāè sā-EE/ surfaces as [bāè sā-ɛ̀] '(they$_i$ said) they$_i$ chose (it)' without r-epenthesis (and with postvocalic shortening), because insertion requires a sequence of long vowels.[4] When the logophoric suffix is preceded by a sequence of adjacent identical vowels – a long vowel, according to the vowel-sequence theory – there is no insertion, as shown by /bāè kɛ̃̄ɛ̃̄-ɛ̃̄-EE/ → [bāè kɛ̃̄ɛ̃̄-ɛ̃̄-ɛ̃̄] '(they$_i$ said) they$_i$ woke (it) up'. Gokana thus phonologically distinguishes true long vowels from adjacent identical-vowel sequences, which is not predicted to be possible by the sequence theory.

The theory that segmental rules always treat long vowels as monosegmental is also falsified by the Mackenzie dialect of Eskimo, noted by Pyle, which also lowers high vowels before uvulars but where /naluuq/ → [naluoq] 'throws out'. Various segmental rules of Southern Paiute treat long vowels as sequences, and can affect just one part of a long vowel, for example final devoicing noted in (6a), as well as a rule whereby /ɨ/ → [i] before [i] (which may later delete), where /ũnikʲaŋumɨɨts·i/ → [ũnikʲaŋumɨits·] 'after they had done so'.

Matumbi presents a worst-case scenario for segmental theories of length which attempt to attribute sequence-like *vs.* feature-like properties of length either to ordering or rule type. As noted above, the language gives strong evidence for the sequence theory, from tone assignment, which counts a long vowel as equivalent to two short vowels. The data of (5) omit one important type of stem, namely CVVCV, where the H would be word-final. In such cases, the predicted H tone is retracted to the penult, being realized as a rising tone.

(12)	u-kaáte	←	/u-kaaté/	'you should cut'
	u-toóle	←	/u-toolè/	'you should take'

To even coherently represent rising tone while explaining the robust constraint that contour tones only exist on long vowels, long vowels must be represented as bisegmental. Yet this tone-retraction rule must refer to the fact that the penult contains a long vowel, arguing for a feature representation of length. The rule only applies after long vowels and not adjacent VV sequences, as shown by [utaunè] 'you should chew', without retraction (cf. [ukataúne] 'you should go chew', showing that *u* can bear H). One cannot even impose an identity condition on the triggering penultimate vowel (i.e. "when preceded by a sequence of identical vowels"), thanks to the fact that the language lexically contrasts identical vowel sequences from long vowels, the former not triggering tone retraction ([baasá] ← /ba.asá/ 'envelope').

Matumbi has a rule which shortens long vowels in a phrasal head followed by a modifier. Vowel sequences do not simplify.

(13)	a.	nika-kálaaŋga	'I will go fry'
		nika-kálaŋga lí	'I will not go fry'
		nika-kálaŋga kinɟáambú	'I will go fry cassava leaves'
		nika-kálaŋga jóopáta eéla	'I will go fry to get money'
	b.	nika-táuna	'I will chew'
		nika-táuna lí	'I will not chew'

[4] Insertion only affects the logophoric suffix and the 2nd plural subject suffix /II/.

Assuming an ordered identical-vowel merger rule, such a rule would have to apply before shortening, so that long vowels could be created and thus referenced by a rule mentioning [+long] in the input and [−long] in the output. This is not ultimately tenable.

Tone assignment in the subjunctive occurs after shortening has applied, as the data below show – H is on the third surface vowel after the subject prefix, regardless of underlying length.

(14)	uʧaŋgaále	'you should wonder'
	uʧaŋgalé lí	'you should not wonder'
	upapaákije	'you should grope'
	upapakíje mundó	'you should grope in the bucket'
	ulaambáte	'you should lick'
	ulambaté mboópo	'you should lick the knife'
	ukeeŋgéembe	'you should dig up'
	ukeŋgembé kindoólo	'you should dig up sweet potatoes'
	ubeénde	'you should shout'
	ubendé ukumú	'you should shout at Ukumu'

In an ordered vowel-merging solution, vowel sequences must merge prior to shortening (to be targeted by shortening), and shortening precedes subjunctive tone assignment (which precedes retraction). But subjunctive tone assignment requires a sequence representation of long vowels, and is flanked by rules requiring a featural representation of vowel length.

Finally, Matumbi contrasts long vowels and adjacent identical vowel sequences, as in (15a), with true long vowels (which condition tone retraction and undergo shortening) *vs.* those of (15b), with sequences which do not trigger retraction or shortening.

(15)	a.	ntuúmbwi	← /ntuumbwi/	'canoe'	ntumbwí waáŋgu	'my canoe'
		ŋgaási	← /ŋgaasí/	'ladder'	ŋgasí jaáŋgu	'my ladder'
		ŋkaáte	← /ŋkaaté/	'bread'	ŋkaté ŋaáŋgu	'my bread'
	b.	baará	'ocean'		baará jaáŋgu	'my ocean'
		baasá	'envelope'		baará jaáŋgu	'my envelope'
		luusá	'permission'		luusá lwaáŋgu	'my permission'

While the insight that long vowels are in some sense like two vowels is correct, there are many reasons not to treat long vowels as literally a sequence of segments. Further developments in representational theory resolve the conflict in these insights.

3 Non-linear representations

When facts of this type were considered from the perspective of autosegmental and metrical phonology in the mid-1970s, it was obvious that, once the presumption of a one-to-one mapping between features and segments is abandoned, non-linear representations provide exactly the flexibility required to resolve these paradoxes. Just as a contour tone is represented as a many-to-one mapping between tones and vowels, a long vowel can be represented as a many-to-one mapping between something and a vowel. The crucial question is, what is that something which long vowels have two of?

3.1 Skeletal accounts

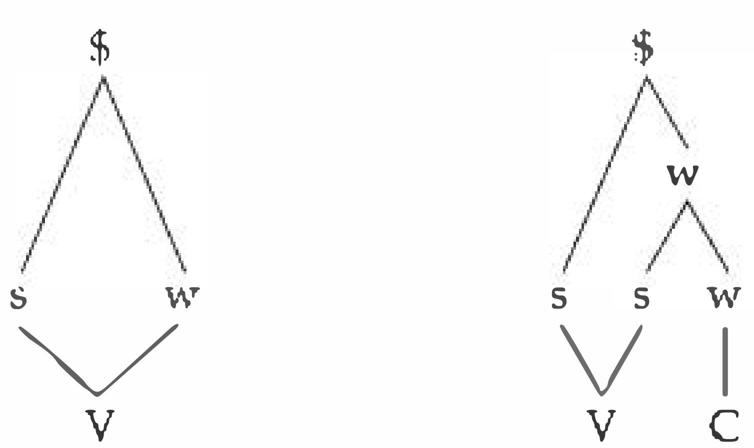
In McCarthy (1979: 34), long vowels were still seen as two-segment sequences, but with the imposition of higher structure including a rhyme. "Long vowel as unit" behavior, such as uniform lowering of long vowels before uvulars in West Greenlandic, is handled by applying the rule at the level of the rhyme. Rules may thus specify whether they apply at the segment level or higher, which yields an ambiguous treatment of long vowels. Long vowels as in Tiberian Hebrew [mala:ki:m] are represented as in (16), where a long vowel is two things (two segments) and one thing (one rhyme) at the same time.

(16)

Vowel lengthening is treated as node-adjunction within the rhyme, where the segment is filled in by automatic feature copying.

Ingria (1980) represents a long vowel as a single segment having two metrical positions.

(17) V: in an open syllable V: in a closed syllable



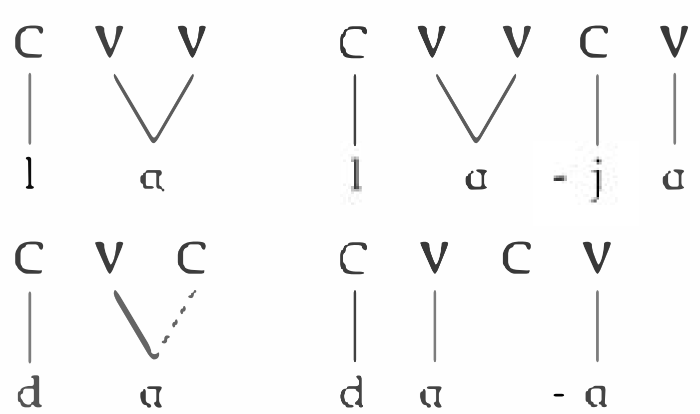
Since the focus of that article is compensatory lengthening arising from consonant deletions, little attention is paid to how underlying length is represented. It is simply noted (1980: 469) that "Each long vowel and consonant must contain the information that it is to be associated with two positions in the syllabic tree, whereas short vowels and consonants will be associated with only one position."

The earliest, most successful non-linear theory of length is CV theory, which is rooted in McCarthy's (1979) prosodic template proposal for Semitic morphology (see CHAPTER 108: SEMITIC TEMPLATES and CHAPTER 105: TIER SEGREGATION). McCarthy made a connection between vowels and V and between consonants and C (1979: 247), such that [syllabic] and [consonantal] are represented on the template tier rather than the segmental tier. McCarthy (1982) drops the feature [syllabic] but reintroduces the *SPE* feature [vocalic] to drive mapping of segments to the CV tier. The Clements and Keyser (1983) version of the theory fully phonologizes the CV tier as a strictly suprasegmental tier, containing only the formal elements C, V, which are independent of morphology, and dispenses with [syllabic]. A segment dominated by V is interpreted as a syllable peak, and a segment dominated only by C is interpreted as a non-peak.[5] The C/V distinction subsumes the functions of [syllabic], but is not identical to it. The feature [long] is also abandoned (1983: 12), and "long vowels are universally represented by means of the multi-attachment of a single vowel matrix to two positions on the CV-tier." The appeal of the theory is that it resolves the longstanding problem of the ambiguity of length, simplifies feature theory, and explains a problematic connection between syllabicity and compensatory lengthening, discussed below.

By reducing length to simple multiple-linking of prosodic atoms to segments, CV theory predicts the possibility of a further distinction in types of long vowels: a long vowel dominated by VC *vs.* one dominated by VV. This actually arises in Turkish, which contrasts two kinds of superficially identical long vowels.

(18)	*nom sg*	*pl*	*dat*	*poss*	
	lɑ:	lɑ:-lɑr	lɑ:-jɑ	lɑ:-sɯ	'la (musical note)'
	dɑ:	dɑ:-lɑr	dɑ-ɑ	dɑ-ɯ	'mountain'
	sɑp	sɑp-lɑr	sɑp-ɑ	sɑp-ɯ	'stalk'

The noun /lɑ:/ has a regular VV long vowel, and by regular principles of Turkish, the dative and possessive have the allomorphs /-jɑ/, /-sɯ/ after that long vowel, and /dɑ:/ has a long vowel arising from association to /ɑ/ of an empty C position which is part of the root.

(19)


The presence of the empty C triggers the suffix variant /-ɑ/, and the empty C syllabifies as the onset of the second syllable.

The theory draws substantial support from the fact that it gives a simple account of the difference – one never squarely dealt with in linear theorizing – between disegmental disyllabic [a.a] and phonetically distinct monosegmental monosyllabic [a:], as exists in Kamba. In principle, a monosyllabic long vowel dominated by VC could also behave differently from one dominated by VV, as is the case in Turkish, but the phonetic interpretation of the two long vowels would be identical, since "syllabicity" requires *some* V node to dominate the segment, not that *every* node dominating the segment must be a V.

[5] It is argued that certain long vowels in Klamath and Turkish are dominated by the sequence VC (see below), which is in line with the interpretation that any linkage to V causes a syllable-peak interpretation of the segment. However, a post-glide epenthesis rule is also posited (Clements and Keyser 1983: 135) in Klamath, whereby C-dominated /w/ becomes CV-dominated, and /wq'as/ → [wuq'əs] 'quartz'. It seems that only VC in the syllable rhyme is interpreted as a uniform syllable peak.

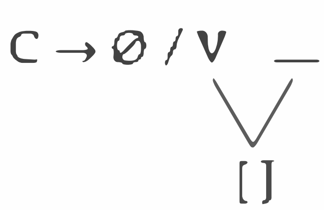
One significant argument for the CV theory of vowel length is its explanation for a Klamath puzzle noted by Kisseberth (1973), a rule shortening long vowels after a heavy syllable under certain conditions. The rule only affects vowels that arise from vocalizing a glide between consonants, not underlying long vowels or long vowels from vowel–glide sequences. This process applying to vocalized glides is seen in (20a), and not in (20b) with other long vowels.

(20)	a.	mbotj'-a	'wrinkle'	/mbodj'-a/
		mbodiː-tk	'wrinkle up'	/mbodj'-dk/
		mbo-mpti-tk	'wrinkled up (DISTR)'	/mbo-mbodj'-dk/
	b.	jə-jdiːs	'spirit stones (DISTR)'	/ja-jadiː-s/
		sdə-sdiːn-k'a	'little heart (DISTR)'	/sda-sdajn-k'a/

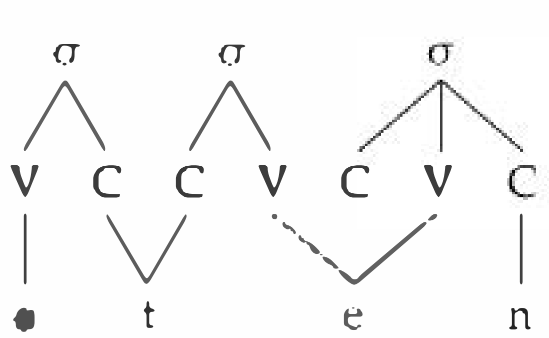
With only one representation of "long vowel," it would be impossible to distinguish shortenable [iː] derived from /CjC/ *vs.* non-shortening underlying /iː/. CV theory explains this representationally, where an underlying long vowel is dominated by VV, but an underlying glide between consonants will, due to a separate rule of pre-glide epenthesis, be represented as VC.

(21)


Clements and Keyser achieve the restriction on shortening in the formulation of the rule Vowel Shortening (1983: 171).

(22)


Parallel facts from Finnish are discussed in Clements and Keyser (1983) and Keyser and Kiparsky (1984), namely the anomalous treatment of certain long vowels with respect to gradation, vowel length alternations, and consonant gemination, which point to another representation of long vowels, as a monosegmental disyllabic VCV span, illustrated in the word [otteːn] 'grip (GEN SG)'.[6]

(23)


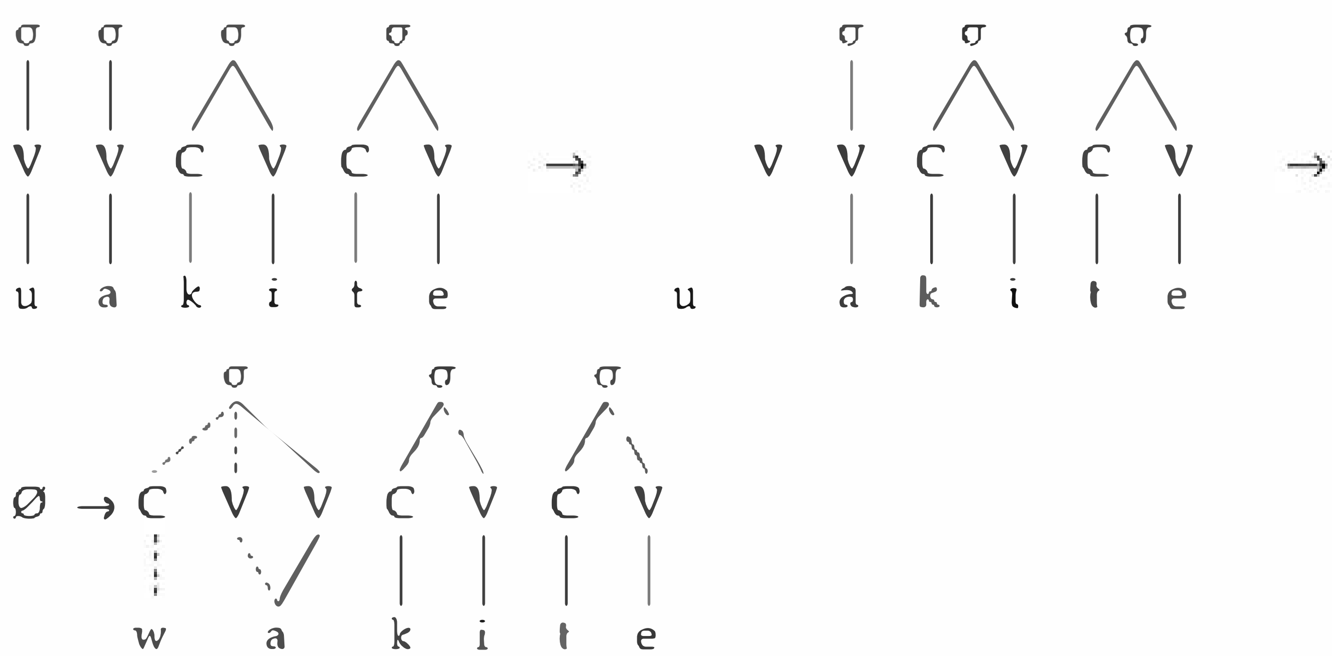
While the extra power of being able to distinguish VV and VC long vowels as well as monosyllabic and disyllabic versions of long vowels might seem to be a

[6] This type of representation is how Strict CV Phonology (Lowenstamm 1996) treats all long vowels.

A problem with the autosegmental theory of CL is the need to posit ad hoc conventions for glide formation in an onsetless syllable. The initial impact of this problem for the theory was reduced by the fact that Ganda does not have CL in such a context. In Ganda (and a set of closely related Bantu languages: see e.g. Odden 1995), there is no CL when an onsetless vowel desyllabifies, thus Ganda /o-a-gula/ → [wagula] 'you bought'; contrast /tu-a-gula/ → [twa:gula] 'we bought'. Clements (1986: fn. 1) treats this by a relabeling rule which directly changes word-initial V into C. Therefore, CL is not always a consequence of GF, even in languages like Ganda with vowel length and CL. Another language which presents this problem is Classical Sanskrit, where prevocalic high vowels become glides and do not cause lengthening of the following vowel (/iti abravit/ → [itj abravit] 'thus he said'). However, some vowel merger processes in Sanskrit do exhibit CL, namely the rule /ai au/ → [e: o:] (/ca iha/ → [ce:ha] 'and here').

In Matumbi, GF does take place in a context analogous to Ganda [wagula], as we observed in (26) with /u-ákite/ → [waákite] 'you hunted'.

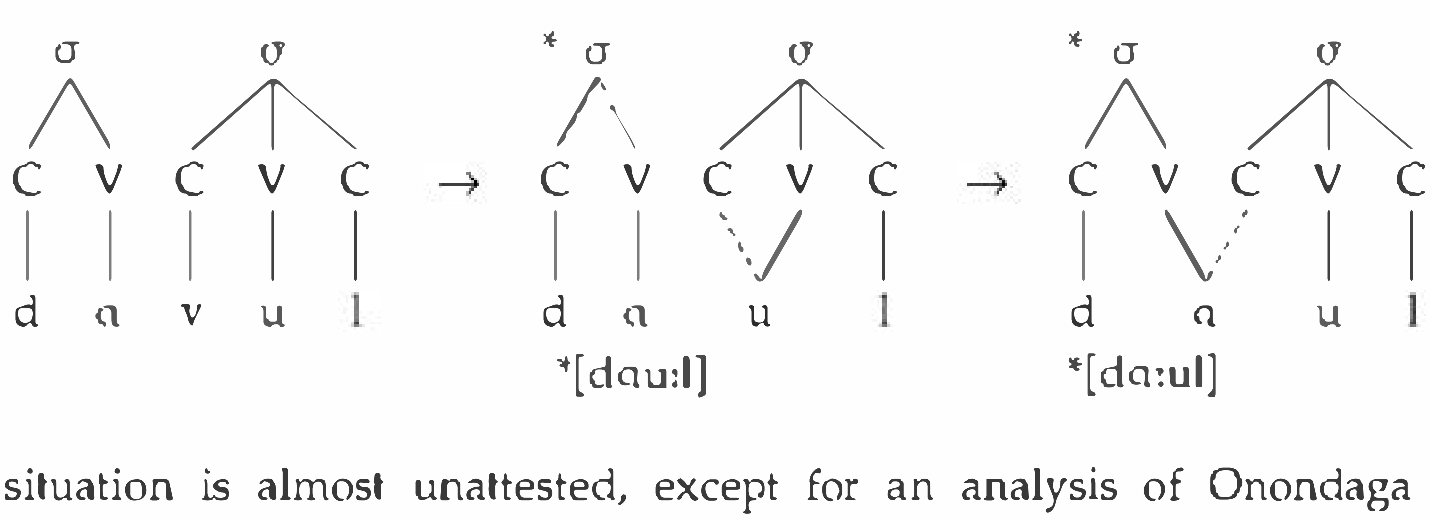
(33)



Creation of a C-slot is necessitated by the fact that CV positions always mediate between the segment and the syllable. This complicates CV theory by necessitating a convention inserting an onset which does not already exist, and is counter to the insight that CL is simply rearrangement of segmental and prosodic material.

CV theory predicts CL from deletion of an onset consonant, which in fact does not happen in Turkish and numerous other languages with CL associated with consonant deletion.

(34)



This situation is almost unattested, except for an analysis of Onondaga in Michelson (1986) where a vowel is lengthened when preceded by an empty C

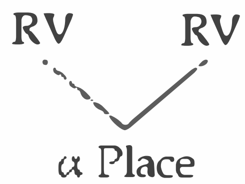
(37)

pɔdok	pɔd-pɔdok	'to be planting'
pa	pa:-pa	'to be weaving'
ca:k	ca:-ca:k	'to be bending'
onop	onn-onop	'to be preparing'
andip	and-andip	'to be splitting'

It does appear, then, that the C/V distinction is not strictly necessary, as long as one enriches underlying representations with higher-level structure with a nucleus node.

A conceptually related theory is that of Selkirk (1990), who dispenses with skeletal positions entirely and relies on "root node" to indicate length, so that a long vowel has two root nodes, but one set of shared place features.[7]

(38)



3.2 *Moraic accounts*

Building on the idea of generic prosodic positions, Hyman (1985) argues that one-to-many mappings between skeletal positions and segments should be further exploited, and proposes that onset consonants lose their skeletal position by joining to the position of the nuclear vowel through a universal Onset Creation Rule (OCR).

(39)

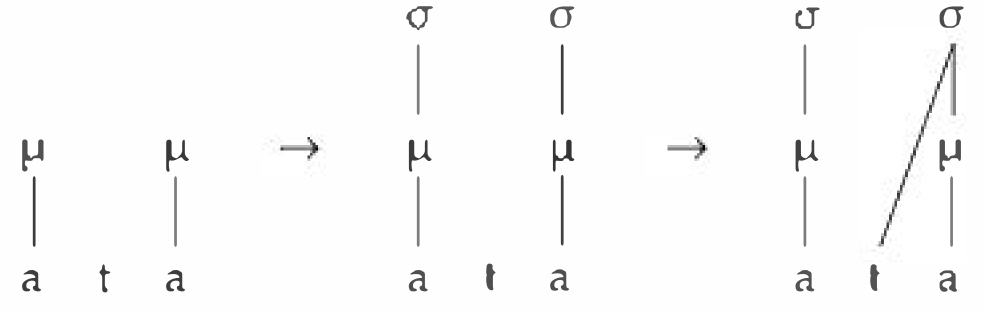
This theory of representation eventually develops into Moraic Theory.

Hyman's main concern is developing a syllable-free theory that still represents the notion of syllable-weight, a notion proposed by Kuryłowicz (1948) to unify syllables with long vowels or coda consonant for purposes of stress, tone, or prosodic morphology. Capturing the notion "heavy" in a theoretically coherent fashion had proven difficult (CHAPTER 57: QUANTITY-SENSITIVITY). Although "heavy syllable" correlates to branching, it has been challenging to say where such branching is relevant, since the onset is irrelevant to syllable weight (but see Davis, forthcoming), and a branching nucleus (long vowel) always defines a heavy syllable if a language has any notion of syllable weight, while a branching rhyme (XVC syllable) only sometimes does, and, furthermore, some languages (such as Lithuanian) distinguish sonorant rhymes as defining heavy syllables *vs.* obstruents for defining light ones. Hyman's theory of weight units (which he equates with the traditional mora) handles these facts quite simply, deriving onset-weightlessness from OCR (39), with a range of options regarding coda consonants which either leave coda consonants with their underlying weight unit (contributors to syllable weight), or merge that weight unit with the preceding vowel (codas are weightless).

[7] In Selkirk's theory, root nodes have intrinsic featural content, so that 'RV' would be [−cons, +son].

The notion of mora as the basic unit of phonological timing is developed more fully in McCarthy and Prince (1986), Hayes (1989), and Zec (1988). A strong argument for Moraic Theory over skeletal theories of length is advanced in Hayes (1989), who shows how a moraic theory of length coupled with the principle that onset consonants have no moraic value explains the onset asymmetry for compensatory lengthening, noted for example in Turkish /savmak/ → [sa:mak], but /davul/ → [daul], *[da:ul], *[dau:l] in Wetzels and Sezer (1986): see also (34) above. A characteristic of these later developments in Moraic Theory (exemplified by Hayes 1989) is that it is unnecessary to underlyingly assign weight units to all segments, and rather than postulating that every segment has a mora and invoking mora-trimming to merge the moras of onset consonants with those of a vowel, underlying representations will *typically* assign one mora to a short vowel and two to a long vowel; a consonant might have an underlying mora if it is geminate or syllabic (see CHAPTER 37: GEMINATES). The weightlessness of onsets then reduces to the fact that consonants ordinarily neither have an underlying mora nor do they receive one by rule, and instead directly syllabify to the left of a syllable by a syllable adjunction rule.

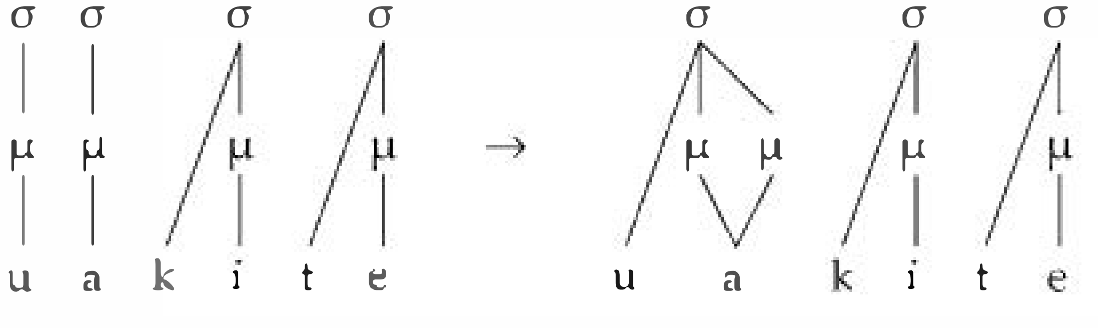
(40)



The generalization that long vowels always define a heavy syllable is a consequence of the moraic characterization of "heavy," namely having two moras within a syllable. Long vowels always have two moras; onsets do not contribute a mora (thus are weight-irrelevant), and coda consonants can vary in their influence on weight, depending on whether they receive a mora.

The moraic theory of direct attachment to the syllable without intervening timing units resolves the onset problem with skeletal theories of timing and CL. It was noted above that skeletal theories of CL do not provide a straightforward account of onset-creation co-existing with CL from GF in an underlyingly onsetless syllable. In (33) it was seen that GF in Matumbi results in compensatory lengthening of the following vowel, so /u-ákite/ becomes [w-aákite] 'you (sc) hunted', which required auxiliary insertion of C to syllabify the onset. The moraic theory accounts for this pattern easily, since the theory does not require prosodic positions for onset consonants: the high vocoid directly attaches to the syllable.

(41)



A final argument for the mora is mora-based tone systems such as that of Matumbi, where H is assigned by counting Vs – but only Vs – from the beginning of the verb. Skeletal theories face the problem of ignoring C-slots in tonal computations. Compare the problem of identifying the fourth "vowel" for Matumbi subjunctive tone assignment, given the representations of skeletal *vs.* moraic theories.

(42)	V C V C V C V C V u p a k a t í k e	μ μ μ μ μ u p a k a t í k e	'shake down'
	V C V C C V V C V u ʧ a ŋ g a á l e	μ μ μ μ μ u ʧ a ŋ g a á l e	'wonder'
	V C V V C C V C V u l a a mb á t e	μ μ μ μ μ u l a a m b á t e	'lick'
	V C V V C C V V C C V u k e e ŋ g é e mb e	μ μ μ μ μ μ u k e e ŋ g é e m b e	'dig up'

Moraic Theory directly represents the prosodic irrelevance of onset consonants via the fact that they project no prosodic unit, thus the fourth timing unit is easily identified. Skeletal theories require special interpretive procedures which render Cs or non-nuclear Xs invisible to the rule which scans for a fourth skeletal position.[8]

Moraic Theory constitutes an advance in our understanding of vowel length on various points: it resolves the problem of onset skeletal-position generation associated with CL, it explains why long vowels always define syllable weight, and it allows a simple representational account of mora-counting processes. Nevertheless, various problems with Moraic Theory have emerged in the literature, and if Moraic Theory is wrong as a theory, then it is wrong as a theory of vowel length. For example: Vago (1992) points to difficulties in a moraic analysis of Hungarian length; Tranel (1991) shows for a number of languages how the predicted weight/length equation is incorrect (geminate consonants do not uniquely contribute to weight); the notion of "weight" is vastly more subtle than the μ/μμ distinction implied by Moraic Theory (Crowhurst and Michael 2005).

It is proposed in Tranel (1991), Hume *et al.* (1997), Odden (1997), and Muller (2001) that prosodic theory needs both skeletal representations of length and moraic structure, a proposal which potentially allows preservation of the insights of both moraic and skeletal theories. A striking problem for Moraic Theory noted by Hume *et al.* is that in Leti, long consonants do not contribute to syllable weight as diagnosed by initial-stress assignment (which only affects long vowels).

(43)	tu'vuri	'kind of shell'
	'rɔː'nenu	'they eat turtle'
	vap'pure	'wild pig'
	ppu'narta	'nest's edge'

However, long vowels and consonants form a natural class in blocking a prosodic reduction process of "downgrading" with a range of segmental and stress

[8] Strict CV theory, which posits that all phonological strings are of the form (CV)*, does not face this problem since each mora is represented as an independent nucleus.

consequences. When the first word in a syntactically eligible structure contains only obviously light syllables, diphthongs, or non-geminate consonant clusters, as in (44a), downgrading is possible. When the first word contains a long segment, either a vowel or a consonant, downgrading is blocked.

(44) a.

Separate phrasing	*Downgraded*	
'sivi 'terannu	sɪvɪ 'terannu	'the egg of the chicken'
'spou 'ttɛnanne	spou 'ttɪnannɪ	'keel of the boat'
'ntutnu 'wai	ntutnu 'wai	'he lights the fire'

b. *No downgrading*

'nva:lu 'vatu	'he flings the stone'
'ppatne 'unne	'truck of the orange'

The problem then is that if length entails moraicity and weight is expressed by bimoraicity, we reach contradictory conclusions for Leti. A mixed theory with both skeletal position and moras, on the other hand, would allow heavy syllables (relevant for stress) to be defined with reference to bimoraicity, and length to be represented skeletally. In Leti, consonants including long consonants have no moraic value, whereas vowels do.

In general, the branching-prosody claim for representing long vowels is well supported, but evidence saying clearly what long vowels have two of is hard to come by.

4 How many lengths are there?

Nearly all languages with vowel length are treated as having two degrees of length, as predicted by Jakobsonian binary analysis, so reputed cases of ternary (or greater) length are of theoretical interest. The descriptive literature includes up to eight degrees of phonetic length in Finno-Ugric practices, reported in Sammallahti (1998). The question is whether that which distinguishes long and short extends in the same way to short, long, and overlong. Most languages with claimed ternary length have so little phonological evidence that we have only transcriptional distinctions to consider.

The most famous case of ternary length is Estonian, illustrated in (45). Diacritic notations of vowel length are regularized so that [vː] refers to long and [vːː] refers to extra-long. Estonian orthography does not represent the distinction between Q2 and Q3, except in stops.

(45)

saada	[saːːda]	'to get'
saada	[saːda]	'send!'
sada	[sada]	'hundred'

Lehiste (1966, 1970, 1978) and Prince (1980) show that segmentalizing this distinction obscures Estonian phonology, which is best treated as binary length intersecting with foot structure, where overlength (Q3) arises when one syllable exhausts the foot and plain length (Q2) results when a heavy syllable is followed by a syllable within the foot. In [saːːda], the long syllable exhausts a foot,

/(sa:)da/, whereas the long vowel in /(sa:da)/ is one of two syllables within the foot.

The strongest arguments for foot plus binary-length treatment come from an analysis of Estonian phonology: overlength only occurs in a stressed syllable; it affects computation of subsequent stresses; it is automatic on all final-stressed vowels; the distribution of overlength relative to plain length is not contrastive within the syllable (long vowels and consonants are either both Q2 or neither Q2, so syllables, and not segments, are overlong). See Prince (1980) for an extensive generative analysis. Foot structure also explains multiple surface lengths in Tiberian Hebrew: see the online appendix to this chapter for details of languages mentioned briefly here.

Applecross Gaelic (Ternes 1973) is another language claimed to have ternary length, but such phonological evidence as is available suggests that this results at least in part from higher-level foot-like prosody. Other languages claimed to have more than three degrees of length have been re-analyzed. Whiteley and Muli (1962) claim an actual four-way length distinction for Kikamba. Phonological analysis of Kikamba (Roberts-Kohno 2000) reveals that the durational differences are a fact, but analytically reduce to the fact that Kikamba has vowel length ([kokɔ̌mã] 'to lie down' *vs.* [kobɔ̌ɔ̀tà] 'to accomplish') as well as heterosyllabic identical vowel sequence ([kokɔ.ɔ̀mà], [kokɔ.ɔɔ̌tà] 'to pull'), pointing to one direction for re-analysis of multiple-length claims.

Multiple-length claims are sometimes grounded in vowel quality and tone facts. The German dialects around Dithmarschen and Stavenhagen (Hock 1986, citing Grimme 1922) have been said to have three vowel lengths, but Kohler (2001) points to vowel quality and tonal distinctions that make the claim for three lengths suspicious. Another case is Hopi, based on Whorf (1946), who mentions three lengths. Jeanne (1982) shows that there are two lengths and two contrastive pitches on long vowels.

Mixe dialects of Coatlán (Hoogshagen 1959) and San José el Paraiso (Van Haitsma and Van Haitsma 1976) present three vowel lengths, illustrated by the following Coatlán data.

(46)	poʃ	'a guava'	po:ʃ	'a spider'	po::ʃ	'a knot'
	pet	'a climb'	pe:t	'a broom'	pe::t	'Peter'

The longest quantity only appears in stressed syllables. Since the language has no vowel sequences, Kamba-style re-analysis into vowel cluster *vs.* long vowel would not be appropriate. Hoogshagen phonemicizes the contrast binarily as respectively /poʃ/, /po:ʃ/, /po:hʃ/. There is no direct evidence from rules that reveals how the three lengths behave phonologically. As discussed in the online version of this chapter, the distinction might lie in how coda consonants are prosodified. Two other languages said to have ternary length, where the phonological nature of that distinction is unclear, are Witchita and Yavapai, the latter case having been connected to distinctive pitch.

The most compelling case for ternary vowel length qua length and not an interaction between binary length plus an orthogonal property comes from Dinka, documented in Andersen (1987, 1993), Remijsen and Gilley (2008), and Remijsen and Manyang (2009). Certain grammatical categories cause vowel lengthening, and there are also lexical vowel length contrasts in verbs. Verbs in the past or 3rd singular have a lengthened vowel, with the consequence that within one and the same category, an underlyingly short vowel becomes long and an underlyingly long vowel becomes overlong.

(47) *Short stems* *Lengthened stems*

NEGATIVE	2SG	PAST	3SG	
tét	-tét	téet	-tèet	'pick'
tèet	-téet	téeet	-tèeet	'divulge'

Data of this sort show by processual analogy that the relationship Q1:Q2 is the same as Q2:Q3.

5 Summary

The first task in treating vowel length is distinguishing length from other properties. Difference in duration is what analysts usually pay attention to, but other segmental properties can have durational consequences. Phonological analysis of processes is necessary to argue for a surface difference of "length" *vs.* some other phonetic feature. Since some apparent instances of vowel length facts can (or must) be reduced to something other than length (foot structure in Estonian, or segment-to-syllable structure differences in Kamba), claims about length cannot gratuitously assume a duration-length equation.

The evidence for not representing vowel length with a feature analogous to [nasal] or [voice] is compelling, and numerous arguments support the "long vowels are two" theory, but it has been difficult to resolve what long vowels are two of. The initially attractive idea of long vowels being two identical segments does not work as a general theory of vowel length, because of the ambiguity problem (long vowels can act both like single segments and like sequences) and because languages can contrast a single long vowel with two identical short vowels.

With the addition of at least syllable structure, a coherent account of vowel length emerges, based on the idea that a long vowel is a single segment associated to two prosodic positions. Then the question arises, what are those positions: do they reflect a basic consonant *vs.* vowel distinction, or are they generic in nature? The evidence for a C/V distinction is weak in the face of a highly articulated theory of syllable structure, but the evidence for a highly articulated theory of the syllable à la X' syllable theory is weak in the face of a more stripped-down theory of the syllable with the somewhat richer theory of the timing positions offered by CV phonology. Moraic Theory has certain advantages over skeletal theories relevant to vowel length, in its treatment of compensatory lengthening of vowels, in particular the problem of onset-irrelevance. Various problems with Moraic Theory have been brought out; but these problems seem to center on problems for consonants, not vowel length. At this point, the conclusion that vowel length is represented by two higher-level entities linked to one lower-level thing seems firmly established, but the precise nature of those higher-level entities remains elusive.

ACKNOWLEDGMENTS

I would like to thank the editors and reviewers of this chapter, as well as Juliette Blevins, Anna Bosch, Sam Meier, Bert Remijsen, and Tobias Scheer for comments.

REFERENCES

Andersen, Torben. 1987. The phonemic system of Agar Dinka. *Journal of African Languages and Linguistics* 9. 1–27.

Andersen, Torben. 1993. Vowel quality alternation in Dinka verb inflection. *Phonology* 10. 1–42.

Cheng, Chin-Chuan & Charles W. Kisseberth. 1979. Ikorovere Makua tonology. *Studies in the Linguistic Sciences* 1. 31–63.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. 1986. Compensatory lengthening and consonant gemination in LuGanda. In Wetzels & Sezer (1986), 37–77.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Crowhurst, Megan J. & Lev Michael. 2005. Iterative footing and prominence-driven stress in Nanti (Kampa). *Language* 81. 47–95.

Davis, Stuart. Forthcoming. Quantity. In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.) *The handbook of phonological theory*. 2nd edn. Malden, MA & Oxford: Wiley-Blackwell.

Fidelholtz, James L. 1971. On the indeterminacy of the representation of vowel length. *Papers in Linguistics* 4. 577–594.

Grimme, Hubert. 1922. *Plattdeutsche Mundarten*. 2nd edn. Berlin & Leipzig: de Gruyter.

Halle, Morris. 1977. Tenseness, vowel shift, and the phonology of the back vowels in Modern English. *Linguistic Inquiry* 8. 611–626.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hock, Hans Henrich. 1986. Compensatory lengthening: In defense of the concept "mora." *Folia Linguistica* 20. 431–460.

Hoogshagen, Searle. 1959. Three contrastive lengths in Mixe. *Zeitschrift für Phonetik* 12. 111–115.

Howard, Irwin. 1972. A directional theory of rule application in phonology. Ph.D. dissertation, MIT.

Hume, Elizabeth, Jennifer Muller & Aone van Engelenhoven. 1997. Non-moraic geminates in Leti. *Phonology* 14. 371–402.

Hyman, Larry M. 1982. The representation of length in Gokana. *Proceedings of the West Coast Conference on Formal Linguistics* 1. 198–206.

Hyman, Larry M. 1985. *A theory of phonological weight*. Dordrecht: Foris.

Ingria, Robert. 1980. Compensatory lengthening as a metrical phenomenon. *Linguistic Inquiry* 11. 465–495.

Jeanne, LaVerne Masayesva. 1982. Some phonological rules of Hopi. *International Journal of American Linguistics* 48. 245–270.

Jensen, John T. & Margaret Stong-Jensen. 1973. Ordering and directionality of iterative rules. *Papers in Linguistics* 6. 66–90.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1985. The internal structure of phonological elements: A theory of charm and government. *Phonology Yearbook* 2. 305–328.

Kenstowicz, Michael. 1970. On the notation of vowel length in Lithuanian. *Papers in Linguistics* 3. 73–113.

Keyser, Samuel J. & Paul Kiparsky. 1984. Syllable structure in Finnish phonology. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*, 7–31. Cambridge, MA: MIT Press.

Kisseberth, Charles W. 1973. On the alternation of vowel length in Klamath: A global rule. In Michael Kenstowicz & Charles W. Kisseberth (eds.) *Issues in phonological theory: Proceedings of the Urbana Conference on Phonology*, 9–26. The Hague: Mouton.

Kohler, Klaus J. 2001. Überlänge im Niederdeutschen? In R. Peters, H. P. Pütz & U. Weber (eds.) *Vulpis Adolatio: Festschrift für Hubertus Menke zum 60. Geburtstag*, 385–402. Heidelberg: Carl Winter.

Krueger, John. 1961. *Chuvash manual: Introduction, grammar, reader, and vocabulary*. Bloomington: Indiana University Press.

Kuryłowicz, Jerzy. 1948. Contribution à la théorie de la syllabe. *Bulletin de la Société Polonaise de Linguistique* 8. 80–113.

Ladd, D. Robert, Bert Remijsen & Caguor Adong Manyang. 2009. On the distinction between regular and irregular inflectional morphology: Evidence from Dinka. *Language* 85. 659–670.

Lehiste, Ilse. 1966. *Consonant quantity and phonological units in Estonian*. Bloomington: Indiana University Press.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, Ilse. 1978. The syllable as a structural unit in Estonian. In Alan Bell & Joan B. Hooper (eds.) *Syllables and segments*, 73–84. Amsterdam: North Holland.

Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.

Lowenstamm, Jean. 1996. CV as the only syllable type. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, 419–441. Salford: ESRI.

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.

McCarthy, John J. 1982. Prosodic templates, morphemic templates, and morphemic tiers. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part I, 191–223. Dordrecht: Foris.

McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague & Paris: Mouton.

Michelson, Karin. 1986. Ghost r's in Onondaga: An autosegmental analysis of *r stems. In Wetzels & Sezer (1986), 147–166.

Muller, Jennifer. 2001. The phonology and phonetics of word-initial geminates. Ph.D. dissertation, Ohio State University.

Odden, David. 1987. Predicting tone in Kikuria. In David Odden (ed.) *Current approaches to African Linguistics*, vol. 4, 311–326. Dordrecht: Foris.

Odden, David. 1995. The status of onsetless syllables in Kikerewe. *Ohio State University Working Papers in Linguistics* 47. 89–110.

Odden, David. 1996. *The phonology and morphology of Kimatuumbi*. Oxford: Clarendon Press.

Odden, David. 1997. Some theoretical issues in Estonian prosody. In Ilse Lehiste & Jaan Ross (eds.) *Estonian prosody: Papers from a symposium*, 165–195. Tallinn: Institute of Estonian Language.

Pam, Martin. 1973. Tigrinya phonology. Ph.D. dissertation, City University of New York.

Prince, Alan. 1980. A metrical theory for Estonian quantity. *Linguistic Inquiry* 11. 511–562.

Pyle, Charles. 1971. On the treatment of length in generative phonology. Ph.D. dissertation, University of Illinois.

Remijsen, Bert & Leoma Gilley. 2008. Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *Journal of Phonetics* 36. 318–344.

Remijsen, Bert & Caguor Manyang. 2009. Luanyjang Dinka: Illustration of the IPA. *Journal of the International Phonetic Association* 39. 113–124.

Roberts-Kohno, R. Ruth. 2000. Kikamba phonology and morphology. Ph.D. dissertation, Ohio State University.

Sammallahti, Pekka. 1998. *The Saami languages: An introduction*. Kárášjohka: Davvi Girji.

Selkirk, Elisabeth. 1990. A two-root theory of length. *University of Massachusetts Occasional Papers* 14. 123–171.

Ternes, Elmar. 1973. *The phonemic analysis of Scottish Gaelic, based on the dialect of Applecross, Ross-shire*. Hamburg: Buske.

Tranel, Bernard. 1991. CVC light syllables, geminates and moraic theory. *Phonology* 8. 291–302.

Vago, Robert M. 1992. The root analysis of geminates in the moraic phonology of Hungarian. In István Kenesei & Csaba Pléh (eds.) *Approaches to Hungarian*, vol. 4: *The structure of Hungarian*, 177–194. Szeged: JATE.

Van Haitsma, Julia D. & Willard Van Haitsma (1976). *A hierarchical sketch of Mixe as spoken in San José El Paraiso*. Norman, OK: Summer Institute of Linguistics.

Wetzels, W. Leo & Engin Sezer (eds.) 1986. *Studies in compensatory lengthening*. Dordrecht, Foris.

White, Robin B. 1972. Klamath phonology. Ph.D. dissertation, University of Washington.

Whiteley, Wilfred H. & M. G. Muli 1962. *Practical introduction to Kamba*. London: Oxford University Press.

Whorf, Benjamin Lee. 1946. The Hopi language, Toreva dialect. In Cornelius Osgood (ed.) *Linguistic structures of native America*, 158–183. New York: Viking Fund.

Zec, Draga. 1988. Sonority constraints on prosodic structure. Ph.D. dissertation, Stanford University.

21 Vowel Height

Douglas Pulleyblank

1 Introduction

While all languages have vowels, and all vowels can be described as having a height,[1] actually defining vowel height turns out to be a rather interesting problem (see also CHAPTER 19: VOWEL PLACE). Consider the systems in (1).

(1) *Sample vowel inventories*

a. *2 vowels*			b. *5 vowels*			c. *7 vowels*			d. *10 vowels*		
ɨ			i		u	i		u	i		u
a			e		o	e		o	ɪ		ʊ
				a		ɛ		ɔ	e		o
							a		ɛ		ɔ
										ɐ	
										ɑ	

It would be fairly uncontroversial to assume that a two-vowel system such as (1a) distinguishes between two vowel heights and that a five-vowel system such as (1b) distinguishes between three heights (although see below). The number of "heights" in the seven- and ten-vowel systems, however, depends on various factors. One view of a ten-vowel system such as (1d) is that it involves three "heights" (high, mid, low) cross-cut by a tongue-root feature; an alternative is that such a system involves a highly differentiated height feature with six heights. A seven-vowel pattern such as (1c) presents more analytic indeterminacy. One possibility is to analyze such a pattern as a simplified version of the ten-vowel system: a three-height system with a tongue-root distinction in the mid vowels, or a four-height system with a differentiated height pattern as in (1d), but with four rather than six distinctions. An additional possibility is to consider the seven-vowel and ten-vowel systems to be qualitatively different: while a ten-vowel inventory like (1d) could be analyzed as involving three heights plus a tongue-root distinction, a seven-vowel inventory like (1c) could be seen as involving four heights (and no tongue-root distinction).

[1] This may depend partly on how we define vowels, since various segments may be syllable nuclei without being vocalic.

The following sections discuss sample phonological patterns involving vowel height, considering the types of factors that bear on the analysis of such systems, both phonetic and phonological, as well as the types of proposals that have been made in the phonological literature for the formal instantiation of vowel height. It will be shown that "height" for some researchers reflects the operation of a single parameter, while for others it is the manifestation of several largely unrelated variables.

2 Phonetics of vowel height

There are three basic ways of defining vowel height phonetically: articulatorily, acoustically, and auditorily (Ladefoged and Maddieson 1996). These three possibilities produce rather different pictures of what vowel height might mean, none of which correspond unproblematically to the categories that seem relevant for phonology. Articulatorily, "vowel height" might be taken to suggest a dimension based on the height of the tongue body. While this is a convenient simplification for pedagogical reasons (see Rogers 1991: 170, who defines [height] as "a multivalued feature distinguishing vowels by the position of the highest point of the tongue"), it turns out that tongue body height is only sometimes correlated directly with phonological vowel height (Ladefoged and Maddieson 1996). Different ways of approaching vowel height articulatorily are possible, but none reproduces straightforwardly the sort of vowel space usually assumed by phonologists.

A plausible explanation for this discrepancy could be that vowel height depends more on the acoustic properties of the vocal tract than on any specific articulatory means of achieving the appropriate effect (Elorrieta Puente 1996). According to this view, vowel height could be acoustically defined in terms of the first formant, i.e. F1. This position was taken by Ladefoged (1975: 265), who defined the "prime feature [height] as the inverse of the frequency of the first formant." Although in later editions of Ladefoged's book the section on prime features was removed, the definition of height as correlating inversely with the value of the first formant was maintained (Ladefoged 2006: 188). While certainly an accurate indication of height in many instances, such an acoustic definition of height also appears to be inadequate in a number of cases. One example involves the determination of vowel height in languages with tongue-root advancement/retraction. As shown by Lindau (1979), Ladefoged and Maddieson (1996), Gick *et al.* (2006), and others, some tongue-root harmony languages exhibit advanced mid vowels such as [e] that are "higher" (that is, lower F1) than the corresponding retracted high vowel such as [ɪ]. Such systems cannot assign vowels to a particular height solely on the basis of the value for the first formant.

Auditory characterizations of vowel height are adopted in much phonological work (Lindau 1978; Casali 1996). From introductory textbooks, to IPA charts, to research articles, we are accustomed to two-dimensional representations of vowels where "height" is represented on the vertical axis and backness is represented on the horizontal axis. In Jones's (1972) proposal for cardinal vowels, reference vowels were established on the basis of articulatory extremes ([i ɑ]), with additional reference points defined in terms of auditorily equidistant steps.[2]

[2] Jones (1972: 32) refers to "acoustically equidistant vowels," but it is clear that he means vowels as perceived by a trained listener.

While the examples in (11a) have a consistent [a] as their second vowel, those in (11b) alternate, exhibiting [u] in non-word-final position and [a] in word-final position. To postulate an underlying /a/ for (11b) would be problematic, since the vowel behaves differently from the consistent [a] of (11a). To postulate an underlying /u/ for (11b) is unproblematic, making the correct prediction that there should be no cases of this type with a word-final [u]. Hence such data motivate a pattern of word-final lowering to [a].

This case of lowering is different from the cases seen so far, in that it is positionally triggered, not assimilatory. In addition, additional forms raise an interesting issue.

(12) *Lardil front vowels*

	uninflected	*non-future*	*future*	
a.	tjempe	ṯjempe-n	ṯjempe-ɽ	'mother's father'
	wiṯe	wiṯe-n	wiṯe-ɽ	'interior'
b.	ŋin̪e	ŋin̪i-n	ŋin̪i-wuɽ	'skin'
	pape	papi-n	papi-ɽ	'father's mother'

Comparable to (11), some forms show an invariant [e], while others show variation between word-final [e] and word-medial [i]. As with (11), this can be analyzed as involving /e/ for (12a) and /i/ for (12b), in conjunction with a rule of lowering.

It appears from such data that the result of lowering is low when /u/ is the target but mid when /i/ is the target. While possible, it should be noted that the vowel inventory of Lardil is {i e a u}. It appears, therefore, that Lardil distinguishes between only two heights, with both [e] and [a] constituting "low" vowels.

4.2.3 *Chain shifts: One-step lowering*

Paralleling the cases of stepwise raising seen above, cases of stepwise lowering are also attested. Esimbi (Bantu, Niger-Congo; Cameroon) is an intriguing case (Hyman 1988). There are three underlying prefix vowels and eight underlying stem vowels, but the underlying pattern is obscured since the height of the stem vowel is transferred to the prefix, the stem itself surfacing uniformly high. This height transfer is illustrated with the prefix vowel /I/ "class 9 singular," "class 10 plural."

(13) *Esimbi class 9/10 nouns*

	singular	*plural*	
/i/	ì-bì	í-bi	'goat'
/u/	ì-sùmù	ì-sumu	'thorn'
/e/	è-gbì	é-gbi	'bushfowl'
/o/	è-nùnù	e-núnu	'bird'
/ə/	è-bi	e-bi	'canerat'
/ɛ/	ɛ̀-ɲìmì	ɛ-ɲimì	'animal'
/ɔ/	ɛ̀-fùmù	ɛ-fumù	'hippo'
/a/	ɛ̀-kìrɨ̀	ɛ-kirɨ	'headpad'

With the highest stem vowels /i u/, the prefix surfaces as [i]; with the next highest stem vowels /e o ə/, the prefix surfaces as [e]; with the lowest stem vowels to height. There are cases, however, where height assimilation involves a target becoming identical to a trigger with respect to its height specifications.

Consider lowering in Matumbi (Bantu, Niger-Congo; Tanzania; Odden 1991). When the causative suffix, which Odden analyzes as being underlyingly high and advanced (16a), (16b), (16g), follows a high retracted vowel, it is realized as high retracted (16c), (16d); when it follows a mid vowel, it is realized as mid (16e), (16f).

(16) *Matumbi*

a.	út-a	'pull'	út-ij-a	'make pull'
b.	jíb-a	'steal'	jíb-ij-a	'make steal'
c.	jʊ́joot-a	'whisper'	jʊ́jout-ɪj-a	'make whisper'
d.	bíɪk-a	'put'	bíɪk-ɪj-a	'make put'
e.	gɔ́ɔnʤ-a	'sleep'	gɔ́ɔnʤ-ɛj-a	'make sleep'
f.	ʧɛ́ɛŋg-a	'build'	ʧɛ́ɛŋg-ɛj-a	'make build'
g.	káat-a	'cut'	káat-ij-a	'make cut'

In a pattern of this type, the height assimilation involves trigger–target identity for height. Moreover, it should be noted that more than two "heights" are involved in this shift; identity is not a result simply because the system distinguishes only between high and non-high vowels.

4.3.2 *Harmony*

Whether harmonic patterns based on vowel height are considered common or rare depends in part on whether "ATR"-like systems are considered to be height harmony. If [ATR] is a height feature, then height harmonies are quite common; if [ATR] is a place feature, then "height"-based harmony systems appear less common (CHAPTER 19: VOWEL PLACE). Nevertheless, even abstracting away from cross-height harmony systems of the "ATR" type, harmonic systems based on height do occur; for discussion of a wide variety of systems, see Linebaugh (2007).

In this section, a small number of harmonic systems based on height are considered, including systems of the cross-height "ATR" type. It should be noted, however, that "harmony" is used as a descriptive term only: there is no incontrovertible reason to think that there is a formally definable type of phonological pattern corresponding to those patterns we consider to be harmonic (Archangeli and Pulleyblank 2007).

To start, C'Lela (Benue-Congo, Niger-Congo; Nigeria) presents a particularly interesting case of height harmony (Dettweiler 2000; Pulleyblank 2002). In C'Lela, within a morpheme, only vowels of like height may co-occur: high vowels may co-occur, non-high vowels may co-occur, but vowels of mixed height may not.[6]

(17) *C'Lela root-internal harmony*

	High vowels			*Non-high vowels*	
a.	d^{ə}tɨndi	'nest'	b.	kʋesa	'show'
	ʧ^{ə}rini	'charcoal'		ʧ^{ə}gjɔmbo	'eyebrows'
	k^{ə}piru	'flower'		d^{ə}ʋɛso	'broom'

[6] The raised schwa represents a short transitional vowel that Dettweiler (2000) considers "non-phonemic," appearing in certain types of consonant clusters.

(20) *Degema stems*

a. *Advanced*		b. *Retracted*	
gbodí	'catch'	fowɔ́	'be cool (of food)'
gurón	'postpone'	ganɛ́	'have as one's lover'
ɗimé	'submerge'	hɔβá	'bail out water from a canoe, etc.'
lebə́	'be partially ripe (of fruits)'	sɪgɔ́m	'jump with one leg folded'

Prefixes in Degema must agree in their tongue-root value with vowels of the stem. Consider, for example, noun class prefixes such as those in (21), where (21a) exhibit advanced values and (21b) retracted values.

(21) *Degema prefixes*

a. *singular*	*plural*		b. *singular*	*plural*	
u-túm	ə-túm	'tail'	ʊ-ɗɔ́m	a-ɗɔ́m	'marriage'
o-kpokí	i-kpokí	'money'	ɔ́-hɔ́hɔ	í-hɔ́hɔ	'fowl'
e-kúnési	i-kúnési	'bed'	ɛ-ɗɔ́ŋʷ	ɪ-ɗɔ́ŋʷ	'throat'
ə́-milíɓə	í-milíɓə	'night heron'	a-ɗʊ́	ɪ-ɗʊ́	'face'

Although less common than prefixes in Degema, suffixes also agree in their tongue-root value with the stem to which they attach. Consider examples of the gerundive.

(22) *Degema suffixes*

a. *Advanced*		b. *Retracted*	
ù-dér-ˈə́m	'cooking'	ʊ̀-tɛ́v-ˈám	'descending'
ù-vój-ˈɔ́m	'fetching'	ʊ̀-sɔ́l-ˈám	'jumping'

Overall, height harmony systems are amply attested and instantiate the typical range of variables discussed in the harmony literature: dominant/recessive *vs.* root-controlled patterns, transparency *vs.* opacity, directionality, and so on. There are also cases illustrating interesting interactions with stress; see, for example, Fitzgerald (2002); Walker (2005).

4.4 Neutralization

A large and interesting class of patterns involving vowel height is that of large-scale neutralization in languages where different vowel inventories are attested in different positions (Crosswhite 2001). An example is Catalan (Romance, Indo-European; Spain; Mascaró 1983):

(23) *Catalan vowel reduction*

Stressed			*Unstressed*		
/i/	'prim	'thin'	[i]	əpri'ma	'to make thin'
/e/	'serp	'snake'	[ə]	sər'pɔtə	'big snake'
/ɛ/	'pɛl	'hair'		pə'lut	'hairy'
/a/	'sak	'sack'		sə'kɛt	'small sack'
/ɔ/	'pɔrt	'harbor'	[u]	purtu'ari	'related to harbor (ADJ)'
/o/	'gos	'dog'		gu'sas	'big dog'
/u/	'ʎum	'light'		ʎumi'nos	'light (ADJ)'

As illustrated in (23), seven vowels are found in stressed positions in Catalan, but these distinctions collapse to three when the corresponding vowels appear in unstressed positions. In stressed positions, three or four vowel heights are distinguished, depending on one's analysis of the vowels [e o] *vs.* [ɛ ɔ]; in unstressed positions, a two-way vowel height distinction is observed. This case illustrates a general pattern where fewer vowel heights are exhibited in positions of reduction than in positions of full contrast (CHAPTER 79: REDUCTION). Note in this regard that the Catalan case exhibits no loss of "color" distinctions, that is, distinctions involving [back] and [round].

4.5 *Interactions between consonants and vowel height*

Although the focus of this chapter is vowels, not all issues of vowel height are restricted to vowels. There are numerous examples of consonant–vowel interactions where vowel height is relevant, involving vowels affecting consonants as well as consonants affecting vowels. Samples of both types are considered in this section. (See also CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS for some related phenomena.)

4.5.1 *Vowels affecting consonants*

As a first example of vowels affecting consonants, consider assibilation in Canadian French (Romance, Indo-European; Canada; Dumas 1978; Walker 1984). The coronal stops [t d] surface without change when before non-high or back vowels (e.g. [taʀt] 'pie', [tɛm] 'topic, theme', [do] 'back', [dɑ̃] 'tooth'), but they become affricates before high front vowels or glides.

(24) *Canadian French assibilation*

[ts]			[dz]	
[i]	patsi	'small'	ʒødzi	'Thursday'
[ɪ]	tsɪg	'tiger'	ʒʏʀɪdzɪk	'legal'
[ɪ^{i}]	etsɪ^{i}ʀ	'stretch'	dzɪ^{i}ʀ	'say'
[j]	tsjɛd	'lukewarm'	dzjɛt	'diet'
[y]	tsynɛl	'tunnel'	ʀɑ̃dzy	'returned'
[ʏ]	tsʏb	'tube'	pɑ̃dzʏl	'pendulum'
[ʏ^{y}]	fʀitsʏʀ	'frying'	dzʏ^{y}ʀ	'hard'
[ɥ]	tsɥɪl	'tile'	ʀedzɥɪ^{i}ʀ	'reduce'

Assibilation takes place whether the trigger is rounded or unrounded, tense or lax, a monophthong or a diphthong, a vowel or a glide. Dumas (1978) takes this as evidence that the distinction between the tense vowels [i y] and their lax counterparts [ɪ ʏ] (which is determined allophonically by syllable structure) is not a difference of vowel height. Auditorily, he argues that there is no significant difference between the mid vowels [e ø o] and the high lax vowels. Consequently, if laxing involved a change in height, we would expect lax vowels not to be assibilation triggers; that they are triggers suggests that the laxing process affects a dimension other than height.

These examples constitute one small example from the large class of palatalization cases attested cross-linguistically (see CHAPTER 71: PALATALIZATION). They are of relevance to vowel height since height is a typical delimiter on the class of vowels that triggers such processes.

A rather different type of consonant–vowel interaction is found in Chumburung (Kwa, Niger-Congo; Ghana; Snider 1984): [l] (a lateral) and [r] (a "lightly-retroflexed alveolar flap") are in complementary distribution. For some speakers, the distribution is unrelated to vowels, with [l] occurring in word and noun-stem-initial position ([lɔ́sɛ́] 'difficult'; [kì-lìmpɔ́] 'shea nut') and in a syllable following another [l] ([lóolí] 'deep'), while [r] occurs elsewhere ([jono-ro] 'in (the) dog'). For some speakers, however, this allophonic pattern between [l] and [r] interacts with vowel harmony, a pattern in Chumburung that is largely analogous to the tongue-root harmony system seen for Degema in §4.3.2, with the difference that Chumburung has a single (retracted) low vowel. For the relevant speakers, medial [l] only surfaces after another [l] if the vowels are in the [+ATR] set:

(25) *Chumburung [l] ~ [r]*

a.	[+ATR]		b.	[−ATR]	
	làlàkwìʔ	'type of tuber'		làarɪ́	'waist'
	lóolí	'deep'		lɔ̀ɔrɪ́	'to remove seeds'
	àlúuláʔ	'red dye'		làarɔ́	'to lie across'

Alternations can be observed under suffixation.

(26) *Chumburung suffix alternations:* /-rɔ/ (LOCATIVE)

a. jono-ro 'in (the) dog'
b. kanɔ-rɔ 'in (the) mouth'
c. lɔ-rɔ 'in (the) sore'
d. laale-lo 'in (the) cattle egret'

Since the suffix-initial consonant is neither word-initial nor stem-initial, it would be correctly expected to surface as [r] in the default case (26a, 26b) and even after [l] if the vowel harmony class is [−ATR] (26c); if the suffix follows an [l], however, and if the harmonic class is [+ATR], then the suffix surfaces with [l] (26d).

This particular kind of case is irrelevant if [ATR] is not a height feature. The point, however, is that height and/or tongue-root distinctions in vowels condition a range of consonantal effects: palatalization, differences in liquids, differences between velars and uvulars (Li 1996), and so on.

4.5.2 *Consonants affecting vowels*

Elorrieta Puente (1996) considers a wide range of cases where consonants affect vowel height, identifying two principal types of cases: (a) raising, triggered typically by palatals, palato-alveolars, or glides; (b) lowering, triggered typically by uvulars, pharyngeals, laryngeals, or rhotics. (See also Vaux (1996, 1998) and Fitzgerald (2002) for cases involving laryngeal features; and CHAPTER 25: PHARYNGEALS.)

Consider first an example of raising, taken from Kikuria (Wiswall 1991; Elorrieta Puente 1996). The relevant example involves a pattern of vowel raising triggered by both vowels any consonants. When a mid vowel prefix appears before a non-high stem vowel, the prefix surfaces as mid.[7]

[7] The k~g alternations observed in the Kikuria data are due to Dahl's Law, an unrelated pattern of voicing dissimilation. See, for example, Davy and Nurse (1982).

(27) *Kikuria mid vowel prefixes*

oko-raara	'sleep'
oko-roma	'bite'
oko-gɛra	'weed'

In contrast, when such a prefix occurs before a high vowel stem, the prefix raises to high.

(28) *Kikuria raising: High vowel triggers*

uku-rugja	'chase'
ugu-siika	'close a door'

Of particular interest, the mid vowel of the prefix is also raised if the first syllable either contains a glide [j w] or a palatal consonant [ʧ ɲ].

(29) *Kikuria raising: Consonantal triggers*

ugu-tweeba	'forget us'
uku-bjoora	'remove from water'
ugu-ʧɔɔra	'draw'
ukuu-ɲandekera	'write for me'

In such cases, it is perhaps unsurprising that glides behave in a manner comparable to vowels, since glides are simply the non-syllabic counterpart of the high vowels that are canonical triggers of raising (CHAPTER 15: GLIDES). For the non-glides, the fact that palatals and alveo-palatals are triggers of raising in Kikuria but that velars are not is typical of such raising patterns cross-linguistically. Being "high" seems to be required of a consonantal trigger, but assuming that velars are also high means that simply being high is not sufficient in Kikuria.

With respect to lowering, the typical case is one where lowering is induced by the class of gutturals. Although this class may vary somewhat from language to language, it is generally defined by segments articulated in the post-velar region. Consider, for example, lowering in Gitksan (Tsimshianic; Canada; Brown 2008), illustrated here with plural reduplication. As seen in (30), there is a class of plurals formed by the prefixation of either C*i*(C)- or C*ix*-.

(30) *Gitksan plural reduplication*

saksxʷ	six-saksxʷ	'be clean'
dzam	dzim-dzam	'cook, boil (TRANS)'
ɬak	ɬi-ɬak	'be crooked'
gʷalkʷ	gʷil-gʷalkʷ	'be dry'

When the vowel of the reduplicative prefix precedes or follows a uvular or a laryngeal, the vowel lowers to [a].

(31) *Gitksan lowering*

dzoq	dzaχ-dzoq	'camp'
ɢetxʷ	ɢa-ɢetxʷ	'be diffiicult, be expensive'
ɢoːt	ɢa-ɢoːt	'heart'
hets	has-hets	'send'
ʔos	ʔas-ʔos	'dog'

The Gitksan pattern is one where lowering is triggered by uvulars and laryngeals. Perhaps more typical would be a case where vowels are lowered by a "guttural" class including pharyngeals (CHAPTER 25: PHARYNGEALS). Setting aside the presence/absence of pharyngeals, languages exhibiting lowering differ as to whether laryngeals are included (e.g. Arabic, Semitic, Afroasiatic; McCarthy 1994) or excluded (e.g. Nuu-chah-nulth, Wakashan; Canada; Wilson 2007); see Rose (1996).

Two final patterns touched upon here are ones where two input vowels merge into a single output vowel (coalescence) and where a single vowel splits into more than one component (diphthongization).

4.6 Coalescence

Patterns of vowel coalescence are frequently height-dependent. Consider the following examples from Anufɔ (Kwa, Niger-Congo; Ghana; Casali 1996).

(32) *Anufɔ vowel contact*

a.	/jɛ-i/	jɛː	'raise it'
b.	/ʤɔ-u/	ʤɔː	'cool you'
c.	/bu-i/	bwiː	'break it'
d.	/fa-i/	fɛː	'take it'
e.	/bo-i/	bweː	'beat it'
f.	/sɔ-i/	swɛː	'carry it'
g.	/fa-u/	fɔː	'take you'
h.	/bo-u/	boː	'beat you'
i.	/n-de-u/	ndoː	'I will take it for you'

Different subcases are of interest for different, though related, reasons. In examples like (32a), (32b), the output contains no discernible trace (other than length) of the high vowel in the input. In cases like (32c), (32e), (32f), labiality is never lost, resulting in a glide–vowel sequence when the first vowel of the sequence is labial and the second vowel is not. The cases of particular interest for vowel height, however, are ones like (32a), (32b), (32e), (32f), (32h), (32i), on the one hand, and (32d), (32g), on the other. In the former type, we see that the combination of a mid vowel with a high vowel results in a long mid vowel; in the latter type, we see that the combination of a low vowel and a high vowel also results in a long mid vowel. Note in particular that the output of a low–high sequence is neither low nor high, but an intermediate vowel.

The pattern observed in Anufɔ is typical in terms of what Casali (1996) refers to as "height coalescence." He observes that coalescence is frequent in cases where V1 is non-high (e.g. [a]) and V2 is higher than V1 (e.g. [i] or [u]), the result being non-high with the place properties of V2. The results in (32d), (32g) are canonical with respect to this pattern. For an in-depth discussion of such patterns, along with the identification of robust subpatterns, see Casali (1996).

4.7 Diphthongization

In patterns of diphthongization, vowel height is frequently an important variable, with diphthongs related to a set of corresponding monophthongs in ways

5.1 Features constituting "height"

As discussed in §2, there is no way a priori to decide that a particular feature should be a height feature. For example, [ATR] might be classified as a "height" feature if acoustics was taken as definitional (due to its effect on F1), but not as a "height" feature if articulatory tongue body height was taken as definitional. Consequently, theories vary as to whether [ATR] is included as a feature or not (Clements and Hume 1995; Vaux 1996), and if it is a feature, whether it is a height feature or not. For example, Clements (1990) and Clements and Hume (1995) discuss the possibility that [ATR] could be eliminated; Odden (1991) maintains a feature [ATR], arguing that it is under a "height" node that dominates [high], [ATR], and (perhaps) [low]; Parkinson (1996) proposes a "height" constituent consisting solely of instances of the feature [closed], with [ATR] analyzed outside of the "height" node as a place feature; Elorrieta Puente (1996: 6) considers that "height" involves a single feature composite of the dimensions "high–low, tense–lax, and advanced–nonadvanced tongue root." Similar issues arise about potential height features involving tenseness and peripherality.

An additional issue affecting featural analyses of height involves the division of vowels, particularly "lowish" vowels, into classes. Parkinson (1996: 117–119), for example, argues that although Njebi, Basaa and Efik all have the vowels {i e ɛ a ɔ o u}, the breakdown of the three inventories into vowel heights is different in each case.[8] For Njebi, as seen in (8) above, four heights need to be distinguished since vowel raising distinguishes between four classes: [a] raises to [ɛ]; [ɛ ɔ] raise to [e o]; [e o] raise to [i u].

(35) *Differing interpretations of comparable vowels*

a. *Njɛbi*		b. *Basaa*		c. *Efik*	
Height 1	[i u]	Height 1	[i u]	Height 1	[i ɨ u]
Height 2	[e o]	Height 2	[e o]	Height 2	[e ɛ a ɔ o]
Height 3	[ɛ ɔ]	Height 3	[a ɛ ɔ]		
Height 4	[a]				

In contrast, raising in Basaa involves only three heights: [a ɛ ɔ] raise to [e o]; [e o] raise to [i u]. Finally Efik distinguishes two heights: [i ɨ u] *vs.* [e ɛ a ɔ o]. Those vowels not distinguished by height are differentiated by place specifications like [labial], [coronal], and [pharyngeal]. Overall, establishing appropriate specifications for vowel height is a complex problem involving the analysis of both height and place and, depending on the theory, analyses of superficially similar inventories may vary considerably from language to language.

A number of proposals for height are considered below, including what the features are and how they are situated within a broader theory of feature composition and structure.

5.2 Features: [high] and [low]

Following Chomsky and Halle (1968; *SPE*), the dominant approach to vowel height over the years has been to base vowel height on the interaction of two binary features, [±high] and [±low].

[8] Not relevant to the point under discussion, Efik also has the central vowel [ɨ] and Njebi also has the central vowel [ə].

(36) *Feature definitions* (*SPE*: 394–395)

a. [±high]
High sounds are produced by raising the body of the tongue above the level that it occupies in the neutral position; non-high sounds are produced without such a raising of the tongue body.

b. [±low]
Low sounds are produced by lowering the body of the tongue below the level that it occupies in the neutral position; non-low sounds are produced without such a lowering of the body of the tongue.

Together, these features define three vowel heights, the possibility of [+high, +low] being ruled out by the definitions of the two features.

(37) *Defining vowel heights*

	High	Mid	Low
[high]	+	–	–
[low]	–	–	+

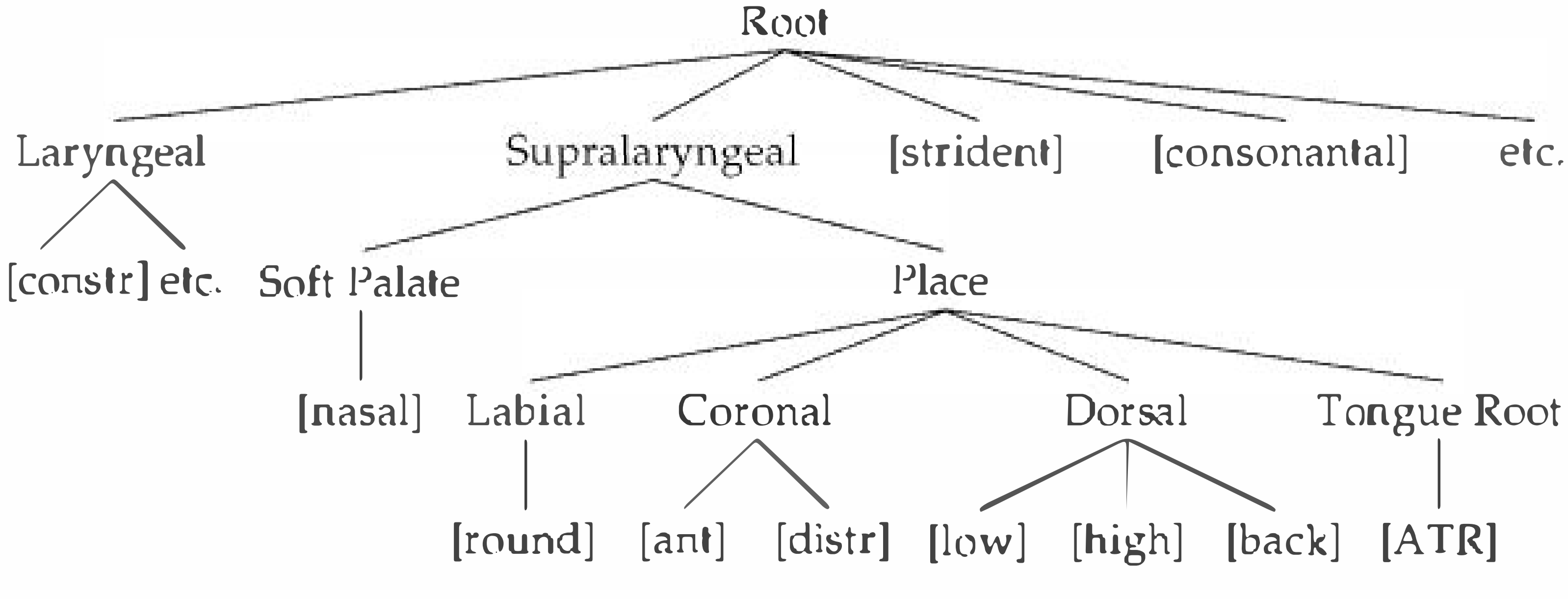
In defining a maximum of three vowel heights, this approach requires that additional, non-height features interact with [±high] and [±low] to define a full range of vowel contrasts. For example, [±tense] and [±covered] are proposed in *SPE*.

A particularly important point to note about this kind of theory (Parkinson 1996) is that vowel height plays no formal role: there is no class of "height" features in any formal sense. While it is possible within the *SPE* model for a rule to refer to the two features [±high] and [±low], such a pair has no formal status that is different from, say, [±high] and [±back], or [±high] and [±nasal]. Within that theory, distinctive features constitute an unstructured set, and any combination of features can constitute the set employed in a given rule.

5.3 *Feature geometry with [high] and [low]*

This lack of formal "height" features is a property that has continued to be adopted in numerous subsequent theories of vowel features. For example, in the influential work of Sagey (1986, 1990) on feature geometry, the features [±high] and [±low] continue to be assumed, assigned as two of the three features of the Dorsal articulator node.

(38) *The feature hierarchy in Sagey* (1990: 113)



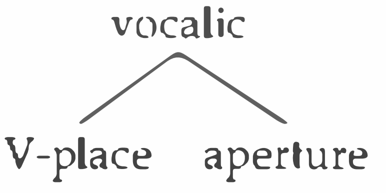
height 2	[e]	əje-é	'it's a spider'
	[ə]	əɲe-é	'it's me'[9]
	[o]	əwo-é	'it's you'
height 1	[ɛ]	ə̀ulɛ-ɛ́	'it's a weaver bird'
	[a]	àgbɛ̀-ɛ̀	'it's a load'[10]
	[ɔ]	ə̀sɔ́-ɛ́	'it's a horse'

If such assimilation is to be given "unit" status, then features for vowel height must be distinguished from vowel place, with a rule spreading specifically vowel height.

5.5 *Properties of a height class*

The response to arguments such as those presented in §5.4 has been to establish a vowel height node, labeled "aperture" by Clements (1990), which is independent of place (Clements 1990; Odden 1991; Wiswall 1991; Goad 1993; Elorrieta Puente 1996; Parkinson 1996; etc.). Typically, both place and height are considered to be daughters of a vocalic node. For example, Clements and Hume (1995) propose a vocalic node branching into place features ("V-place") and height features ("aperture").

(41) *Distinguishing place and height*



Of specific importance with respect to the characterization of vowel height, the featural content of the aperture/height node must be determined.

While the formal instantiations of this kind of proposal are relatively new, the conceptualization of speech sounds being composed of specifications for classes of features goes back to the International Phonetic Alphabet. As argued by Ladefoged and Halle (1988), the IPA constitutes a theory of segment structure, and in the IPA theory, vowel height is explicitly recognized as an important class. In the following sections, a sample of the proposals for formally characterizing the vowel height class are laid out.

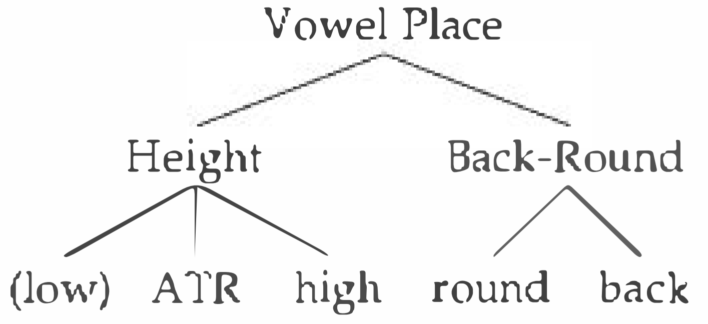
Ideally, a survey such as this one would carefully examine the arguments that have been presented for specific models, examining both successes and problems. Given the limited scope of this chapter, however, much of the discussion below will not go beyond a brief introduction. The reasons for this go beyond simple limitations on space. To adequately address the success of a given proposal, one needs to consider at least three things: (a) the validity of the empirical claims; (b) the predictions concerning natural classes, and how such predictions match the empirical generalizations; (c) how the theory of vowel height interacts with the overall theory of phonology. Examples of these considerations will be given below, but they will not be investigated in detail for each theory presented.

9 The stem vowel [ə] undergoes fronting by an independent rule (Clements 1990).
10 As with [ə], the stem vowel [a] undergoes fronting by an independent rule (Clements 1990).

5.5.1 *Grouping [high] and [ATR]*

Odden (1991) argues for grouping [ATR] with [high] under a "height" node.[11]

(42) *Height node*



As seen in Matumbi (§4.3.1), both the features [high] and [ATR] together are affected by height assimilation but not the feature(s) for backness and rounding; see the data in (16). On the basis of such data, Odden proposes that it is the "height" node that assimilates. In making the argument, Odden explicitly addresses the alternative possibility that assimilation is the combined result of assimilating [ATR] and [high] independently. He rejects such a "two-rule" analysis on two grounds. First, neither assimilation of [high] nor assimilation of [ATR] applies when the potential trigger is [ɛ] and the potential target is [u]; Odden suggests that this condition should be stated once, not twice. Second, morphemes that are exceptional to one type of assimilation are exceptional to both, again suggesting a unified analysis of the "two" processes.

The hierarchy in (42) is a natural development of the approach to vowel height in *SPE*, in that the features [low], [high], [back], and so on continue to be posited; see also Wiswall (1991) and Goad (1993). The distinction is that these features are assigned class status, with a bifurcation between the features defining height and those defining place.

A fundamental observation that can be made about this type of proposal is that vowel height as a class is defined by the interaction of formally independent sub-height features. Features like [high], [low], and [ATR] may function as three individual, independent features, or they may function together as the height class. In contrast with such an approach, numerous researchers have proposed that height should be viewed in a more integrated fashion, as will be discussed below. In discussing the various possible approaches that have been taken to gradient height, it is important to keep one central point in mind. Independent of one's precise proposals for vowel height, it is possible to consider [ATR] as being a height feature (and therefore integrated into the gradient height framework) or as being a place feature (and therefore independent of gradient height). The issue of tongue-root involvement is therefore a matter to be resolved in one way or the other for all proposals.

5.5.2 *Gradient height*

Amongst the early proposals for a gradient height feature are Ladefoged (1975) and Lindau (1978). A central argument in favor of gradience comes from chain shift cases such as those seen in §4.1.3, §4.2.3, and §4.7. Lindau (1978) discusses

[11] The feature [low] is parenthesized because Odden (1991) considers the evidence he presents to be inconclusive in this regard.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 245-306. Cambridge, MA & Oxford: Blackwell.

Crosswhite, Katherine. 2001. *Vowel reduction in Optimality Theory*. New York: Routledge.

Crothers, John. 1978. Typology and universals of vowel systems. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 2: *Phonology*, 93–152. Stanford: Stanford University Press.

Davy, J. I. M. & Derek Nurse. 1982. The synchronic forms of Dahl's Law, a dissimilation process, in Central and Lacustrine Bantu languages of Kenya. *Journal of African Languages and Linguistics* 4. 159–195.

Dettweiler, Stephen H. 2000. Vowel harmony and neutral vowels in c'Lela. *Journal of West African Languages* 28. 3–18.

Dumas, Denis. 1978. Phonologie des réductions vocaliques en français québécois. Ph.D. dissertation, Université de Montréal.

Dumas, Denis. 1981. Structure de la diphtongaison québécoise. *Canadian Journal of Linguistics* 26. 1–61.

Dumas, Denis. 1987. *Nos façons de parler*. Sillery: Presses de l'Université du Québec.

Elorrieta Puenta, Francisco Javier. 1996. A bipolar model of height in vowels and consonants. Ph.D. dissertation, University of Texas, Austin.

Fitzgerald, Colleen M. 2002. Vowel harmony in Buchan Scots English. *English Language and Linguistics* 6. 61–79.

Gick, Bryan, Douglas Pulleyblank, Fiona Campbell & Ngessimo Mutaka. 2006. Low vowels and transparency in Kinande vowel harmony. *Phonology* 23. 1–20.

Gnanadesikan, Amalia E. 1997. Phonology with ternary scales. Ph.D. dissertation, University of Massachusetts. (ROA-195.)

Goad, Heather. 1993. On the configuration of height features. Ph.D. dissertation, University of Southern California.

Guthrie, Malcolm. 1968. Notes on Nzɛbi (Gabon). *Journal of African Linguistics* 7. 101–129.

Hale, Kenneth. 1973. Deep-surface canonical disparities in relation to analysis and change: An Australian example. In Thomas A. Sebeok (ed.) *Current trends in linguistics*, vol. 11: *Diachronic, areal, and typological linguistics*, 401–458. The Hague: Mouton.

Halle, Morris. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26. 1–46.

Hayes, Bruce. 1990. Diphthongisation and coindexing. *Phonology* 7. 31–71.

Hualde, José Ignacio. 1989. Autosegmental and metrical spreading in the vowel-harmony systems of northwestern Spain. *Linguistics* 27. 773–805.

Hualde, José Ignacio. 1991. *Basque phonology*. London & New York: Routledge.

Hyman, Larry M. 1988. Underspecification and vowel height transfer in Esimbi. *Phonology* 5. 255–273.

Hyman, Larry M. 1999. The historical interpretation of vowel harmony in Bantu. In Jean-Marie Hombert & Larry M. Hyman (eds.) *Bantu historical linguistics: Theoretical and empirical perspectives*, 235–295. Stanford: CSLI.

Hyman, Larry M. 2001. Vowel harmony in Gunu. *Studies in African Linguistics* 30. 147–170.

Jones, Daniel. 1972. *An outline of English phonetics*. 9th edn. Cambridge: Heffer & Sons.

Kari, Ethelbert. 1997. *Degema*. Munich & Newcastle: Lincom Europa.

Kari, Ethelbert. 2003. Clitics in Degema: A meeting point of phonology, morphology, and syntax. Ph.D. dissertation, Tokyo University.

Kenstowicz, Michael & Charles W. Kisseberth. 1979. *Generative phonology: Description and theory*. New York: Academic Press.

Ladefoged, Peter. 1975. *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Ladefoged, Peter. 2006. *A course in phonetics*. 5th edn. Boston: Thomson Wadsworth.

Ladefoged, Peter & Morris Halle. 1988. Some major features of the International Phonetic Alphabet. *Language* 64. 577–582.

22 Consonantal Place of Articulation

Keren Rice

1 Introduction

An examination of the International Phonetic Alphabet chart (IPA 2007) yields a large number of consonantal places of articulation that are phonetically possible: bilabial, labio-dental, dental, alveolar, post-alveolar, retroflex, palatal, velar, uvular, pharyngeal, and glottal, as shown in Table 22.1.

Some identify even more places of articulation. Ladefoged and Maddieson (1996: 44), in a major work on sounds of the world's languages, present 17 places of articulation: bilabial, labio-dental, linguo-labial, interdental, laminal dental, laminal alveolar, lamino-post-alveolar (palato-alveolar), apical dental, apical alveolar, apical post-alveolar, sub-apical palatal (retroflex), palatal, velar, uvular, pharyngeal, epiglottal, and glottal. They divide these into five major target regions: labial (bilabial, labio-dental), coronal (laminal [linguo-labial, interdental, laminal dental, laminal alveolar, laminal post-alveolar], apical [apical dental, apical alveolar, apical post-alveolar], sub-apical [sub-apical palatal (retroflex)]), dorsal (palatal, velar, uvular), radical (pharyngeal, epiglottal), and laryngeal (glottal).

Table 22.1 The International Phonetic Alphabet (revised to 2005)

	bilabial	labio-dental	dental	alveolar	post-alveolar	retroflex	palatal	velar	uvular	pharyngeal	glottal
plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
trill	ʙ			r					ʀ		
tap/flap		ⱱ		ɾ		ɽ					
fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
lateral fricative				ɬ ɮ							
approximant		ʋ		ɹ		ɻ	j	ɰ			
lateral approximant				l		ɭ	ʎ	ʟ			

My goal in this chapter is to examine several issues surrounding consonantal place of articulation. I begin with an overview of the types of evidence used to justify the major place groupings, and then examine the evidence for subclasses within these and asymmetries in patterning between different places of articulation. I then review the features used to describe places of articulation. I end with a discussion of some additional issues relevant to the study of place of articulation.

2 The major places of articulation: A phonological perspective

The goal of a chart like that in Table 22.1 or a list like that of Ladefoged and Maddieson (1996) is to characterize locations where constriction is possible. From the perspective of phonology, the points of constriction group into classes, with sub-places within a class, based on natural class patterning (CHAPTER 17: DISTINCTIVE FEATURES). As discussed above, Ladefoged and Maddieson (1996: 44) identify five major places of articulation: labial, coronal, dorsal, radical, and laryngeal. Other theories note a similar small number of major places. For instance, Articulator Theory and Revised Articulator Theory (e.g. Sagey 1990; Halle *et al.* 2000: §4) distinguish labial, coronal (tongue blade), dorsal (tongue body), and tongue root or radical. Element theory distinguishes labials, palatals, coronals, and velars (e.g. Harris and Lindsey 1995). While there are differences between these theories, they agree on the existence of major zones, and I begin with evidence for these.

In the following discussion I examine types of evidence for the division into labial, coronal, dorsal, radical (also called tongue root and pharyngeal), and laryngeal places of articulation, as well as the sub-places within each. Each section is organized as follows. I first present a phonetic description of the class. I next examine phonological evidence for the class, including distribution, harmony patterns, co-occurrence restrictions, and other phonological processes. I then introduce sub-places within the major place, asking if what are categorized as place distinctions are best analyzed as such from a phonological perspective.

Before turning to evidence for the individual features, I review one argument for places of articulation in the oral cavity being divided into labial, coronal, and dorsal, based on the major articulator involved. Sagey (1990), in an important work on place, presents an argument for this division based on complex articulations (CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION). In a cross-linguistic survey, she finds that only a limited number of complex articulations occur. There are labial-velars (e.g. [kp]), probably the best studied. In addition, there are labial-coronals (e.g. [pt]), coronal-velars (e.g. [ǁ]), and labial-coronal-velars (tkw). Unattested are, for instance, complex articulations consisting of two labial type articulations or two coronal type articulations. The possible place combinations found, Sagey argues, present an argument for these three major places of articulation.

2.1 *Labial place of articulation*

Labial includes bilabial, articulated with the upper and lower lips, and labio-dental, articulated with the tip of the tongue and the lower lip; in addition, Ladefoged and Maddieson (1996: 44) include a linguo-labial place of articulation, articulated with the upper lip and the tongue, but they treat this as a type of coronal. This sound has not figured in the phonological literature, and I do not discuss it here.

(5) a. *Rotokas (North Bougainville)*

alveolar
t

b. *Yareba (Papuan)*

alveolar
t d dz s

c. *Finnish (Uralic)*

dental	alveolar
t̪ (d̪)	s

d. *Hungarian (Uralic)*

dental	palato-alveolar	palatal
t̪ d̪ t̪s̪ d̪z̪ s̪ z̪	ʧ ʤ ʃ ʒ	cç ɟʝ

e. *Pashto (Indo-Iranian)*

dental/alveolar	palato-alveolar	retroflex
t d ts dz s z	ʧ ʤ ʃ ʒ	ʈ ɖ ʂ ʐ

f. *Western Desert (Pama-Nyungan)*

dental	alveolar	retroflex
t̪	t	ʈ

g. *Nunggubuyu (Australian)*

dental	alveolar	palatal	retroflex
t̪	t	c	ʈ

h. *Malayalam (Dravidian) voiceless stops/affricates*

dental	alveolar	palato-alveolar/ palatal	retroflex
t̪ d̪	t	ʧ ʤ	ʈ ɖ

i. *Kalasha (Indo-Aryan; only voiceless stops/affricates shown)* (Arsenault and Kochetov, forthcoming)

dental	palatal	retroflex
t ts	tɕ	ʈ ʈʂ

The above are representative of the coronal obstruent inventories found cross-linguistically, from Rotokas with a single coronal obstruent to the coronal inventories of languages such as Pashto, Malayalam, and Kalasha, with affricates and fricatives as well as stops.

2.2.2 *Evidence for coronals as a phonological class*

There is considerable evidence that coronals pattern as a class. I summarize some below.

A well-known argument for the class of coronals involves co-occurrence restrictions. Hall (1997: 5), following Davis and Hammond (1995: 163–164), argues that there are co-occurrence restrictions in most varieties of American English against the sequence coronal + [j] + vowel.

(6) p b m f v k g + j + vowel
*θ ð t d s z ʃ ʒ ʧ ʤ n l r + j + vowel

The prohibition includes all English coronal consonants, and holds of syllable-initial clusters. Thus, clusters with /j/ are not allowed if the first segment is coronal, irrespective of its sub-place of articulation; they are otherwise.

Dixon (1980) discusses a variety of types of evidence from Australian languages for grouping coronal consonants as a class. One argument is based on neutralization and variation amongst coronals (CHAPTER 80: MERGERS AND NEUTRALIZATION). Typical Australian languages show a four-way coronal contrast, subdivided into apicals and laminals, with each further subdivided. Dixon (1980) points out that the coronal subclasses in some Banjalang dialects, apical and laminal, neutralize intervocalically; the place contrast is maintained elsewhere and other places of articulation are not involved. There is often free variation between apical and laminal articulations for the stop, with the choice differing by dialect.

Hamilton (1993) offers evidence from Australian languages for the class of coronals based on phonological patterning. For instance, coronals provide an environment for a phonological process in Nunggubuyu (Heath 1984: 69–71), where a palatal deletes in a derived environment before a coronal; with a following labial or velar, deletion does not occur. In Walmatjari (Hudson 1978: 11–12), there are suffixes beginning with a retroflex lateral; following a coronal, the initial consonant of the suffix assimilates in place of articulation to the coronal; following a non-coronal, other changes occur, but the consonant does not assimilate.

Moving away from Australia, there are languages with co-occurrence restrictions on different coronal places of articulation. In Päri (Western Nilotic), dental and alveolar stops and nasals contrast, and do not co-occur within a root (Andersen 1988; Hansson 2010). This harmony is actively enforced in suffixed forms. In another Nilotic language, Dholuo, the co-occurrence restriction holds of the dental and alveolar stops; there is a single nasal that does not participate. Pohnepian (Austronesian; Rehg and Sohl 1981) has contrastive dental and retroflex stops; these do not co-occur within a root.

Assimilation provides evidence for the coronal class, as in the Sanskrit *nati* process. Basically, a retroflex triggers retroflexion of a following /n/. Retroflexion is transparent with respect to intervening labial and dorsal consonants, but is blocked if a coronal of any place of articulation (dental, palatal, retroflex) intervenes; see Hansson (2010).

(7)	a.	*non-coronal consonants between trigger and target*		
		iʂ-naː	iʂɳaː-	'seek (PRESENT STEM)'
		vrk-na	vrkɳa-	'cut up (PASSIVE PARTICIPLE)'
		kʂubh-aːna	kʂubhaːɳa-	'quake (MIDDLE PARTICIPLE)'
	b.	*coronal consonant between trigger and target*		
		mrd-naː	mrdnaː-	'be gracious (PRESENT STEM)'
		maɪj-aːna-	maɪjaːna-	'wipe (MIDDLE PARTICIPLE)'

In addition, in assimilation in Sanskrit a dental assimilates to an adjacent coronal, but not to other places of articulation; see §4.

There is thus clear evidence from several sources that in languages with more than one coronal place of articulation these can pattern as a class.

While in some cases evidence exists that all coronals of a language are in a single class – for instance, English (6) – the class of coronals may pattern together, but only within a manner class. An example comes from Nilotic languages (Mackenzie 2009; Hansson 2010), where harmony occurs between coronal stops and nasals, or just stops, depending on the language, but liquids and continuants do not enter into the harmony.

Stratification by manner exists in the consonantal root consonant co-occurrence restrictions in many Semitic languages, where sounds within a class are dispreferred within a root. While coronal sonorants and stops, for instance, co-occur, coronal fricatives rarely co-occur within a root.

(8) *Arabic coronal co-occurrence restrictions* (Rose 1996: 75)

coronal sonorants	l r n
coronal stops	t d (plus emphatic counterparts)
coronal fricatives	ʃ ʒ s z θ ð (plus emphatic counterparts)

See Coetzee and Pater (2008) for discussion of place/manner effects in Muna.

Phonological evidence thus exists for the class of coronals as a whole; in some cases the patterning of coronals is stratified by manner.

2.2.3 *Coronal sub-places of articulation*

Because coronals allow several subclasses, the internal structure of the coronal region has received detailed attention. For this we look to languages with four distinctive places of articulation within coronals, where the distinction commonly called apical/laminal receives wide support. Evans (1995: 727) notes that "There is an enormous amount of evidence – from phonotactics, morphophonemics, diachronic changes, and synchronic variation for these groupings," referring to apical (alveolar, retroflex) and laminal (dental, palatal) in the coronals of languages of Australia. Hamilton (1993) presents a variety of kinds of evidence for this subdivision.

Allophony presents one type of evidence. If there is a single apical and/or a single laminal place, allophonic variation may exist within it. For instance, in single laminal languages such as Watjarri (Douglas 1981: 203–204), alternations or variation between alveo-palatal and dental articulations occur, often conditioned by the following vowel. Some languages show variation in apical articulations, with non-contrastive alternation between alveolar and retroflex articulations in different vocalic environments; Wargamay is an example (Dixon 1980: 155–156).

Neutralization is a second type of evidence. In many Australian languages the two apical articulations neutralize to a single non-contrastive series word-initially, often symbolized as a retroflex but sometimes as an alveolar; Hamilton (1993: 32) remarks that only a few languages show an apical contrast word-initially. In some languages the laminal articulations neutralize syllable-finally, generally reported as palatal; Hamilton (1993: 33) notes that only a few languages exhibit this contrast syllable-finally.

Phonotactics treat apicals and laminals as natural classes: apical consonants are permitted in certain environments to the exclusion of laminals, and vice versa. For instance, in some languages both apicals occur as the initial member of a cluster, while laminals are not permitted in this position (e.g. Kalkatungu; Blake 1979); in other languages both laminals occur as the second member of a heterorganic cluster, while apicals are not allowed (e.g. Nunggubuyu; Heath 1984).

Phonological processes can reference apical or laminal. For example, Mara (Heath 1981) has a dummy syllable in certain cases between a prefix and a root when the root begins with an apical sonorant; it does not occur with other places of articulation, including laminals. In some languages a lenis laminal stop becomes a glide intervocalically; apicals are not affected (e.g. Djapu; Morphy 1983).

Local harmony occurs within the apical set and within the laminal set in some languages. In particular, clusters of heterorganic apical clusters and heterorganic laminal clusters are not permitted in some languages, while clusters of an apical and a laminal are allowed; Dhuwala-Dhuwal (Morphy 1983) is an example.

Dravidian languages also exhibit an apical/laminal contrast, as argued by Arsenault (2008). For instance, word-initial apicals, both alveolar and retroflex, are dispreferred in these languages, with only laminals occurring.

Serbian provides evidence for apical and laminal classes, and for cross-classification between them; see Morén (2006) and Radišić (2009). Serbian has coronal stops/affricates as follows (only voiceless stops illustrated).

(9) t ts ʧ tŝ

Evidence for constituency within the coronals comes from several processes. A process called iotation groups together the first and last of these as opposed to the other two: /t tŝ/ *vs.* /ts ʧ/. Mid-vowel fronting provides evidence for /ʧ tŝ/ as a class, with /ɔ/ fronted to [ɛ] in the environment of these places of articulation. Assuming that the first and fourth columns represent laminal articulations and the middle two apical articulations, based on descriptions of the sounds, apical/laminal provides the classes involved in iotation. The further back of the apicals and laminals (last two columns) provide the environment for mid-vowel fronting. Thus, while the Australian languages do not appear to provide evidence for cross-classification within the apical and laminal subgroups, Serbian does provide for such classification.

While the division of coronals into apical and laminal receives support from languages with four coronal sub-places, other divisions appear to be possible. In Tahltan (Athabaskan), there are four coronal places of articulation (Shaw 1991; also §6.1); only voiceless stops and affricates are indicated.

(10) t tθ ts ʧ

The latter three enter into harmony, with the plain /t/ excluded. Assuming that the affricates are stops in Tahltan, with the consonants in (10) distinguished solely by place of articulation, as Shaw argues, then the phonology does not appear to support a primary apical/laminal distinction.

There may be asymmetries between the number of coronal places of articulation available at different manners of articulation. Polish obstruents offer an interesting example. Voiceless symbols are shown; all have voiced counterparts.

(11)	t	ts	tʂ	tɕ
		s	ʂ	ɕ

The stops/affricates and fricatives at a place of articulation do not always pattern together phonologically. For instance, /t/ and /s/ undergo palatalization, but /ts/ (and other stop/affricates and fricatives) does not. In second velar palatalization, /k/ shifts to [ts], while [x] shifts to [ʂ]. Similarly, [t] and [s] both occur with high vowels [i] and [ɨ], while other coronal obstruents are restricted, with [ts ʂ tʂ]

(12)	t	tʰ	tʼ	alveolar
	s			alveolar
	s̪		s̪ʼ	dental
	sʷ		sʷʼ	labialized alveolar
	ʂ	ʂʰ		retroflex
	ɕ	ɕʰ		alveolopalatal
	ɬ		ɬʼ	lateral

In addition, it has palatals.

(13)	ç	çʰ		palatal

Hall, following Colarusso, argues that palatal fricatives are not coronal, based on phonotactic generalizations about consonant clusters. In a cluster, the first consonant can be a labial, a voiceless uvular fricative, or a voiceless coronal stop or fricative. However, clusters cannot begin with a palatal fricative. As Hall (1997: 16) notes, this would be surprising if the palatal fricatives were coronal, since all other coronal stops and fricatives occur as the first consonant in a two consonant cluster. This argument perhaps establishes the status of the palatal as a non-coronal, but does not establish what it is; dorsal is a logical choice.

2.3.2 *Evidence for dorsals as a phonological class*

Considerable evidence exists for dorsals (velars, uvulars, palatal fricatives) as a class from allophonic patterning, co-occurrence restrictions, and harmony.

Velars, uvulars, and palatal fricatives enter into allophonic relations (CHAPTER 11: THE PHONEME). Hall (1997) notes that in German, the fricatives [ç x χ] are in complementary distribution, suggesting a common feature: they are dorsals, varying by sub-place.

Co-occurrence restrictions present another type of evidence for a single class. Rose (1996), after Bessell (1992), shows that Interior Salish has co-occurrence restrictions on place and manner; the place restrictions are summarized in (14). Identical places of articulation are disallowed within a manner; in addition, velar and uvular consonants fail to co-occur.

(14) *labial V labial
*coronal V coronal
*velar V velar
*uvular V uvular
*laryngeal V laryngeal
also
*velar V uvular
*uvular V velar

Thus the velars and the uvulars interact in excluding one another.

In Arabic, co-occurrence restrictions hold of dorsals just as they do of labials and coronals, dispreferring the co-occurrence of velar and uvular stops [g k q]. Note also the co-occurrence restrictions on gutturals, discussed below.

Dorsal harmony, involving velars and uvulars, provides evidence for velars and uvulars forming a class. Hansson (2010) discusses several cases. In Misantla

Totonac (Totonacan), velar and uvular stops harmonize. The primary facts are given below; see MacKay (1999) and Hansson for details.

(15) *Misantla Totonac* (MacKay 1999)

/ut maka-ʃqat/	'ʔut ma'qaʃ'qɛt	's/he scratches X (with hand)'
/ut maka-paʃ/	'ʔut maka'paʃ	's/he bathes his/her hand'

A similar harmony occurs in the related Tlachichilco Tepehua.

(16) *Tlachichilco Tepehua* (Watters 1988)

/ʔuks-laqts'in/	ʔoqslaqts'in	'look at Y across surface'
/ʔuks-k'atsa:/	ʔuksk'atsa:	'feel, experience sensation'

Hansson (2010) argues that Bolivian Aymara (Aymaran) has dorsal harmony. The language has plain, aspirated, and glottalized stops at velar and uvular places of articulation with restrictions on phonation type as well as place; I abstract away from the former. While velars co-occur and uvulars co-occur, a velar and a uvular do not.

(17) *Bolivian Aymara* (de Lucca 1987)

 q-q, qh-qh, q'-q'
 k-k, kh-kh, k'-k', k'-k
*k-q, q-k (any laryngeal properties)

Hansson notes that dorsal consonant harmony is cross-linguistically rare. In order to ascertain whether this is true, it is necessary to examine the occurrence of dorsal harmony in languages with contrastive velars and uvulars.

Given the debate about whether the bilabial/labio-dental distinction involves place or manner (§2.1), one might ask if the velar/uvular distinction always involves place. Some processes point to the conclusion that place is the primary dimension of contrast in languages with this distinction. In spirantization processes, just as /p/ spirantizes to [f], /k/ spirantizes to [x], and /q/ to [χ]. In the Athabaskan D-effect, /x/ hardens to [k] and /χ/ to [q] (e.g. Ahtna). This process shows the importance of place of articulation in distinguishing velars and uvulars.

When there is not a contrast between a velar and uvular within a manner, an inventory with velar stops and uvular fricatives is sometimes found. For instance Welsh (Celtic; Ball and Müller 1992) has voiced and voiceless velar stops, [k g], and a voiceless uvular fricative, [χ]. Welsh has mutations (see CHAPTER 117: CELTIC MUTATIONS); in Soft Mutation, /k/ becomes [g], while in Aspirate Mutation, /k/ spirantizes, becoming the voiceless uvular fricative [χ]. This parallels mutation at other places of articulation: under Soft Mutation, /p/ becomes [b] and /t/ becomes [d]; under Aspirate Mutation, /p/ becomes [f] and /t/ mutates to [θ]. Basically, mutation triggers a manner change, maintaining place. Thus, while the stop and fricative differ in place and manner phonetically, phonologically they differ in manner, and the organization in (18) is appropriate:

(18) dorsal stop k g
 dorsal fricative χ

One can ask if a different phonological organization is possible, with a primary place difference realized as a manner distinction, as in (19).

(19)	velar	uvular
	k	x

If, for instance, a language existed in which there were co-occurrence restrictions on obstruents of like place, and [k]–[x] combinations were allowed, we might conclude that the [x] patterned as a distinct place of articulation, realized phonetically as a spirant.

To summarize, dorsal includes velars, uvulars, and likely palatal fricatives. Evidence for this class is based on allophony, co-occurrence restrictions, harmony, and other phonological processes.

2.3.3 *Laryngeals*

I now consider laryngeal consonants. There is considerable discussion of laryngeal consonants in the literature, where they are accorded two treatments. They are often considered to lack a place of articulation (e.g. Steriade 1987), and they are also considered to be a type of pharyngeal (e.g. Lombardi 2002). In this section I present evidence for placelessness; evidence for their pharyngeal nature is given in §2.3.4.

The primary argument for the laryngeal class, glottal stop, and [h] being placeless comes from laryngeal transparency. The following evidence, from Kashaya (Pomoan), is from Buckley (1994), summarized by Rose (1996: 100–101). Within a morpheme, vowels are identical across a laryngeal consonant (20a), while there are no general co-occurrence restrictions on vowels (20b).

(20)	a.	s'iʔi	'flesh'	he'ʔen	'how'	inaʔa	'food, eat'
		ni'hin	'to oneself'	behe	'bay nut'	juhu	'pinole'
		ʔaha	'mouth'				
	b.	du'weʔ	'yesterday'	n'uqa:ʧ	'get lost'	maʧe-	'hold with foot'

This is an argument for laryngeal placelessness, under the assumption that specified places of articulation block assimilation. If laryngeals lack a specified place of articulation, vowel features can harmonize across them.

Laryngeals can pattern as a class. There may be co-occurrence restrictions, as in Interior Salish (14). In addition, laryngeals often result from obstruent debuccalization, with stops neutralizing to [ʔ] and fricatives to [h], as in the historical development of Kelantan Malay (Trigo 1988).

(21)	*Standard Malay*	*Kelantan Malay*	
	ʔasap	ʔasaʔ	'smoke'
	kilat	kilaʔ	'lightning'
	balas	balah	'finish'

2.3.4 *Pharyngeal (radical, tongue root)*

The final articulation region identified by Ladefoged and Maddieson (1996: 37), radical, includes pharyngeal and epiglottal sounds, i.e. sounds articulated in the region below the uvula.

Transguttural vowel harmony occurs in some languages, illustrated with Iraqw (Cushitic).

(27) *Iraqw* (van der Hulst and Mous 1992: 103–104; Mous 1993: 37; Rose 1996: 77)

no harmony across non-gutturals			*harmony across gutturals*		
hamaatl-iim	[hamtliim]	'wash'	buuʔ-iim	[buʔuum]	'harvest pay'
baal-iim	[baaliim]	'defeat'	ufaaħ-iim	[ufaħaam]	'blow'

In D'opaasunte (Eastern Cushitic), a vowel is lower when the preceding consonant is guttural, ([χ ɢ ʕ' ħ ʔ h]); it is higher after other consonants (Hayward and Hayward 1989). Hayward and Hayward use "A," an archiphoneme, to represent this vowel.

(28) *D'opaasunte* (Hayward and Hayward 1989: 183–184)

ʔan ˈkod[ɛ]j	'I am working'	ʔan ˈʧ'oχ[a]j	'I am milking'
ʔan ħull[ɛ]j	'I am entering'	ʔan fAʔ[a]j	'I am loading'

In Standard Somali (Cushitic), a suffix has the form [-d] after a guttural ([q χ ʕ ħ ʔ g]) and [-t] elsewhere (Hayward and Hayward 1989: 184).

Laryngeals thus pattern as if they were placeless in some languages and as if they were guttural in others. Rose asks whether it is accidental how they pattern, or if there is a systematic way to determine this. She argues that laryngeals pattern with pharyngeals when pharyngeals or uvular continuants are present in a system; otherwise they pattern as if they lacked a specified place.

2.3.5 *Summary*

Dorsal consonants (velar and uvular) interact in many languages, and they also interact with further back consonants in some languages. As with labials and coronals, many questions remain, and I conclude with one. The sound commonly written with the symbol /ŋ/ is ambiguous in its patterning, in some languages patterning as a velar and in some as if it were placeless or laryngeal (e.g. Trigo 1988; Rice 1996; de Lacy 2006). Is this an appropriate analysis? Is it only the manner that represents these two places of articulation, perhaps better written as /ŋ/ and /N/, or can other manners of articulation also show this kind of ambiguity between patterning as a laryngeal and patterning as a velar?

2.4 Conclusion

In this section I have examined phonological evidence for the major places of articulation: labial, coronal, dorsal, radical, and laryngeal. In all cases, there is clear evidence for the class, and, at the same time, complexities involving such things as stratification by manner and interactions between place classes require further study.

3 Further constituency

We have seen evidence for five major places of articulation, three in the oral cavity and two further back. While there is general acceptance of this division in the literature, debate exists about whether these major places of articulation enter into relationships with one another, or, to put this another way, whether there is constituency among places of articulation (see CHAPTER 27: THE ORGANIZATION OF FEATURES). One might imagine that there is no relationship between them, with a flat structure. Alternatively, one might imagine that some of the places of articulation are more closely linked to one another than others are, with a constituent structure. I examine this with respect to labial, coronal, and dorsal, where two possible groupings are proposed. One involves grouping labials and dorsals to the exclusion of coronals, under a feature [grave] or [peripheral]. The other involves grouping coronals and dorsals to the exclusion of labials; the feature has been called [lingual]. One might imagine a grouping of coronals and labials to the exclusion of velars, a proposal that has not received support; see §5 on the controversy about the feature [anterior], which groups labials and front coronals.

3.1 *Oral and pharyngeal*

Before turning to these proposals, I briefly review proposals for the overall organization of place. McCarthy (1994: 223) argues within feature geometry for the bifurcation of place into two major constituents, which he terms Oral and Pharyngeal. Oral dominates Labial, Coronal, and Dorsal. Others have refined this proposal to, for instance, allow [RTR] as a dependent of Pharyngeal; see Rose (1996) for discussion of how to accommodate the guttural class.

3.2 *Grouping labials and dorsals*

Evidence is found in the literature for grouping labials and dorsals; see Hall (1997) for detailed references, including Jakobson *et al.* (1952), Jakobson and Halle (1956), Hyman (1973), Campbell (1974), Odden (1978), and Rice (1994). Hall gives the following evidence for labials and velars (and/or uvulars) patterning together, among others.

In Yurok (Algic), the 3rd person prefix is [ʔu] before non-coronals (labials, velars, labial-velars) and [ʔwe] before coronals. In Lhasa Tibetan the consonants [p k q] spirantize and voice intervocalically, while coronals do not. A vowel shift occurred in the history of Korean before labials and velars but not before dentals and alveo-palatals.

Many of the arguments presented by Hall are diachronic. There are also synchronic arguments for grouping labial and velar. One of the best-known arguments comes from Korean, where, in some speech forms, the coronal stop and nasal assimilate in place to a following adjacent consonant, and the labial stop and nasal assimilate to a following velar; the velar stop and nasal do not assimilate. If labial and velar are grouped together, their patterning can be understood: coronals assimilate to these places, identified as grave. In (29), a place of articulation assimilates to those to its right, but not vice versa. Thus a coronal (T) assimilates to both a labial (P) and a velar (K), while a labial assimilates only to a velar, and a velar does not assimilate.

(29) [T [P K]]

Lombardi (1996) identifies cases that suggest the need for grouping labials and velars (her [–coronal]) in post-lexical rules. For instance, between word sequences in Pohnepian (Rehg and Sohl 1981; Rehg 1984), a labial is realized as a labial nasal when followed by a labial, and a velar is realized as a velar nasal before another velar.

(30) *labial + labial*
e kala/p p/aan soupisek 'he'll always be busy'
e kala[m p]aan soupisek
velar + velar
e sai/k k/eŋwini 'he hasn't yet taken medicine'
e sai[ŋ k]eŋwini

However, with a sequence of coronals, the fiirst does not become a nasal.

(31) e mei/t t/aŋaaŋa 'Aren't you lazy!'
*e mei[n t]aŋaaŋa

Thus evidence for grouping labials and dorsals is both diachronic and synchronic.

3.3 Grouping coronals and dorsals

Arguments for grouping coronals and velars as linguals are also found. Rubach (1993), followed by Lombardi (1996), adduces evidence from Slovak (Slavic), where /æ/ backs to [a] post-lexically in an environment that Rubach calls [−labial]. There is a contrast between these vowels following a labial (32), but not elsewhere.

(32) masa 'mass'
mæso 'meat'

The diminutive suffix illustrates the alternation between [a] and [æ]. The suffix is a front vowel after a labial (33a), following a non-labial it is [a], with palatalization of the preceding consonant (33b, 33c).

(33)	a.	holub	'pigeon'	holub-æ	'young pigeon'
	b.	had	'reptile'	haad'a	'young reptile'
	c.	vnuk	'grandchild'	vnuutʃa	'young grandchild'

Rubach argues that the suffix is underlyingly front; it triggers palatalization and is backed following a coronal and a velar (see CHAPTER 121: SLAVIC PALATALIZATION).

The facts are complicated by the presence of a diphthongization process where /æ/ becomes [ia]. The forms in (34) show lengthening in the genitive plural, with /a/ lengthening to [aa].

(34) lano 'strong' laan (GEN PL)
pivo 'beer' pŭv (GEN PL)

While generally vowels lengthen in the genitive plural, diphthongization occurs in some environments, including with /æ/, which diphthongizes to [ia].

(35) hovædo 'beast' hoviad (GEN PL)
mæso 'meat' mias (GEN PL)

Many Slavic languages exhibit a third pattern: the velar can undergo a shift in place of articulation, resulting in a coronal. This is exemplified for Serbian in (45).

(45) a. *Serbian: First palatalization* (Radišić 2009)
ru/k/a + itsa   ru[ʧ]itsa   'hand'
pra/x/ + iti   pra[ʃ]iti   'dust'
b. *Serbian: Second palatalization* (Radišić 2009)
ɔra/x/ + i   ɔra[s]i   'nut'
pɛ/k/ + i   pɛ[ts]i   'to bake'

Coronals also shift their place of articulation in iotation, but remain coronal.

Thus, while in languages such as Catalan coronals shift to another major place of articulation in assimilation, in Sanskrit coronals are assimilation targets, but only assimilate to other coronals, and in many Slavic languages, coronals shift to other coronals and velars shift to coronals. To use the terminology of phonological markedness, based on such shifts we might identify the velar as unmarked in languages like Serbian, and the coronal as unmarked in languages like Catalan and Korean. Given the nature of the coronal inventories in languages like Sanskrit and Serbian, with more than one coronal sub-place allowed in assimilator position, one might conclude based on target asymmetries, as did Avery and Rice (1989), that coronals are unmarked generally, attributing their patterning in languages with more than one coronal sub-place to the fact that there is more than one sub-place. The existence of more than one coronal sub-place indeed appears to affect the patterning of coronals, and in the following discussion I set these languages aside and examine assimilation and neutralization in the absence of a rich coronal inventory, beginning with neutralization. I consider languages with labial, coronal, and dorsal places of articulation.

4.2.2 *Neutralization, epenthesis, and the emergence of the unmarked*

Neutralization is considered to yield the unmarked (CHAPTER 80: MERGERS AND NEUTRALIZATION). Languages with word-final neutralization are shown in (46). These languages exhibit passive neutralization, with morphotactics that disallow other places of articulation for stops word-finally. In addition to coronal stops, languages exist where only labial stops occur in this position, or where only dorsal stops are allowed in this position.

(46) a. languages allowing only a coronal stop: Saami, Finnish
b. languages allowing only a labial stop: Nimboran (Papuan; Anceaux 1965), Basari (Niger-Congo; Abbott and Cox 1966)
c. languages allowing only a dorsal stop: Quichua (Quechua; Orr 1962)

While there are statistical differences, with coronals (and laryngeals) more common as the only place of articulation allowed when no contrasts exist in stops word-finally, overall, the range of places of articulation occurs cross-linguistically in this position. If neutralization is a diagnostic for unmarkedness, and if coronal is universally unmarked, these patterns are unexpected. Epenthesis mirrors neutralization, with at least laryngeal, coronal, and dorsal consonants serving as epenthetic. While de Lacy (2006) and de Lacy and Kingston (2006) argue that dorsal

The sound pattern of English (*SPE*; Chomsky and Halle 1968) features have encountered serious criticism on both empirical and conceptual grounds. With respect to place, one challenge is the inability to define a class of labial consonants and round vowels (e.g. Campbell 1974). The feature [anterior] is also problematic, since anterior sounds in the *SPE* sense do not pattern as a natural class (e.g. Dixon 1980; Gnanadesikan 1994).

Sagey (1990), building on work by Halle (e.g. 1983), argues for four major places of articulation – Labial, Coronal, Dorsal, and Tongue Root (Radical) – expressed by monovalent features, dominating binary features. Building on Sagey, Halle, and others, Hall (2007: 332) gives a chart of obstruent places of articulation, which is shown in Table 22.2. I follow him in separating stops and fricatives, and show voiceless consonants only. The table perhaps represents an overall North American consensus about place of articulation features. Nevertheless, many issues remain; I list a few in (47).

(47) a. What distinguishes bilabials and labio-dentals? Hall leaves this open, suggesting that a feature [labio-dental] might be required.
b. How are palatals represented? Are all palatals coronal (see §2.2; Hall 1997)? Do all palatal sounds have the same representation? Similar questions can be raised for retroflexes.
c. Are [anterior] and [distributed] appropriate to distinguish coronal sub-places? In particular, the role of [anterior] remains controversial.
d. How are gutturals distinguished featurally?
e. Do features group into classes? In particular, do lingual and/or grave classes exist?
f. Is a mix of monovalent and bivalent features appropriate?

Table 22.2 Features for obstruent places of articulation (after Hall 2007). A check (✓) in a cell indicates the presence of a unary feature; in rows with checks, the absence of a check indicates the absence of this feature. In rows with + and −, absence of any mark indicates that the feature is not relevant for that sound

	p	t̪	t	ʈ	c	k	q	ts	tʃ	ɸ	f	θ	s̪	s	ʂ	ʃ	ç	x	χ	ħ
continuant	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+
labial	✓									✓	✓									
round	−									−	−									
coronal		✓	✓	✓	✓			✓	✓			✓	✓	✓	✓	✓	✓			
anterior		+	+	−	−			+	−			+	+	+	−	−	−			
distributed		+	−	−	+			+	−			+	+	−	−	+	+			
dorsal						✓	✓											✓	✓	
back						+	+											+	+	
low						−	−											−	−	
high						+	−											+	−	
pharyngeal																				✓
strident								+	+	−	−	−	+	+	+	+	−	−	−	−

(49) q χ
k s
ts s
tɬ ɬ

The spirantization of the front velar to [s] is surprising, but otherwise a stop/affricate is realized as a fricative of the same place of articulation. If lateral represents a place of articulation, this pattern is unsurprising. Note also the development of the Proto-Athabaskan *ts series (*ts *ts^h *ts' *s *z) to a /tl/ series ([tɬ tɬ^h tɬ' ɬ l]) in Koyukon (Jetté and Jones 2000), a development which can be explained as a shift in place of articulation of the series.

6.2 *Consonant–vowel interactions*

I have not discussed consonant–vowel interactions, since this chapter concerns consonantal place; see CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS. Flemming (2002) also addresses consonant–vowel interactions and the types of features required. To give a few examples, vowel-triggered palatalization can involve the addition of a secondary articulation; it can also involve a shift in primary place, with perhaps the most dramatic being the shift from a velar to a coronal, found in many Slavic languages; e.g. (45). There is often interaction between low vowels and dorsal sub-places, with low vowels conditioning uvulars (e.g. Xibe (Tungusic); B. Li 1996) or vice versa (e.g. Totonac; MacKay 1994). Low vowels or retracted tongue root vowels often interact with pharyngeal consonants; e.g. (26). Round vowels and labial consonants can pattern together, as in Igbo (§2.1). In addition, there are interactions between back vowels and retroflex consonants. Vowels and consonants may pattern together in natural classes as well, for instance, in conditioning rules.

There has been considerable debate as to how to handle such interactions. Clements (1991) argues for a unified theory of consonant and vowel place, as do element-based theories (see Harris and Lindsey 1995 for an overview), while others argue for distinct place features (e.g. Padgett 2002) for consonants and vowels.

6.3 *Evidence for innate features*

An important premise since distinctive features were introduced is that features are innate, with a small universal set. Much research has been directed at defining what set and identifying its phonetic foundations, with the understanding that an appropriate and complete set of features would allow an account of cross-linguistic sound patterning.

I have pointed out in various places that what is characterized as the same sound from a phonetic perspective can show different phonological patterning. To some degree this is dependent on the sound system of the language, a point that has long been observed, as the following quote from Trubetzkoy shows.

> The ambiguous character of lateral articulation, which causes such difficulties in *phonetic* systematization, is something that can quite easily be resolved in *phonological* systemization, the more so since the important thing here is only to establish to which other phoneme the particular "lateral" phoneme stands in a relation of opposition, and to determine the nature of such an oppositive relationship. (Trubetzkoy 1969: 140)

In the discussion of evidence for places of articulation, we have seen debate about how to classify particular sounds. For instance, what are phonetically two places of articulation might pattern as one, as with [p] and [f], where [p] is bilabial and [f] labio-dental, but they enter into co-occurrence restrictions, for instance, as a single place of articulation. Similarly, [k] and [χ] can form a pair in spirantization, despite the fact that one is velar and the other uvular.

In addition, what looks like a single place of articulation might pattern as two distinct places. This is dramatic in Polish (11), where the retroflex stop and the retroflex fricative differ in patterning.

Furthermore, a single place of articulation may be classified in two groups. For instance, there is evidence from co-occurrence restrictions that uvulars are both dorsal and radical, with stops patterning with velars and fricatives with radicals. Palatals appear to be divided between dorsal (fricatives) and coronal (other manners).

Cross-linguistically, coronals do not appear to be divided in the same way in all languages. While in many languages there is evidence for segregating coronal inventories into two main places of articulation, often characterized as apical and laminal, there are languages with similar contrasts, but this categorization does not seem to be appropriate.

Such variation in cross-linguistic patterning might be responded to in different ways. One path is to continue the search for a small set of universal features, seeking to revise and refine it and find ways of understanding whether there are foundations for different patternings, based, for instance, on inventory contrasts or prosodic position. The understanding of consonantal place has deepened over the years through this method.

An alternative has recently been proposed, to abandon the assumption that features are innate. Mielke (2005, 2008), among others, argues that features are not innate, but emergent (see also CHAPTER 17: DISTINCTIVE FEATURES). The limited number of features that is observed is not surprising, given the shape of the vocal tract and perceptual apparatus. Some ambivalence in patterning is expected for sounds that are phonetically ambiguous. It is thus perhaps time to challenge universality and focus on differences in patterning and what they reveal.

7 Summary

Consonantal place features are perhaps the best studied of all features, and the understanding of place of articulation has deepened over the years. There is little disagreement about the major regions of labial, coronal, dorsal, radical, and laryngeal, with evidence from numerous languages to support these distinctions. At the same time, some sounds do not appear to fit neatly into these classes. Labio-dentals might involve both a labial and a coronal component, and some gutturals pattern with both dorsals and radicals. The sub-places within these major places are less well established, with continuing debate particularly around the nature of the coronal sub-places.

There has been an attempt to identify a particular place of articulation as universally unmarked, based on patterning asymmetries. A good understanding of this requires a careful study of coronal inventories and contrasts, and it appears that much is determined by the language, although there are statistical differences

Davis, Stuart & Michael Hammond. 1995. On glides in American English. *Phonology* 12. 158–182.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology.* Cambridge: Cambridge University Press.

de Lacy, Paul & John Kingston. 2006. Synchronic explanation. Unpublished ms., University of Massachusetts, Amherst & Rutgers University.

Dixon, R. M. W. 1980. *The languages of Australia.* Cambridge: Cambridge University Press.

Douglas, Wilfred H. 1981. Watarri. In R. M. W. Dixon & Barry J. Blake (eds.) *Handbook of Australian languages,* vol. 2, 197–271. Amsterdam: John Benjamins.

Dresher, B. Elan. 2008. The contrastive hierarchy in phonology. In Peter Avery, B. Elan Dresher & Keren Rice (eds.) *Contrast in phonology: Theory, perception, acquisition,* 11–33. Berlin & New York: Mouton de Gruyter.

Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology.* Cambridge: Cambridge University Press.

Duanmu, San. 2000. *The phonology of Standard Chinese.* Oxford: Oxford University Press.

Evans, Nicholas D. 1995. Current issues in the phonology of Australian languages. In Goldsmith (1995), 723–761.

Ferrell, Raleigh. 1982. *Paiwan dictionary.* Canberra: Australian National University.

Flemming, Edward. 2002. *Auditory representations in phonology.* London & New York: Routledge.

Gnanadesikan, Amalia E. 1994. The geometry of coronal articulations. *Papers from the Annual Meeting of the North East Linguistic Society* 24. 125–139.

Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory.* Cambridge, MA & Oxford: Blackwell.

Hall, T. A. 1997. *The phonology of coronals.* Amsterdam & Philadelphia: John Benjamins.

Hall, T. A. 2007. Segmental features. In Paul de Lacy (ed.) *The Cambridge handbook of phonology,* 311–334. Cambridge: Cambridge University Press.

Halle, Morris. 1983. On distinctive features and their articulatory implementation. *Natural Language and Linguistic Theory* 1. 91–105.

Halle, Morris. 2002. *From memory to speech and back: Papers on phonetics and phonology 1954–2002.* Berlin & New York: Mouton de Gruyter.

Halle, Morris, Bert Vaux & Andrew Wolfe. 2000. On feature spreading and the representation of place of articulation. *Linguistic Inquiry* 31. 387–444.

Hamilton, Philip. 1993. On the internal structure of the coronal node: Evidence from Australian languages. Unpublished ms., University of Toronto.

Hansson, Gunnar Ólafur. 2010. *Consonant harmony: Long-distance interaction in phonology.* Berkeley: University of California Press.

Harris, John & Geoff Lindsey. 1995. The elements of phonological representation. In Jacques Durand & Francis Katamba (eds.) *Frontiers of phonology: Atoms, structures, derivations,* 34–79. Harlow: Longman.

Hayward, Katrina & Richard Hayward. 1989. "Guttural": Arguments for a new distinctive feature. *Transactions of the Philological Society* 87. 179–193.

Heath, Jeffrey. 1981. Basic materials in Mara: Grammar, texts, and dictionary. (*Pacific Linguistics* C60.) Canberra: Australia National University.

Heath, Jeffrey. 1984. *Functional grammar of Nunggubuyu.* Canberra: Australian Institute of Aboriginal Studies.

Hualde, José Ignacio. 1992. *Catalan.* London & New York: Routledge.

Hudson, Joyce. 1978. *The core of the Walmatjari language.* Canberra: Australian Institute of Aboriginal Studies.

Hulst, Harry van der & Maarten Mous. 1992. Transparent consonants. In Reineke Bok-Bennema & Roeland van Hout (eds.) *Linguistics in the Netherlands 1992,* 101–112. Amsterdam & Philadelphia: John Benjamins.

Hume, Elizabeth. 2003. Language specific markedness: The case of place of articulation. *Studies in Phonetics, Phonology and Morphology* 9. 295–310.

Hyman, Larry M. 1973. The feature [grave] in phonological theory. *Journal of Phonetics* 1. 329–337.

Hyman, Larry M. 1975. *Phonology: Theory and analysis*. New York: Holt, Rinehart & Winston.

IPA. 2007. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Jakobson, Roman & Morris Halle. 1956. *Fundamentals of language*. The Hague: Mouton.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Jarmasz, Lidia-Gabriela. 2009. Natural classes of coronal obstruents in Polish. *Toronto Working Papers in Linguistics* 30. 35–52.

Jetté, Jules & Eliza Jones. 2000. *Koyukon Athabaskan dictionary*. Fairbanks: Alaska Native Language Center.

Kari, James. 1990. *Ahtna Athabaskan dictionary*. Fairbanks: Alaska Native Language Center.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford & Malden, MA: Blackwell.

Li, Bing. 1996. *Tungusic vowel harmony: Description and analysis*. The Hague: Holland Academic Graphics.

Li, Paul Jen-kuei. 1995. Atayal. In Darrell T. Tryon (ed.) *Comparative Austronesian dictionary: An introduction to Austronesian studies*. Part 1: *Fascicle 1*, 283–292. Berlin & New York: Mouton de Gruyter.

Lombardi, Linda. 1996. Postlexical rules and the status of privative features. *Phonology* 13. 1–38.

Lombardi, Linda. 2002. Coronal epenthesis and markedness. *Phonology* 19. 219–251.

Lucca, Manuel de. 1987. *Diccionario práctico aymara–castellano castellano–aymara*. La Paz: Editorial Los Amigos del Libro.

MacKay, Carolyn. 1994. A sketch of Misantla Totonac phonology. *International Journal of American Linguistics* 60. 369–419.

MacKay, Carolyn. 1999. *A grammar of Misantla Totonac*. Salt Lake City: University of Utah Press.

Mackenzie, Sara. 2009. Contrast and similarity in consonant harmony processes. Ph.D. dissertation, University of Toronto.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Marlett, Stephen A. 1981. The structure of Seri. Ph.D. dissertation, University of California, San Diego.

Marshall, Christina. 2008. The phonetics and phonology of Esan nasalization. M.A. thesis, University of Toronto.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.) *Phonological structure and phonetic form: Papers in laboratory phonology III*, 191–233. Cambridge: Cambridge University Press.

Mielke, Jeff. 2005. Ambivalence and ambiguity in laterals and nasals. *Phonology* 22. 169–203.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Morén, Bruce. 2006. Consonant–vowel interactions in Serbian: Features, representations and constraint interactions. *Lingua* 116. 1198–1244.

Morphy, Frances. 1983. Djapu, a Yolngu dialect. In R. M. W. Dixon & Barry J. Blake (eds.) *The handbook of Australian languages*, vol. 3, 1–188. Amsterdam: John Benjamins.

Mous, Maarten. 1993. *A grammar of Iraqw*. Hamburg: Buske.

Odden, David. 1978. Further evidence for the feature [grave]. *Linguistic Inquiry* 9. 141–144.

Odden, David. 2005. *Introducing phonology.* Cambridge: Cambridge University Press.
Orr, Carolyn. 1962. Ecuador Quichua phonology. In Benjamin F. Elson (ed.) *Studies in Ecuadorian Indian languages,* vol. 1, 60–77. Norman, OK: Summer Institute of Linguistics.
Padgett, Jaye. 2002. Feature classes in phonology. *Language* 78. 81–110.
Paradis, Carole & Jean-François Prunet (eds.) 1991. *The special status of coronals: Internal and external evidence.* San Diego: Academic Press.
Radišić, Milica. 2009. The double nature of the velar /g/ in Serbian. *Toronto Working Papers in Linguistics* 30. 91–103.
Rehg, Kenneth L. 1984. Nasal substitution rules in Ponapean. In Byron W. Bender (ed.) *Studies in Micronesian linguistics,* 317–337. Canberra: Australian National University.
Rehg, Kenneth L. & Damian G. Sohl. 1981. *Ponapean reference grammar.* Honolulu: University of Hawai'i Press.
Rice, Keren. 1989. *A grammar of Slave.* Berlin & New York: Mouton de Gruyter.
Rice, Keren. 1994. Peripheral in consonants. *Canadian Journal of Linguistics* 39. 191–216.
Rice, Keren. 1996. Default variability: The coronal–velar relationship. *Natural Language and Linguistic Theory* 14. 493–543.
Rice, Keren. 2009. On coronals: Are they special? *Toronto Working Papers in Linguistics* 30. 1–15.
Rose, Sharon. 1996. Variable laryngeals and vowel lowering. *Phonology* 13. 73–117.
Rubach, Jerzy. 1993. *The Lexical Phonology of Slovak.* Oxford: Clarendon Press.
Sagey, Elizabeth. 1990. *The representation of features in nonlinear phonology: The articulator node hierarchy.* New York: Garland.
Shahin, Kimary. 2002. *Postvelar harmony.* Amsterdam & Philadelphia: John Benjamins.
Shaw, Patricia A. 1991. Consonant harmony systems: The special status of coronal harmony. In Paradis & Prunet (1991), 125–157.
Steriade, Donca. 1987. Locality conditions and feature geometry. *Papers from the Annual Meeting of the North East Linguistic Society* 17. 595–617.
Svantesson, Jan-Olof, Anna Tsendina, Anastasia Karlsson & Vivan Franzén. 2005. *The phonology of Mongolian.* Oxford: Oxford University Press.
Trigo, Loren. 1988. On the phonological derivation and behavior of nasal glides. Ph.D. dissertation, MIT.
Trubetzkoy, Nikolai S. 1969. *Principles of phonology.* Translated by Christiane A. M. Baltaxe. Berkeley & Los Angeles: University of California Press. Originally published 1939 as *Grundzüge der Phonologie.* Göttingen: van der Hoeck & Ruprecht.
Uhlenbeck, Eugenius M. 1949. *De structuur van het javaanse morpheem.* Bandoeng: Nix.
Watters, James. 1988. Topics in Tepehua grammar. Ph.D. dissertation, University of California, Berkeley.
Wu, Yuwen. 1994. Aspects of Mandarin segmental phonology. Ph.D. dissertation, University of Toronto.
Yip, Moira. 1989. Feature geometry and cooccurrence restrictions. *Phonology* 6. 349–374.
Zuraw, Kie & Yu-An Lu. 2009 Diverse repairs for multiple labial consonants. *Natural Language and Linguistic Theory* 27. 197–224.

23 Partially Nasal Segments

Anastasia K. Riehl
Abigail C. Cohn

1 Introduction

Partially nasal segments, or those that are nasal for only a portion of their duration, contribute many interesting puzzles to the study of phonology. Prenasalized stops, such as /^{ŋ}g/ in the Fijian word /^{ŋ}gone/ 'child', or prestopped nasals, such as /^{p}m/ in the Arrernte word /a^{p}mər/ 'camp', challenge our notion of the segment. If a segment, which we typically think of as being composed of a bundle of features each with a single specification, can be composed of contrasting values for a particular feature (in this case [nasal]), this raises a whole array of questions about the nature of segments. We explore some of these questions in this chapter.

The question of what defines a segment has been a topic of considerable interest in the phonological literature, especially since Chomsky and Halle's (1968; *SPE*) characterization of segments as unordered bundles of features (building on Jakobson *et al.*'s 1952 characterization). Much attention in the phonological literature in the 1970s was given to rethinking and enriching this strict definition of phonological representation, leading to developments including autosegmental representation of tone and other phenomena, formally incorporating syllable structure and higher-order prosodic representation, and so forth. (See Anderson 1985 and Kenstowicz 1994 for a review of these developments, as well as CHAPTER 14: AUTOSEGMENTS; CHAPTER 17: DISTINCTIVE FEATURES; CHAPTER 27: THE ORGANIZATION OF FEATURES; CHAPTER 54: THE SKELETON; and many other contributions to the *Companion to Phonology*.) Particularly relevant in this regard is Anderson's (1974, 1976) discussion of partially nasal segments and his conclusion that such differences cannot be adequately captured through static features (such as *SPE*'s [delayed release]), but rather require direct reference to timing in the phonology, forcing a rethinking of the fundamental definition of the "segment."

There has also been an ongoing discussion in both the phonological and phonetic literature about the phonetic realization of partially nasal segments and nasal–stop clusters and whether or not these two types of NC sequences are phonetically distinct (Herbert 1986; Maddieson and Ladefoged 1993; amongst others). To the degree that such discussions address phonological entities, these issues are also directly relevant to our discussion.

We begin by framing some background issues in the remainder of this introduction, including phonological segmenthood and phonological *vs.* phonetic entities. In §2, we consider the types of partially nasal segments that have been reported in the literature. In the case of nasal–oral entities (NC), one of the central issues is whether these entities are unary segments or clusters, while in the case of oral–nasal entities (CN), a central question is whether such elements are part of the phonology as contrastive elements or only occur allophonically. In §3, we pursue the questions raised by the observed NC patterns and turn more specifically to the issue of unary segments *vs.* clusters, including the issue of their phonetic characteristics. Next, in §4, we address the topic of representations, considering the fit between the attested patterns and the various proposed representations. Finally, in §5, we conclude and offer directions for further research, considering the implementation of the phonological patterns, the role of co-articulation, and possible pathways of change over time.

At this point, we would like to introduce some terminology that we will use frequently throughout the chapter. The abbreviation NC stands for *nasal–oral sequence*, and similarly, CN stands for *oral–nasal sequence*, whether they constitute unary elements or clusters. *Unary* NC and *unary* CN (sometimes abbreviated ^{N}C and ^{C}N) refer to partially nasal sequences that constitute a single segment, while NC *cluster* and CN *cluster* refer to cases where the sequence constitutes two segments. When we wish to refer to a sequence in a generic sense, regardless of its phonological status, we use the terms NC *sequence* and CN *sequence*. We reserve the term *segment* for the unary cases. Where voicing is relevant, we indicate a voiced obstruent portion with a "D" (e.g. D^{N}) and a voiceless oral portion with "T" (as in e.g. NT).

The phonological segment, sometimes equated with the phoneme, is defined by Anderson (1976: 326) as follows: "A segment is usually taken to be a self-contained portion of an utterance which can be characterized in terms of one basic position of the articulatory organs." This is captured in *SPE* with unordered bundles of distinctive features, which are taken to define both the abstract contrastive elements – the phonemes – as well as the actual phonetic events, or phones. At issue is whether such representations – as phonological representations – include internal structure, capturing segment-internal timing.

Before we consider the types of partially nasal segments that exist in the world's languages, we need to be more explicit about the levels of representation for which such structures are being posited. Which types are relevant to the phonology? In other words, which may constitute phonologically distinctive elements, and which are only observed as a phonetic or allophonic result of the phasing of gestures? Clarifying this question is critical for characterizing a typology of such patterns, which in turn is needed to evaluate various representational proposals.

In the literature, many types of partially nasal elements have been described, but often it is not clear whether or not these are really phonological elements. Our interest here is in those cases where timing internal to the segment can result in a phonological contrast. In order to accurately assess which cases are contrastive, we need to separate out both those that result in partially nasal surface elements that are predictable (arguably either the result of phonological rules/constraints yielding allophony or gestural timing patterns), as well as those that have been mischaracterized as unary when they are actually clusters. The issue of allophony is taken up in §2, and the issue of unary segments *vs.* clusters is taken up in §3. We turn now to a review of possible cases.

Table 23.3 Inventory of prestopped nasals, oral stops and nasal stops in Central/Eastern Arrernte

	bilabial	*dental*	*alveolar*	*retroflex*	*palatal*	*velar*
prestopped nasal	ᵖm	ᵗ̪n̪	ᵗn	ʈɳ	ᶜɲ	ᵏŋ
oral stop	p	t̪	t	ʈ	c	k
nasal stop	m	n̪	n	ɳ	ɲ	ŋ

begins, so that by the end of the segment, it has the qualities of a plain nasal, and progressive nasalization may be manifested on following vowels. Prestopped nasals have been much less widely reported than prenasalized stops, and are not found amongst the sample of languages in UPSID. The best-known descriptions are found in Australian languages (Butcher 1999). They have also been reported in Borneo and other languages of the Austronesian family, as well as in some languages of the Austro-Asiatic and Amazonian families (see Blust 1997 for a review). These latter cases, however, are allophonic, and therefore do not provide evidence for prestopped nasals as phonological entities.

We now consider Arrernte, an Arandic language of Australia, as described by Butcher (1999) and Breen (2001), a language where prestopped nasals appear to be phonemic (see also Breen and Dobson 2005). The Central and Eastern dialects of Arrernte have a series of prestopped nasals at various places of articulation, in contrast with plain voiceless stops and plain nasals. No primary voicing contrast is observed for stops, as is characteristic of these languages. Table 23.3 contains the relevant portion of the consonant inventory.

These sounds occur word-initially and word-medially, although an inserted vowel usually precedes all initial consonants, and the language has been analyzed as having only VC(C) syllables (see Breen and Pensalfini 1999). Arrernte allows a limited range of consonant clusters as well, including sequences of nasals and homorganic stops. (4) contains (near-)minimal triplets of these sounds and corresponding plain stops and nasals.

(4) *Arrernte* (Butcher 1999)

	bilabial		*alveolar*	
prestopped nasal	aᵖmər	'gum tree'	aᵗnəm	'yamstick'
oral stop	apər	'camp'	atəɪn	'bursting'
nasal stop	aməl	'nest'	anəm	'staying'

The prestopped nasals in Arrernte are believed to have originated from plain nasals that were allophonically produced with a preceding oral portion when following a stressed oral syllable. Today these dialects are analyzed as having phonemic prestopped segments. Most other southern Australian languages, however, including for example Nukunu and Gupapuyŋu, have allophonic, not phonemic, prestopping.

As mentioned above, prestopped nasals have also been described in several Austronesian languages. For example, in Bonggi, a language of Sabah, Malaysia, within the oral–nasal group, where the nasal and oral portions are more or less prominent in comparison to one another. A distinction for the nasal–oral cases between a prenasalized stop and what is sometimes referred to as a *post-stopped nasal* (also *post-ploded nasal* (Blust 1997), *funny nasal* (Durie 1985), or *obstruent nasal* (Durvasula 2009)) has been claimed, while a distinction within the oral–nasal category has not, to our knowledge, been reported.

As we discuss in Cohn and Riehl (2008), we know of no languages where a compelling case for a phonological *contrast* has been made. Several Austronesian languages, including Rejang and Acehnese, have been argued to have sounds described as primarily nasal, with only a brief oral release (see Blust 1997 for recent review and references to specific cases). However, in phonetic and phonological studies of two putative cases, Acehnese and Sundanese, we find that the NCs are better analyzed, both phonologically and phonetically, as nasal–stop clusters, given that phonologically they pattern like other clusters and phonetically they have the duration of clusters rather than single segments, and their internal nasal–oral timing patterns do not show consistent differences with either ND clusters or prenasalized segments (as discussed in §3). The other type of case where such phonetic differences have been claimed is clearly allophonic in nature (e.g. the post-ploded nasals in some dialects of Chinese; see Chan 1987 and Chan and Ren 1987).

Thus we conclude that finer divisions of this sort within the phonology are not warranted. We follow Anderson (1976) in positing only two phonologically relevant types of partially nasal segments – *pre*nasalized and *pre*stopped segments. We know of no languages that contrast these two types, but we do not predict a priori that such a contrast would not be possible. Rather we expect this is because prestopped nasals are themselves very rare.[4]

2.2 *Other factors: Voicing, place, and manner*

2.2.1 *Voicing*

In the case of prenasalized stops, the majority of the segments reported are fully voiced. In the best-attested cases, the nasal and oral portions are not divisible in the sense that the oral stop portion cannot occur alone. This is the case for Tamambo, discussed above, as well as many other Oceanic languages. In such languages, the prenasalized stop series *is* the voiced stop series. Fully voiced prenasalized stops are also found in some languages that have plain voiced stops and nasals occurring independently, including Wolof. Many Bantu languages are also identified as falling into this category. (However, we take an alternative view, that the NCs are actually clusters (see §3).)

There are also descriptions of putative cases of prenasalized stops with a voiced nasal portion followed by a voiceless oral portion. However, there is a clear asymmetry between these two sorts of cases. While there are a number of clear and widely accepted cases of fully voiced unary ^{N}Ds, the reports of prenasalized voiceless stops are questionable. Sequences of nasals and voiceless stops do

[4] A proposal for laryngeal specification made by Golston and Kehrein (1998) and Kehrein and Golston (2004) that contrast is at the level of the subsyllabic unit, not at the level of the segment, cannot be extended to the nasal specification, as this would predict only one, not two, types of partially nasal segments.

phonologically relevant. We have argued that neither prenasalized stops nor prestopped nasals show a contrast in voicing. Both types of partially nasal segments may occur at various places of articulation, and most have the manner of articulation of stops or affricates, with prenasalized trills also being attested.

In conclusion then, we find that internal structure matters, but only in a very limited way. Specification for [nasal] affects a full segment or a part of a segment, but there are no finer distinctions in terms of the timing of the parts and no additional complexities in terms of differences in place, manner, or voicing. Partially nasal segments are unified in their specifications for voicing, place, and continuancy, as schematized in (5). As noted above, we know of no languages where both ^{N}D and D^{N} are attested contrastively, but we also know of no reason why a priori this would not be possible.

(5) *Possible phonological specifications for nasal*

	N	D	T	^{N}D	D^{N}
nasal	+	−	−	+ | −	− | +
cont	−	−	−	−	−
voice	+	+	−	+	+

In what follows, we focus primarily on those types that are potentially phonologically contrastive, but we will also have occasion to consider the role of timing as it leads to allophonic as well as contrastive patterns. We turn now to the issue of unary partially nasal segments *vs.* clusters, including a discussion of their phonetic properties, and how these may bear on the issue of phonological status.

3 Unary *vs.* cluster NC sequences

Now that we have documented the existence of at least some partially nasal segments, this opens a broader set of questions regarding the relationship of partially nasal segments to NC/CN clusters. The possibility of one sequence of sounds exhibiting a distinction between a single unit *vs.* two units raises various questions of interest to phonology.

There are two distinct yet interrelated issues to address. The first is whether or not the phonology provides evidence for two different patterns. The second is whether or not these different patterns, should they occur, have distinct phonetic realizations. In terms of phonetic characteristics, duration is the factor most commonly discussed (although other possible phonetic differences have been considered).

This discussion ties into broader concerns about the relationship between phonological timing and phonetic duration. Such issues have been widely discussed with regard to geminates *vs.* singletons. In *SPE*, two representations were available for geminates – two identical adjacent features matrices or a single matrix specified as [+long]. Subsequent work has strongly suggested that both aspects play a role (see Broselow 1995; CHAPTER 37: GEMINATES; CHAPTER 47: INITIAL GEMINATES for reviews). In the case of unary ^{N}Cs *vs.* NC clusters, some of the relevant questions parallel these discussions. However, we will argue that in the case of NCs, it is timing and not weight that is primarily at issue, even though, as discussed below, both timing and weight may play a role in a single language.

The phonetic realization of singletons *vs.* geminates has been discussed in a number of cases (see Ham 1998). However, the prior interpretations concerning the phonetic realization of unary NCs *vs.* clusters are much murkier. As discussed in §3.3, we argue that, if appropriate comparisons are made, there *is* a systematic durational difference between unary NCs and clusters.

In the following discussion we focus on prenasalized stops and NC clusters, rather than prestopped nasals and CN clusters, as all of the discussion that we are aware of has focused on the former. In addition, since we know of no cases where a contrast between ^{N}T and NT has been claimed, the discussion of the unary–cluster distinction is limited to voiced NCs, in other words ^{N}D *vs.* ND.

3.1 *The phonological distinction*

The first issue to consider is whether or not there are distinct phonological ND patterns. Here it is useful to observe data from two languages where ND patterns are quite different, English and Fijian.

In English, nasals and stops occur independently, and nasals and stops combine in both orders ND and DN. Importantly, the sequences are limited to positions where they will not violate the language's sonority sequencing constraints (CHAPTER 49: SONORITY; CHAPTER 53: SYLLABLE CONTACT). NDs may be medial where they are usually interpreted as heterosyllabic, as in *am.ber*; final where they decrease in sonority, as in *amend* (though /mb/ and /ŋg/ do not occur word-finally in English synchronically); but not word-initial, where they would violate constraints on sonority in English. NDs also appear across morpheme boundaries (*in-decisive*). These patterns suggest that the sounds are independent segments that cannot be tautosyllabic in an onset. Considering the facts as a whole, English NDs receive a straightforward cluster analysis.

The ND patterns in a language like Fijian look quite different (see e.g. Schütz 1985). In this language, all voiced stops and the alveolar trill are preceded by a nasal component; that is, [b d g] never occur independently, but only as [^{m}b ^{n}d ^{ŋ}g]. Voiceless stops, on the other hand, never occur with a preceding nasal. The ^{N}Ds occur word-initially, as in /^{m}bonu/ 'eel (sp.)', and medially, as in /vu^{n}di/ 'banana'. The language has no word-final consonants and no other consonant sequences, suggesting that all syllables are open. Taking the evidence as a whole, Fijian NCs are straightforwardly analyzed as unary. Based upon such data, there appears to be ample evidence for two different phonological patterns. We argue in §4, following many others, that these two different patterns require distinct phonological representations.

The discussion of ND patterning is complicated by a number of languages where the unary *vs.* cluster status of the sequences is less clear. One widely discussed case involves the Bantu languages. In many Bantu languages, ND sequences are found word-medially in intervocalic position (and sometimes to a limited degree word-initially, where the nasal is part of a prefix). Although their patterning suggests a cluster analysis (see Herbert 1986), the sequences are often analyzed as unary segments, due to the lengthening of vowels observed before the NDs. The following data are from Jita, a Bantu language, spoken in Tanzania (Downing 1996):

(6)	a.	oku-cuma	'to get wealthy'	oku-cu:ma	'to jump'
		oku-loja	'to try'	oku-lo:ja	'to visit the sick'

that can be interpreted as having such a contrast. One such language is Sinhala, which has series of voiced stops, voiceless stops and plain nasals, as well as a singleton–geminate contrast throughout the consonant inventory, and it allows for a wide range of consonant combinations, including ND and NT (see Coates and de Silva 1961; Feinstein 1979; Maddieson and Ladefoged 1993; Letterman 1997). While the NTs are uncontroversially considered clusters, there is an interesting distinction among the NDs, illustrated by pairs of words often transcribed as follows (following Letterman's 1997 transcription):

(7)	a.	/la^{n}da/	'thicket'	/landa/	'blind'
	b.	/ka^{n}de/	'truck (GEN)'	/kande/	'hill (GEN)'

In (7), there appears to be a contrast between a unary and a cluster ND. Further consideration of the phonotactic patterns reveals that the prenasalized stops pattern with singleton consonants and clusters pattern with geminates, for example:

				singular	*plural*
(8)	a.	/balal/	'cat'	[balala]	[balallu]
	b.	/poloŋg/	'viper'	[polə^{ŋ}ga]	[poloŋgu]

In (8a), addition of the plural suffix causes gemination of a preceding stem-final consonant, resulting in a long [ll]. When this suffix follows a prenasalized stop, the result is described as an ND cluster, as in (8b).

Given data such as that above, both Letterman (1997) and Maddieson and Ladefoged (1993) analyze the contrast in Sinhala as being one of a short *vs.* a long ND (although the details of their analyses differ), with Letterman describing the difference as one between a prenasalized stop and a geminate prenasalized stop. This analysis is consistent with Riehl's (2008) conclusion that ^{N}D and ND can only contrast in languages with phonemic length (see also CHAPTER 20: THE REPRESENTATION OF VOWEL LENGTH; CHAPTER 37: GEMINATES). In all cases known to us where a unary–cluster ND contrast has been posited, the languages also have phonemic length, including Fula (Arnott 1970; Maddieson and Ladefoged 1993) and Selayarese (Mithun and Basri 1986) (and likely other languages closely related to Sinhala). The linkage of the unary–cluster ND contrast with phonemic length appears to have the following explanation: since the only phonetic difference between the two cases is length of the nasal (discussed in §3.3 below), this is sufficient for a contrast only in those languages where length independently forms the basis of a contrast.[5]

3.3 *The phonetics of NC*

In recent years, how the phonological and phonetic components of grammar are related has been a topic of interest. The phonetic realization of nasal–obstruent sequences has been a part of these discussions, given that the same apparent

[5] Both an analysis that treats the contrast as one between a singleton and geminate prenasal and one that treats the contrast as between a prenasal and a cluster, could be accommodated under our view. Phonetically, we would expect the same realization, given the internal nasal–oral timing of all NDs, as discussed in §3.3.

sequence of phonetic elements (nasal and obstruent) can pattern as either one or two units in the phonology. Similar attention has been paid to the timing of geminates (see Ham 1998).

It has long been assumed that a unary NC, such as a prenasalized stop, should differ phonetically from a nasal–obstruent cluster in being of shorter overall duration. This assumption is entailed in the distinct symbols used for these sounds, e.g. [ⁿd] *vs.* [nd], where the raised nasal in the unary case implies a less significant nasal portion and shorter overall duration. Herbert (1986), in his monograph on prenasalized stops, includes "shorter overall duration" as one of the defining characteristics of unary segments.

More recently, however, the assumptions about differing phonetic characteristics between these two NC types have been challenged. Maddieson and Ladefoged (1993), for example, state that there is no clear evidence of a phonetic difference and that the issue is rather one for the phonology. However, a review of the literature reveals that there have in fact been very few studies that systematically investigate the question of phonetic difference. Perhaps the most commonly cited is an article by Browman and Goldstein (1986), comparing bilabial NC clusters in English with claimed unary bilabial NCs in Chaga. However, this study, which examines labial trajectories rather than duration per se, is far from conclusive. In short, claims about phonetic characteristics of NCs, whatever they may be, have until recently been unsupported by sufficient phonetic data.

Riehl (2008) conducted parallel phonological and phonetic studies of NC types in four languages, examining phonetic characteristics of NC duration, preceding vowel duration, and preceding and following vowel nasalization. Riehl found that total duration *did* differ significantly between unary and cluster NCs when appropriate comparisons were made (most relevant being the ratio of the duration of plain nasals to NC sequences), while the other characteristics were not consistently different. The graph in Figure 23.1 contains duration data from four languages,

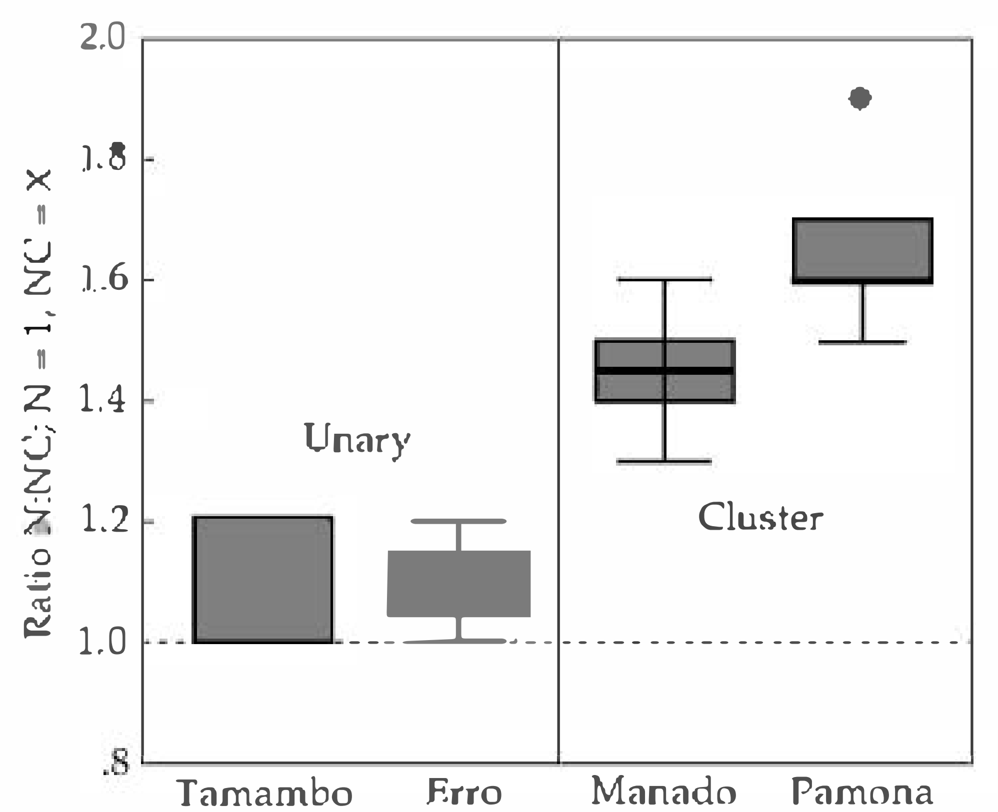


Figure 23.1 Ratio of total duration of plain N to ND sequences at the alveolar place of articulation in unary Tamambo and Erromangan and cluster Manado Malay and Pamona. Average values over ten repetitions of each target word by 4–6 speakers of each language. Duration of [n] = 1

5.2 The asymmetry between prenasalized and prestopped segments

Throughout the chapter, we have discussed the two types of partially nasal segments attested phonologically – prenasalized segments and prestopped nasals – in parallel, despite much greater documentation of ^{N}Cs than ^{C}Ns cross-linguistically. While we have acknowledged the paucity of information on the ^{C}N cases, we have not addressed the source of this imbalance. This imbalance raises several questions. First, are prenasalized stops truly more common cross-linguistically, or do we simply know more about the languages that contain them than we do about those that contain prestopped nasals? Second, if prestopped nasal segments are less common, is this simply accidental, or is there a principled explanation for why this is so? Third, if indeed the only evidence for prestopped nasal segments comes from those few cases identified above, is it possible that further examination of those cases will reveal that the contrastive cases are actually clusters and that there is no evidence for phonological unary prestopped segments at all? Having reviewed the literature as carefully as possible, we believe that the imbalance is neither due to a lack of appropriate documentation nor accidental, but more data is required to figure out why.

5.3 Pathways to change

The observed asymmetries between ^{N}C and ^{C}N and the possible phonetic motivations for these patterns lead us to consider what the pathways to change might be which result in partially nasal segments as part of a phonemic inventory of a language. Notably, the pathways of ^{N}C and ^{C}N seem to be different.

As discussed in §2.1 above, one characteristic pattern of unary ^{N}Cs is that they often constitute the only series of voiced stops in the language. As discussed above, such cases arguably arise as an enhancement effect where the addition of nasal airflow helps maintain full voicing as an enhancement of the voicing contrast (Ohala 1983; Iverson and Salmons 1996; Nasukawa 2005). The historical pathway for ^{C}Ns appears to be a result of co-articulatory timing. This pathway seems less likely to result in contrastive partially nasal segments, suggesting parallels to Cohn's (1993) observations of what circumstances result in contrastively nasalized approximants where weakening, not contextual nasalization, seems to result in the attested cases. Another pathway noted by a reviewer is partial denasalization of nasal stops, as observed for example in Athabaskan (e.g. Leer 1996). As we come to better understand the synchronic typologies of such cases, we can start to address the question of possible pathways to change.

5.4 Further directions

In closing, this chapter highlights the need for more data to better understand the cross-linguistic patterns of partially nasal segments and the implications of such patterns for phonological representations. This includes the continued documentation of less well-studied languages especially in underdocumented language families; the critical importance of understanding such data as part of a phonological system, not just as isolated elements; and finally the need for additional careful phonetic analysis that can inform not only the phonetic description of such cases, but also the phonology.

Clements, G. N. 2003. Feature economy in sound systems. *Phonology* 20. 287–333.
Coates, William A. & M. W. S. de Silva. 1961. The segmental phonemes of Sinhala. *University of Ceylon Review* 18. 163–167.
Cohn, Abigail C. 1990. Phonetic and phonological rules of nasalization. Ph.D. dissertation, University of California, Los Angeles.
Cohn, Abigail C. 1993. The status of nasalized continuants. In Huffman & Krakow (1993), 329–367.
Cohn, Abigail C. & Anastasia K. Riehl. 2008. The internal structure of nasal–stop sequences: Evidence from Austronesian. Paper presented at the 11th Conference on Laboratory Phonology, Wellington.
Downing, Laura J. 1996. *The tonal phonology of Jita*. Munich: Lincom Europa.
Downing, Laura J. 2005. On the ambiguous segmental status of nasals in homorganic NC sequences. In Marc van Oostendorp & Jeroen van de Weijer (eds.) *The internal organization of phonological segments*, 183–216. Berlin & New York: Mouton de Gruyter.
Durie, Mark. 1985. *A grammar of Acehnese on the basis of a dialect of North Aceh*. Dordrecht: Foris.
Durvasula, Karthik. 2009. Understanding nasality. Ph.D. dissertation, University of Delaware.
Feinstein, Mark. 1979. Prenasalization and syllable structure. *Linguistic Inquiry* 10. 245–278.
Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT. Published 1979, New York: Garland.
Golston, Chris & Wolfgang Kehrein. 1998. Mazatec onsets and nuclei. *International Journal of American Linguistics* 64. 311–337.
Halle, Morris & G. N. Clements. 1983. *Problem book in phonology*. Cambridge, MA: MIT Press.
Ham, William. 1998. Phonetic and phonological aspects of geminate timing. Ph.D. dissertation, Cornell University.
Hayes, Bruce & Tanya Stivers. 1995. The phonetics of post-nasal voicing. Unpublished ms., University of California, Los Angeles.
Herbert, Robert K. 1986. *Language universals, markedness theory, and natural phonetic processes*. Berlin: Mouton de Gruyter.
Hubbard, Kathleen. 1995a. Prenasalized consonants and syllable timing: Evidence from Runyambo and Luganda. *Phonology* 12. 235–256.
Hubbard, Kathleen. 1995b. Toward a theory of phonological and phonetic timing: Evidence from Bantu. In Bruce Connell & Amalia Arvaniti (eds.) *Papers in laboratory phonology IV: Phonology and phonetic evidence*, 168–187. Cambridge: Cambridge University Press.
Huffman, Marie K. 1993. Phonetic patterns of nasalization and implications for feature specification. In Huffman & Krakow (1993), 303–327.
Huffman, Marie K. & Rena A. Krakow (eds.) 1993. *Nasals, nasalization, and the velum*. Orlando: Academic Press.
Hyman, Larry M. 2001. The limits of phonetic determinism in phonology: *NC revisited. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 141–185. San Diego: Academic Press.
Iverson, Gregory K. & Joseph C. Salmons. 1996. Mixtec prenasalization and hypervoicing. *International Journal of American Linguistics* 62. 165–175.
Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.
Jauncey, Dorothy. 1998. A dictionary of Tamambo, Malo Island, Vanuatu. Unpublished ms., Australian National University.
Ka, Omar. 1994. *Wolof phonology and morphology*. Lanham, MD: University Press of America.
Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1985. The internal structure of phonological elements: A theory of charm and government. *Phonology Yearbook* 2. 305–328.
Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology* 7. 193–231.

Kehrein, Wolfgang & Chris Golston. 2004. A prosodic theory of laryngeal contrasts. *Phonology* 21. 325–357.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kreitman, Rina. 2008. The phonetics and phonology of onset clusters: The case of Modern Hebrew. Ph.D. dissertation, Cornell University.

Leer, Jeff. 1996. The historical evolution of the stem syllable in Gwich'in (Kutchin/ Loucheux) Athabaskan. In Eloise Jelinek, Sally Midgette, Keren Rice & Leslie Saxon (eds.) *Athabaskan language studies: Essays in honor of Robert W. Young*, 193–234. Albuquerque: University of New Mexico Press.

Letterman, Rebecca. 1997. The effects of word-internal prosody in Sinhala: A constraint-based analysis. Ph.D. dissertation, Cornell University.

Maddieson, Ian & Peter Ladefoged. 1993. Partially nasal consonants. In Huffman & Krakow (1993), 251–301.

Maddieson, Ian & Kristin Precoda. 1990. Updating UPSID. *UCLA Working Papers in Phonetics* 74. 104–111.

Mithun, Marianne & Hasan Basri. 1986. The phonology of Selayarese. *Oceanic Linguistics* 25. 210–254.

Nasukawa, Kuniya. 2005. *A unified approach to nasality and voicing*. Berlin & New York: Mouton de Gruyter.

Ndiaye, Moussa D. 1995. Phonologie et morphologie des alternances en Wolof. Ph.D. dissertation, Université de Québec à Montréal.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.) *The production of speech*, 189–216. New York: Springer.

Ohala, John J. & Manjari Ohala. 1991. Nasal epenthesis in Hindi. *Phonetica* 48. 207–274.

Pater, Joe. 1999. Austronesian nasal substitution and other NÇ effects. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*, 310–343. Cambridge: Cambridge University Press.

Pater, Joe. 2001. Austronesian nasal substitution revisited. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 159–182. Cambridge: Cambridge University Press.

Piggott, Glyne L. 1989. Prenasalization and feature geometry. *Papers from the Annual Meeting of the North East Linguistic Society* 19. 345–357.

Quick, Philip A. 2003. A grammar of the Pendau language. Ph.D. dissertation, Australian National University.

Riehl, Anastasia K. 2008. The phonology and phonetics of nasal-obstruent sequences. Ph.D. dissertation, Cornell University.

Riehl, Anastasia K. & Dorothy Jauncey. 2005. Illustrations of the IPA: Tamambo. *Journal of the International Phonetic Association* 35. 255–259.

Rosenthall, Sam. 1988. The representation of prenasalized consonants. *Proceedings of the West Coast Conference on Formal Linguistics* 7. 277–291.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT. Published 1990, New York: Garland.

Schütz, Albert J. 1985. *The Fijian language*. Honolulu: University of Hawaii Press.

Steriade, Donca. 1993. Closure, release and nasal contours. In Huffman & Krakow (1993), 401–470.

Stetson, Raymond H. 1951. *Motor phonetics: A study of speech movements in action*. 2nd edn. Amsterdam: North-Holland.

Storto, Luciana. 1999. Aspects of Karitiana grammar. Ph.D. dissertation, MIT.

Storto, Luciana & Didier Demolin. 2002. The phonetics and phonology of unreleased stops in Karitiana. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 28. 487–497.

Teoh, Boon Seong. 1988. Aspects of Malay phonology revisited: A non-linear approach. Ph.D. dissertation, University of Illinois.

24 The Phonology of Movement in Sign Language

WENDY SANDLER

To watch a person communicating in a sign language is to observe a well-coordinated, multi-channel display of bodily motion. Most salient in this display is movement of the hands, which transmit lexical, morphological, and timing information. In coordination with the hands, motion of the mouth and lower face performs phonological, morphological, and gestural functions. Simultaneously, movement of the head and body provides a kind of prosodic shell to house the signing hands. Prominently embedded in this outer shell are the brows and eyes, whose movements provide intonation, the visual "tunes" of the message. The same physical articulators are all exploited by speakers as well, to augment the linguistically organized vocal–auditory signal.[1] But in sign language, it is these visually perceived actions that convey the linguistic signal itself, in a synchronized panoply of motion.

What are the primitives of movement at each level of linguistic structure? How are they organized? In what ways does the phonological organization of movement correspond to phonological properties of spoken languages? Can a comparison of the two help to separate universal linguistic properties from modality effects in language generally? How does the organization of such a system arise? These are the issues addressed in this chapter, which focuses on the phonological category of movement. The other two major phonological categories – hand configuration and location – are not the main objects of interest here.[2]

The first section describes the nature of movement in lexical signs, providing evidence for the existence and unity of movement as a phonological category. The section also provides a model for the representation of movement, briefly noting its strengths and weaknesses and those of alternative models. The goal of that section is to give the reader a feel for the nature of movement in lexical signs and for some of the issues involved in representing it phonologically.

[1] The term "speakers" as used here refers only to producers of spoken languages, while "signers" refers to producers of sign language.
[2] See Sandler and Lillo-Martin (2006) for a detailed overview of sign language phonology.

The morphology of sign languages exploits movement forms that are not found in simplex signs, a peculiarity of sign languages that is described in §2, where movement patterns in temporal aspect morphology are analyzed by means of morphological templates. The section also shows that morphological complexity is achieved not only by movement of the hands, but by movement of the mouth as well. A subsystem of the grammar in which movement appears to have certain analogic properties completes §2.

Also playing a linguistic role in sign language is movement of the upper face, head, and torso. These articulators provide prosodic and other cues, simultaneously with the lexical and timing information conveyed by the hands. The organization of movement in the service of prosody is the subject of §3, followed by §4, a brief comparison with the way speakers use visible movement.

While all established sign languages seem to exploit movement in similar ways, most likely due to the interaction of particular iconic and motoric underpinnings, evidence from a newly emerging sign language described in §5 suggests that these typological features are not "given" at the outset. Instead, lexical movement, obligatory in signs of established sign languages, takes phonological form only gradually, as the linguistic system as a whole emerges. §6 concludes the chapter by considering what the study of sign language phonology and movement might teach us about phonology in general, through a comparison of movement in sign language with phonological elements of spoken languages.

1 Movement in lexical signs

Examples of basic types of manual movement and their combinations that occur in lexical signs are illustrated in Figure 24.1. The examples are from Israeli Sign Language (ISL), except for Figure 24.1g, which is from American Sign Language (ASL), but the same types are found in other sign languages as well. The two types of lexical movement are path movement and internal movement of the hand or hands. Path movement is generated at the shoulder or elbow and results in moving the hand in a path through space. Internal movement is generated either by the wrist, resulting in orientation change, or by the fingers, resulting in a change in the shape of the hand.

Path movement can have the shapes [straight] (the default, shown in Figure 24.1a), [arc], or, in ASL, "7", the latter used primarily for initialized city names, in which the handshape is taken from the fingerspelled first English letter of the name (see CHAPTER 9: HANDSHAPE IN SIGN LANGUAGE PHONOLOGY for *initialization*), and also a rare "Z" shape, e.g. in NEVER and LIGHTNING in ASL. Lexical movement is constrained in the following ways: there is typically only one movement in a sign, performed on or near a single area of the body, the head, torso, arm, or hand. Another way to create a path movement is by moving the hand/s in a circle. Some researchers consider circling movement to be an additional movement feature in addition to [straight] and [arc], while others favor an analysis of circles as consisting of a sequence of arcs with different values for concavity, a point to which we return in §3. Lexical movements can also be characterized by the features [tense] and [doubled].[3]

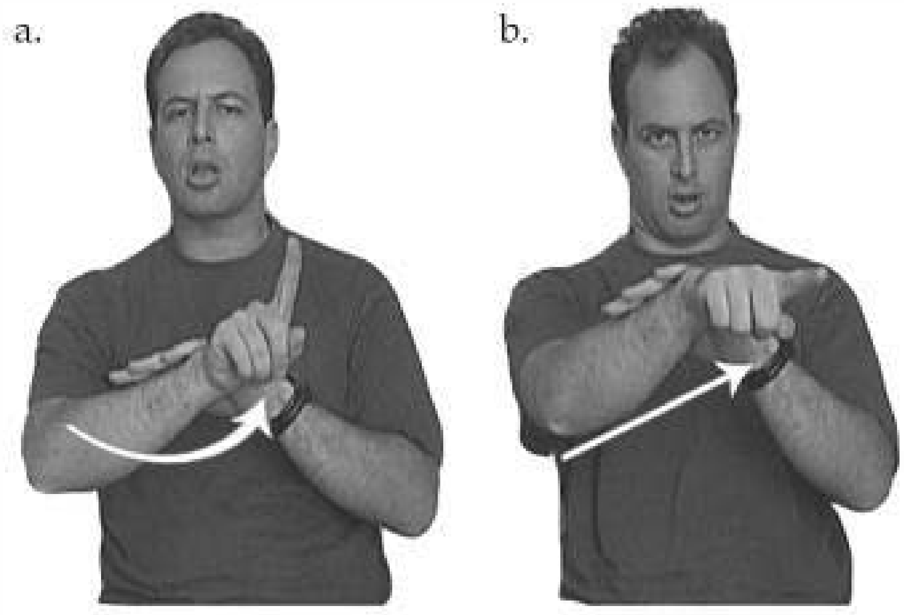


Figure 24.2 An ISL minimal pair differing only in movement shape

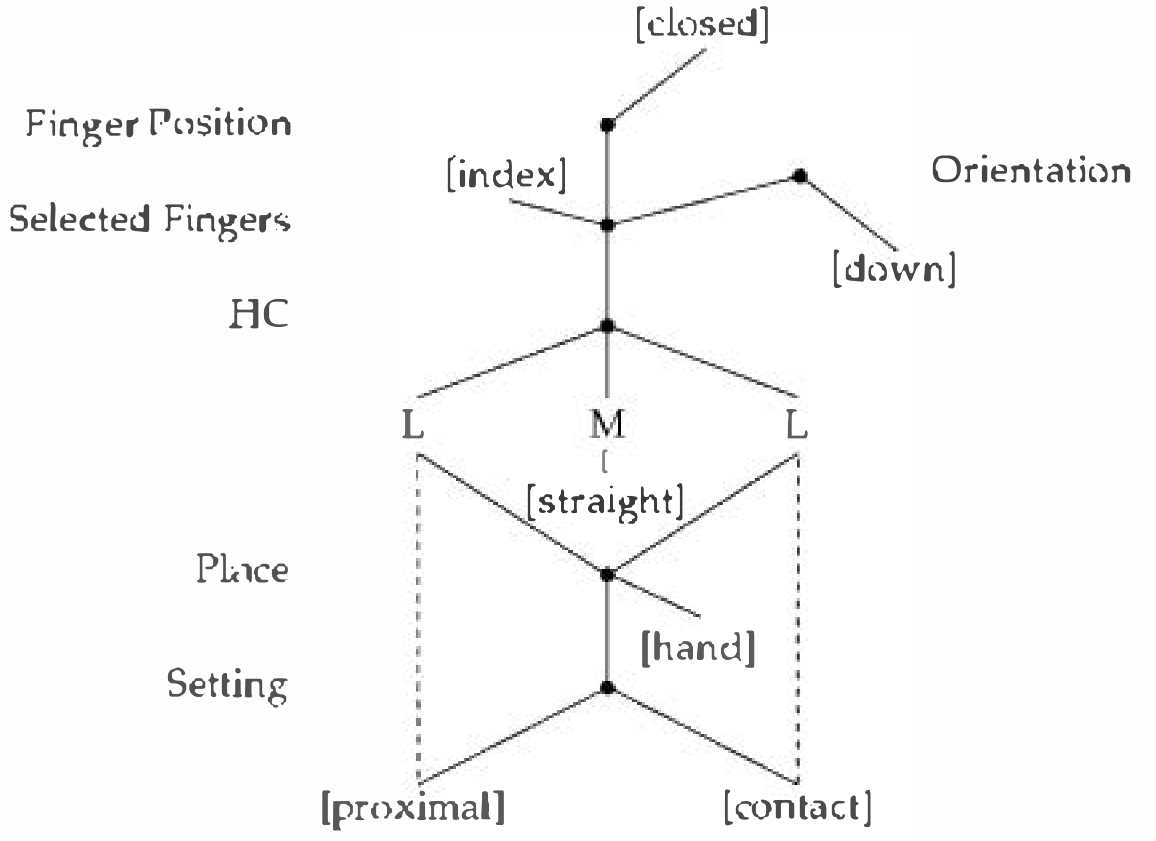
distinguished by single or doubled movement can also be found, but are also very uncommon.

A straightforward linguistic argument for a movement category is reference to movement in phonological constraints and processes. A constraint on a rule involving movement is the blocking of a particular agreement marker in verbs that have specified movement. In ISL, verbs are blocked from undergoing inflection for multiple object agreement if they are specified for any feature belonging to the class of (non-default) lexical movement features, such as [arc] (in the vertical plane), [tense], or [doubled][4] (Sandler 1996). In such signs, the verb is not inflected for multiple object agreement; instead, an object pronoun is used, and it carries the inflection. Since the multiple agreement form consists of a horizontal arc movement which is superimposed on the sign, it appears that its blocking by a specified movement feature such as [tense] or [doubled] is a constraint on movement complexity. Verbs of the relevant type with the most common path movement, the default straight path movement, all do undergo this inflection. Apparently, a similar constraint is active in ASL (Padden 1988).

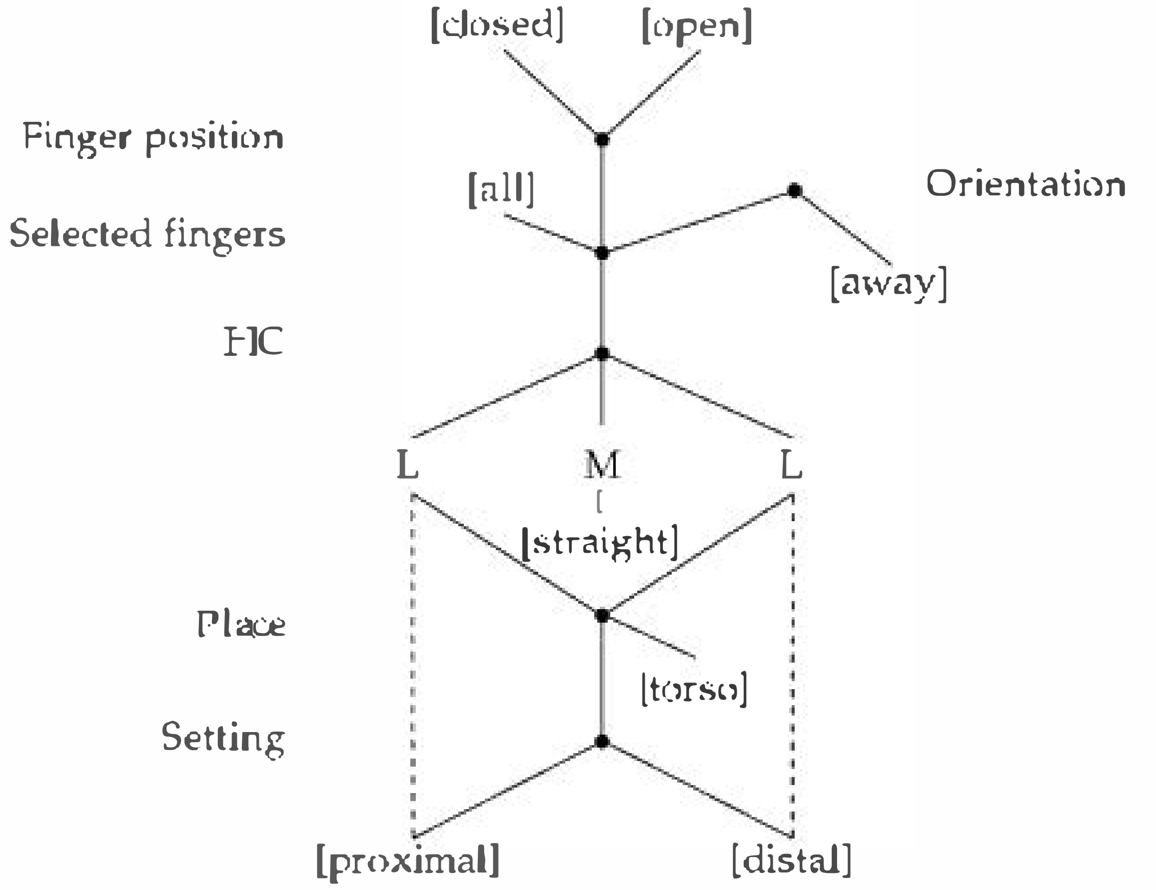
A morphological rule that refers to the movement category provides further evidence for the category in sign language phonology. This morphological rule of reduplication for Iterative and other temporal aspects reduplicates the final syllable of a sign. The rule refers to movement in the sense that a syllable is defined as consisting either of a single movement (path or internal) or of a simultaneously executed path and internal movement. Two movements in a sequence constitute two syllables. Most signs are monosyllabic, and are fully reduplicated when inflected for various temporal aspects. However, compounds may be disyllabic, and, if so, then the final syllable – with movement as its nucleus – is reduplicated (Sandler 1989). This is the case whether the movement in that syllable is path or internal, which in turn is an argument for unifying both types of movement in a single category.

4 [doubled] movement, named [restrained] in Sandler (1996), refers to lexical doubled movement, which differs from morphological reduplication both phonetically and phonologically.

(2) a. *Representations of ISL* JUST-THEN, *a sign with plain path movement*



b. *Representations of ISL* SEND, *a sign with path and internal movement*



The aspect of internal movement captured by this representation is its intimate connection to the hand, both physically and behaviorally, in the phonology. While path movement is achieved by movement at higher joints (elbow or shoulder), internal movement is achieved by movement of the fingers (handshape change; Figure 24.1b) or wrist (orientation change; Figure 24.1c). An argument for the intimate connection between internal movement and the category of hand configuration comes from assimilation in lexical compounds, where all hand configuration features spread, including their movement. A version of a compound meaning FAINT, composed of the signs MIND and DROP, illustrates this clearly, as seen in Figure 24.3.

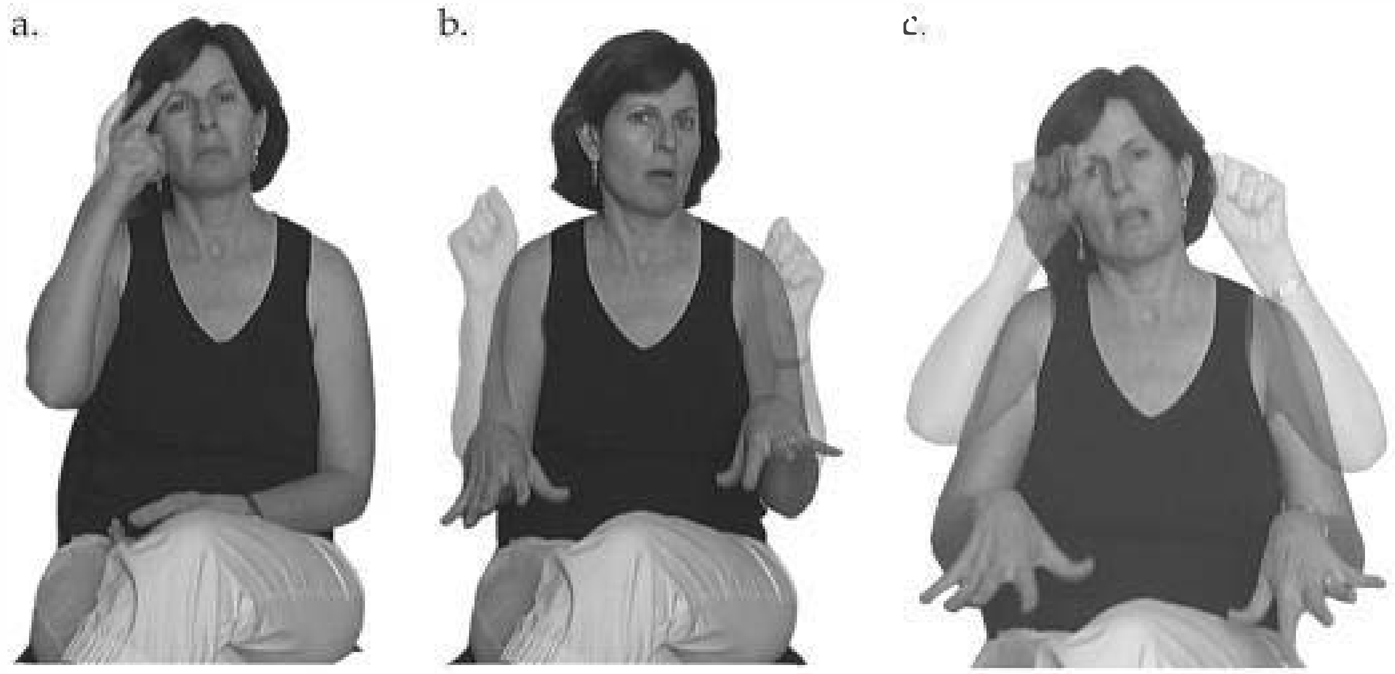
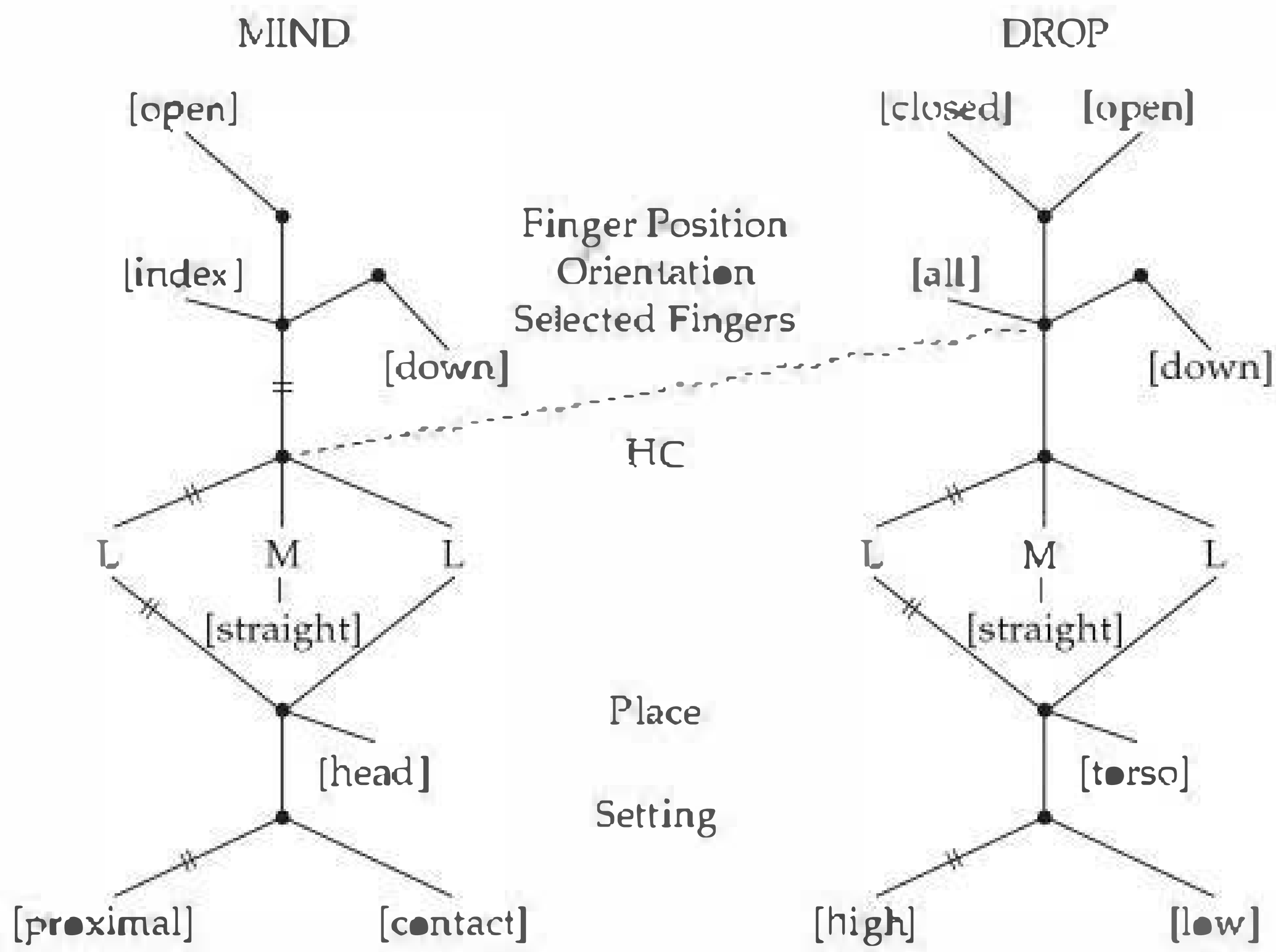


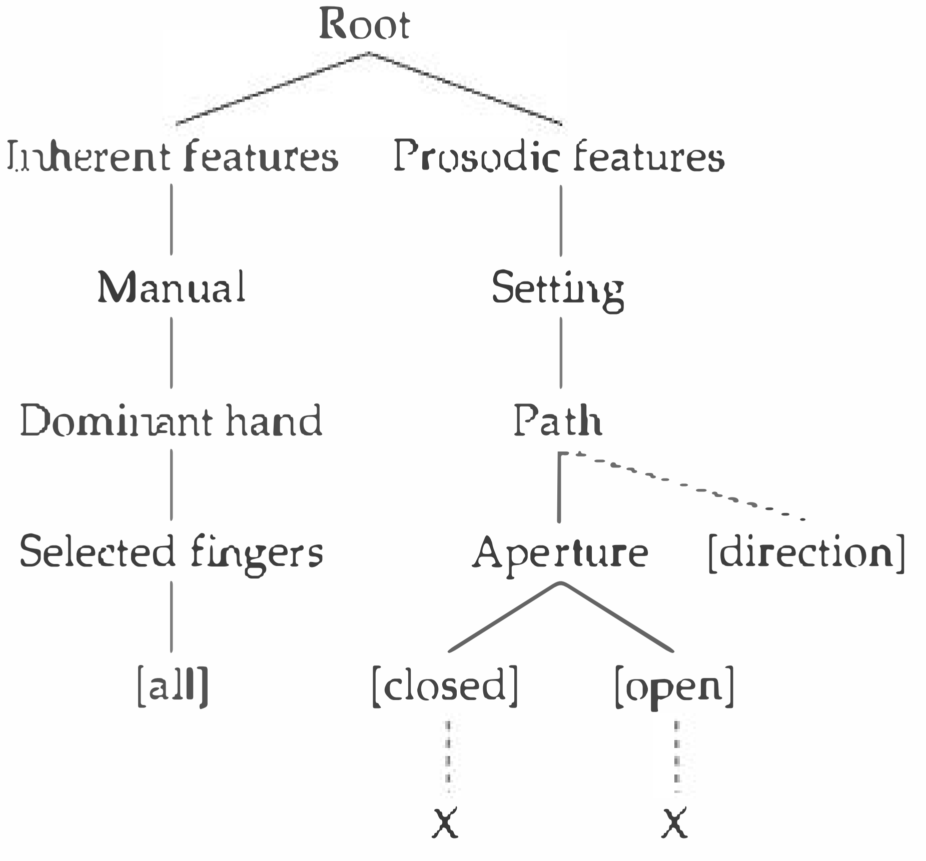
Figure 24.3 The ASL lexical compound, MIND^DROP = FAINT

Example (3a) shows that the entire hand configuration node, with its branching features for internal movement, spreads regressively (dashed line), associating to the second location of MIND. The surface form of the compound is represented in (3b).

(3) a. *Representations of compound reduction: Truncation and assimilation of* MIND^DROP



(5) *Prosodic Model schema with internal movement (handshape change)*



Each of the two models accounts for a number of different theoretical and empirical phenomena in sign language phonology, and both models alleviate certain problems inherent in the Hand Tier model, in particular, the linearization problem, in which path and internal movement are not accommodated in a unified way. But each has disadvantages as well. Omitting a movement segment from the representation, as the Dependency Model does, makes it difficult to accommodate the durational effects on movement described in §3, and the distribution of the feature [contact], which can be realized on either of the Ls, on the M, or on all segments (Sandler and Lillo-Martin 2006). A coherent account of the syllable reduplication facts described above also provides a challenge for that model.

A drawback of the Prosodic Model lies in the segregation of the internal movement features from the hand configuration features in two separate branches of structure. In this way, the model sacrifices the phonetic and phonological integrity of hand configuration (the latter supported by the compound assimilation facts, for example). Doing away with a Movement segment, a representation inherited from the Dependency Model, presents the same problems for the Prosodic Model.

No model comfortably accommodates all the facts. Unfortunately, a detailed review of existing models of movement would take us too far from the goal here of introducing movement in sign language to a general linguistic audience, and interested readers are referred to the sources cited above, and to Sandler and Lillo-Martin (2006: ch. 13), for further illustration and discussion.

1.3 *Lexical movement and meaning*

A central property of phonological units is that they are meaningless. This "duality of patterning" of language – the division into a meaningful level of morphemes, words, and sentences, as well as a meaningless level of auditory or visual formational elements – has been called a basic design feature of human language (Hockett 1960). This property is what gives human language its ability to create vast vocabularies from a small number of primitive elements. Yet sign and locations remain unchanged, while the movement pattern alone is affected – by changes in shape, duration, rhythm, and number of iterations. Klima and Bellugi describe 11 aspects for ASL, which they distinguish from one another by means of the following binary features: reduplicated, even, tense, end-marked, fast, and elongated.

The system has certain key elements in common with the root and pattern morphology of Semitic languages, in which the consonants of the root remain constant, while the vowel pattern and the rhythm (expressed through gemination of different consonant or vowel positions in the morphological template) are altered to create different verb forms (see CHAPTER 105: TIER SEGREGATION; CHAPTER 108: SEMITIC TEMPLATES). The abstract vocalic and timing patterns constitute morphemes of voice or aspect and of verb form, such as (roughly) reflexive, causative, etc. Influenced by McCarthy's model for Standard Arabic (McCarthy 1981), a templatic treatment of ASL aspect proposes similar timing templates and movement features for the aspectual system of that language. Associating morphological features to the movement segment and geminating either locations or movements produce aspectual patterns (Sandler 1990). For example, the citation forms of the verbs GIVE and LOOK-AT are shown in Figure 24.4. Each has a plain straight movement with no special shape or timing qualities. If these verbs are inflected for temporal aspect, then the form gets its shape from the path movement feature and its timing from the aspectual template. For example, inflection for Durational adds a reduplicated circular movement to the base forms, illustrated in Figure 24.5. For any verb so inflected, the movement pattern for any given aspect is the same.

Both GIVE and LOOK-AT belong to the class of verbs which in many sign languages agree for subject and object or, more precisely, for source and goal (Padden 1988; Meir 2002). Agreement is realized by moving the hand(s), configured according to the feature specifications of the base verb, from a spatial locus associated with the source to a locus associated with the goal. The citation forms shown in Figure 24.4 are very similar to the inflected forms for I-GIVE-YOU and I-LOOK-AT-YOU. That is, the initial locus (1st person) is typically near the torso of the signer, the final locus (2nd person) is at a distal location in front of the signer and movement is from the signer toward the addressee. The spatial loci for 3rd person referents are typically established during the discourse, and consist of points in space which defy precise phonological representation.

a.

b.

Figure 24.4 Citation forms of ASL (a) LOOK-AT and (b) GIVE. Reprinted with permission from Ursula Bellugi

in the class of agreement verbs), while there is a conventionalized set of movement shapes and timing templates which determine aspect.

2.2 Mouth movement in sign language morphology

In addition to the infrequent lexical specification of mouth and movements, noted in §1 above, repeated tongue flap ("lalalala") plays a grammatical role in many sign languages, often marking predicates for iterative or exhaustive inflection, sometimes with additional manually produced inflection such as reduplication or horizontal arc. For example, in ISL, the horizontal arc of the exhaustive verb inflection can be accompanied by a rapid tongue movement of this kind (Meir and Sandler 2008).

A different tongue movement – side-to-side tongue wag – has been closely investigated in the Delayed Completive aspect in ASL (Brentari 1998). In this language, tongue wag is an allomorph of finger wiggle, marking the Delayed Completive aspect. This operation productively applies to telic verbs, with some phonological restrictions, and adds the meaning "delay the completion of x," where x is the base verb.

In addition to morphological functions, mouth shapes and movements perform prosodic and gestural functions (Sandler 1999, 2009; Boyes Braem and Sutton-Spence 2001). In fact, in these "manual" languages, the mouth is often in constant motion – not in the service of speech, but in the service of language.

2.3 Predicates of motion and location: A partly linguistic subsystem

Established sign languages typically have a subsystem within their grammars for designating location and motion of classes of referents. This system, called verbs of motion and location, classifier predicates, or classifier constructions, combines classifier handshapes with motions and locations in space. Identified and analyzed in the early years of sign language research (by Supalla 1982, 1986, and others), the system has subsequently been tackled by a number of other researchers of ASL and several other sign languages (see Emmorey 2003). The organization of these constructions clearly has conventionalized properties, and the system is apparently difficult to acquire by both first and second language learners.

In classifier predicates, there is a conventional set of handshapes corresponding to classes of referents (hence the term "classifiers"). Unlike the handshape category in lexical signs, which is a meaningless phonological unit, the classifier handshapes are meaningful morphemes, which fall into three categories. Size and shape classifiers class objects according to physical properties, like small and round, flat, or cylindrical. Handling classifiers mimic the shape of the hand or object (like a hook or tongs) handling another object, be it flat and thin, like paper, or thick, like a book, for example. Entity classifiers refer to semantic classes such as humans, small animals, legs, eyes, or vehicles. The handshapes in classifier systems are conventionalized, and vary from sign language to sign language; they are part of the grammatical system.

These classifier handshapes can combine with other classifier shapes on the non-dominant hand, an impossibility in lexical signs, in which the non-dominant hand too has meaningless phonological status and is strictly constrained in shape

[IF HE INVITE-ME PARTY HIS]$_I$	[I COME-TO PARTY HIS]$_I$	
Brow Raise		
Wide Eyes		
Tilt right		*Tilt left*
Head forward	*Head back*	
Head down	*Head up*	
Hold		*Hold*

Figure 24.9 Alignment of facial and head positions with the timing cues of the hands at the end of each intonational phrase produces salient and synchronized movement at the boundary. The timing cue in the text is cessation of movement (holding the signing hands still)

4 Movement of hands, face, and body in spoken language: A comparison

Unlike sign language, spoken language is usually not described in terms of movement of the face, hands, and body. Instead we think of language in the auditory modality as produced by the largely hidden vocal tract, suggesting that hearing people are not conduits of the variegated motion ensembles described above – and speakers may indeed look rather strange to deaf people. But that is not because speakers are static. They, too, make use of the head, body, face, and hands when they communicate, as the burgeoning fields of co-speech gesture and "visual prosody" reflect. Speakers of all languages and cultures use their hands rhythmically and imagistically to augment the linguistically organized signal in ways that have motivated a theory of co-speech gesture as integral to language (see e.g. McNeill 1992; Goldin-Meadow 2003; Kendon 2004). Speakers use the face and body as well to "punctuate" and augment their speech (see Swerts and Krahmer 2009 for a recent collection of articles on visual prosody). But the motions of the face and body that accompany speech, though rich, are not as finely tuned, conventionalized, or systematically coordinated as are the motions of prosody accompanying signed utterances. In sign language, these signals are organized into a linguistic system.[6] Current work on a new sign language gives clues to how such a linguistic system emerges.

5 Emergence of the movement category

The foregoing exposition leapt back and forth in rather cavalier fashion from ASL to ISL, with forays to SLN, as if the system being described were common to all sign languages. And indeed, while there certainly are language-specific differences, most of the characteristics of movement described in the broad description

[6] While sign languages organize motion of the hands, face, and body into a linguistic system, this does not mean that they do not also avail themselves of paralinguistic gestural elements corresponding to co-speech gesture. In fact, they do. A particular kind of iconic co-sign gesture in sign languages is commonly made with the mouth (Sandler 2009).

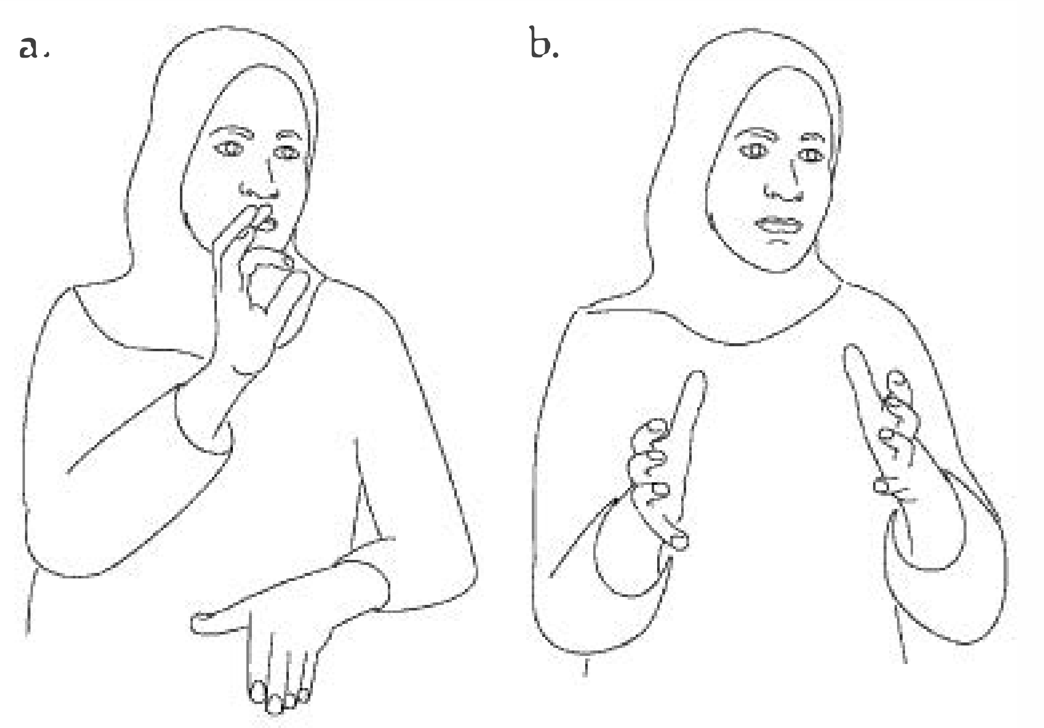


Figure 24.10 Al-Sayyid Bedouin Sign Language (ABSL) KETTLE, comprised of CUP and ROUND-OBJECT, as signed in an extended family with many deaf members

at some location with a transitional movement only – a phonetic possibility that established sign languages do not exploit. A system like that of ABSL does not need to count syllables, unlike ASL, for example (see §1), and has not arrived at LML signs as a (possibly perceptually advantageous) alternation between static and dynamic elements. In such a system, movements are not mandatory. Interestingly, in the third generation, among children who dwell in households with a deaf parent and deaf siblings, we begin to see mandatory movement emerging. Movement segments are introduced into everyday signs such as GOAT and KETTLE in these children, performed without movement by others in the village (Sandler *et al.*, forthcoming). KETTLE as signed in one household, a compound made up of the signs CUP and ROUND-OBJECT, is shown in Figure 24.10.

The second part of the sign KETTLE is typically signed by placing the hands in front of the torso in the configuration shown in Figure 24.10b. But a young girl in this household signed it differently, introducing a salient movement of the two hands toward each other, as shown in Figure 24.11. This movement is counter-iconic, as the kettle does not become any smaller. An epenthetic movement is introduced in the sign GOAT by two children in different deaf households that include a deaf parent (an unusual situation, as the gene is recessive). That sign is typically made with an "L" hand, by placing the thumb in contact with the temple – an iconic representation of a goat's horn. But the two children from deaf households whom we recorded introduced a repeated tapping movement at the temple. Again, the movement is not iconic. A goat's horn does not tap its head, and this movement can be seen as introducing a formal element that will ultimately be exploited in a phonological system.

6 Summary and conclusion

Perhaps a key to understanding sign language phonology, and movement as part of it, is to recognize the fact that no potential visual signal goes unexploited. But these signals, available to all humans, are molded in sign languages into a

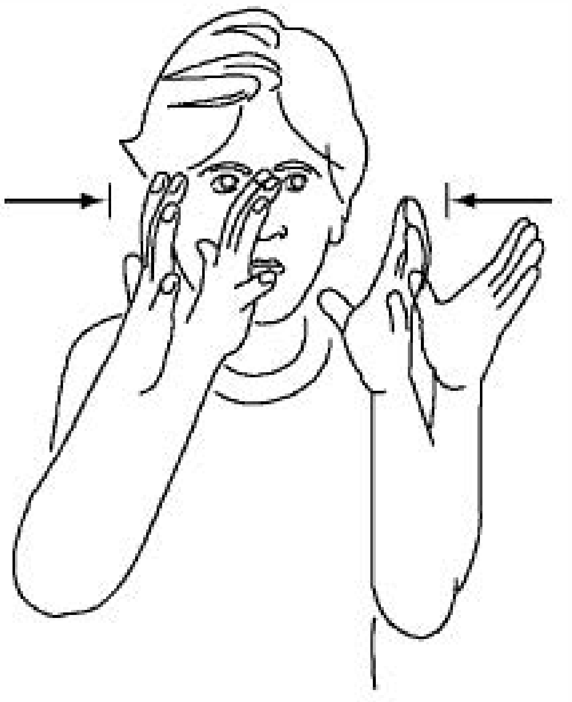


Figure 24.11 Movement epenthesis in ROUND-OBJECT by a small child in the family

well-disciplined vehicle for conveying particular kinds of linguistic information at different levels of structure.

Lexical signs must have some movement, even if it is only a default straight path, a requirement that implies an organizational advantage for the category – perceptual, articulatory, contrastive, or all three. Within signs, the primary movement is manual – either path, hand internal, or the two simultaneously – characterized by a small set of movement features.

More evidence for the existence of a movement category comes from morphological constraints and rules that refer to it. Aspectual morphology makes systematic use of movement patterns – templates of movement shape and timing that are comparable in some respects to the templatic morphology of Semitic languages. More freedom of movement appears to be available to the classifier predicate subsystem, which is exploited for types of discourse in which spatial relations and manners of movement are prominent, and which provides a vehicle for stylistic creativity, as in poetry.

As the hands move in space, assuming various configurations, they are sometimes accompanied by coordinated movement of the mouth, which can serve lexical, morphological, prosodic, or gestural functions. Above the lexical level, prosodic constituents such as the intonational phrase are marked by movement of the head and torso, and meaningful intonational arrays are overlaid on these displays by actions of the upper face.

Non-trivial similarities to spoken language phonological categories can be identified here: (i) signs must have an alternation between static and dynamic components, as spoken words must have consonants and vowels;[8] (ii) there is a limited inventory and combination of features in a given phonological category (here the movement category) within lexemes (CHAPTER 17: DISTINCTIVE FEATURES); (iii) constraints and rules may be appealed to in order to provide linguistic

[8] This comparison between Ls and Ms and Cs and Vs is not meant to imply that these categories have similar properties or distribution in the two modalities, because they do not. Rather, the comparison is much broader, reflecting the fact that these phonological categories manifest different rhythmic properties.

Kooij, Els van der. 2002. Phonological categories in Sign Language of the Netherlands: The role of phonetic implementation and iconicity. Ph.D. dissertation, University of Leiden.
Liddell, Scott K. 1984. THINK and BELIEVE: Sequentiality in American Sign Language. *Language* 60. 372–399.
Liddell, Scott K. 2003. Sources of meaning in ASL classifier predicates. In Emmorey (2003), 199–220.
Liddell, Scott K. & Robert E. Johnson. 1989. American Sign Language: The phonological base. *Sign Language Studies* 64. 197–277.
McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12. 373–418.
McNeill, David. 1992. *Hand and mind*. Chicago: Chicago University Press.
Meir, Irit. 2002. A cross-modality perspective on verb agreement. *Natural Language and Linguistic Theory* 20. 413–450.
Meir, Irit & Wendy Sandler. 2008. *A language in space: The story of Israeli Sign Language*. Mahwah, NJ & New York: Lawrence Erlbaum/Taylor & Francis.
Meir, Irit, Mark Aronoff, Wendy Sandler & Carol Padden. Forthcoming. Sign languages and compounding. In Sergio Scalise & Irene Vogel (eds.) *Compounding*. Amsterdam & Philadelphia: John Benjamins.
Nespor, Marina & Wendy Sandler. 1999. Prosody in Israeli Sign Language. *Language and Speech* 42. 143–176.
Padden, Carol. 1988. *Interaction of morphology and syntax in American Sign Language*. New York: Garland. (1983 Ph.D. dissertation, University of California, San Diego.)
Padden, Carol, Irit Meir, Mark Aronoff & Wendy Sandler. 2010. The grammar of space in two new sign languages. In Diane Brentari (ed.) *Sign languages*, 570–592. Cambridge: Cambridge University Press.
Perlmutter, David. 1992. Sonority and syllable structure in American Sign Language. *Linguistic Inquiry* 23. 407–442.
Quer, Josep (ed.) 2008. *Signs of the time: Selected papers from TISLR 2004*. Hamburg: Signum.
Reilly, Judy, Marina McIntire & Ursula Bellugi. 1990. The acquisition of conditionals in American Sign Language: Grammaticized facial expressions. *Applied Psycholinguistics* 11. 369–392.
Sandler, Wendy. 1986. The spreading hand autosegment of American Sign Language. *Sign Language Studies* 50. 1–28.
Sandler, Wendy. 1989. *Phonological representation of the sign: Linearity and non-linearity in American Sign Language*. Dordrecht: Foris.
Sandler, Wendy. 1990. Temporal aspect and American Sign Language. In Susan Fischer & Patricia Siple (eds.) *Theoretical issues in sign language research*, vol. 1, 103–129. Chicago: University of Chicago Press.
Sandler, Wendy. 1993. A sonority cycle in American Sign Language. *Phonology* 10. 243–279.
Sandler, Wendy. 1996. Phonological features and feature classes: The case of movements in sign language. *Lingua* 98. 197–220.
Sandler, Wendy. 1999. Cliticization and prosodic words in a sign language. In T. A. Hall & Ursula Kleinhenz (eds.) *Studies on the phonological word*, 233–254. Amsterdam & Philadelphia: John Benjamins.
Sandler, Wendy. 2009. Symbiotic symbolization by hand and mouth in sign language. *Semiotica* 174. 241–275.
Sandler, Wendy. Forthcoming a. Prosody and syntax in sign language. *Transactions of the Philological Society*.
Sandler, Wendy. Forthcoming b. The uniformity and diversity of language: Evidence from sign language. *Lingua*.
Sandler, Wendy & Diane Lillo-Martin. 2006. *Sign language and linguistic universals*. Cambridge: Cambridge University Press.

Sandler, Wendy, Irit Meir, Carol Padden & Mark Aronoff. 2005. The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences* 102. 2661–2665.

Sandler, Wendy, Mark Aronoff, Irit Meir & Carol Padden. Forthcoming. The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory*.

Schembri, Adam. 2003. Rethinking "classifiers" in signed languages. In Emmorey (2003), 3–34.

Shepard-Kegl, Judy. 1985. Locative relations in ASL word formation, syntax and discourse. Ph.D. dissertation, MIT.

Stokoe, William C. 1960. *Sign language structure: An outline of the visual communication systems of the American deaf.* Silver Spring, MD: Linstok Press.

Stokoe, William C., Dorothy Casterline & Carl Croneberg. 1965. *A dictionary of American Sign Language on linguistic principles.* Washington, DC: Callaudet College Press. Reprinted 1976, Silver Spring, MD: Linstok Press.

Supalla, Ted. 1982. Structure and acquisition of verbs of motion and location in American Sign Language. Ph.D. dissertation, University of California, San Diego.

Supalla, Ted. 1986. The classifier system in American Sign Language. In Collette Craig (ed.) *Noun classes and categorization*, 181–214. Amsterdam: John Benjamins.

Supalla, Ted & Elissa Newport. 1978. How many seats in a chair? The derivation of nouns and verbs in American Sign Language. In Patricia Siple (ed.) *Understanding language through sign language research*, 91–132. New York: Academic Press.

Swerts, Marc & Emil Krahmer (eds.) 2009. Special issue on visual prosody. *Language and Speech* 2.

Wilbur, Ronnie. 1987. *American Sign Language: Linguistic and applied dimensions*. Boston, MA: College Hill Press.

Wilbur, Ronnie. 1999. Stress in ASL: Empirical evidence and linguistic issues. *Language and Speech* 42. 229–250.

Wilbur, Ronnie. 2008. Complex predicates involving events, time and aspect: Is this why sign languages look so similar? In Quer (2008), 217–250.

Woll, Bencie. 2001. The sign that dares to speak its name: Echo phonology in British Sign Language. In Boyes Braem & Sutton-Spence (2001), 87–98.

Zwitserlood, Inge. 2003. Classifying hand configurations in Nederlandse Gebarentaal (Sign Language of the Netherlands). Ph.D. dissertation, Utrecht University.

25 Pharyngeals

Kimary Shahin

1 Introduction

Pharyngeals are segments with primary or secondary articulation in the pharynx. They are sometimes referred to as "post-velars" (e.g. Bessell and Czaykowska-Higgins 1991; Bessell 1992). It is usually understood that the uvula is the topmost structure of the pharynx, and that the pharynx extends to, but excludes, the glottis (e.g. Carmody 1941 and Maddieson 2009, who states that the upper boundary is "around the uvula"). The term"pharyngeals" has been used in a second sense, to refer to segments, like [ħ ʕ], which are "pharyngeal" according to the IPA classification of places of articulation. They have been presumed to be articulated above the epiglottis and below the uvula. However, this is a misnomer because the place of articulation of those sounds is actually aryepiglottal, as will be explained (see CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION for more general discussion).

Figure 25.1 provides a general idea of the geographic distribution of pharyngeal consonants, including the pharyngeal glide but excluding non-uvular rhotics like [r ɹ]. Figure 25.1 is based on the observations in this chapter, and the data of Maddieson (2005a, 2005b) and Chambers and Trudgill (1998). Regarding non-uvular rhotics, pharyngeal constriction has been observed for English [r ɹ] (e.g. McMahon *et al.* 1994; Scobbie 2009). However, identification of all languages with a pharyngeally articulated non-uvular rhotic is left for work elsewhere, and the areas for English are not marked in Figure 25.1.

Pharyngeal vowels, i.e. "low back" and "lax"/"–ATR" vowels, as will be explained below, probably have extensive geographic distribution.

Regarding inventory configurations, languages with primary pharyngeal consonants at more than one place of articulation tend to have large consonantal inventories, typically utilizing secondary articulations like lip rounding or glottalization, but not necessarily having a voicing opposition for obstruents. The presence of secondary pharyngeal consonants implies the presence of primary pharyngeal consonants, and it seems that no language with pharyngealized vowel phonemes also has pharyngeal consonant phonemes; see Maddieson (2009). Languages with pharyngeal consonants typically have derived secondary pharyngeal, i.e. "lax," vowels, due to vowel pharyngealization in the environment of a pharyngeal consonant (CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION).

Figure 25.1 Geographical distribution of pharyngeal consonants, except for non-uvular rhotics

The aim of this chapter is to review the main issues in pharyngeal phonology. These can be expressed by the following questions: (i) what is the pharyngeal typology?; (ii) what is the articulation of pharyngeals?; (iii) what is their representation?

The answers to these questions as indicated by research so far will be summarized below. Where data examples are given, they will mostly just illustrate various patternings – for data that establish the patterns, see the sources cited for each pattern. Our discussion will incorporate the findings of Esling (1996, 1999, 2005) on the actual nature of pharyngeal articulation. It will also incorporate the findings of Miller-Ockhuizen (2003) regarding a superordinate class that includes both pharyngeals and glottals. Finally, although some comment will be made on the phonological representation of non-pharyngealized ("+ATR") vowels, *vis-à-vis* the phonological representations of pharyngeal segments, +ATR and −ATR vowel patternings (e.g. Casali 2008 and references therein) will not be illustrated, as full integration of Niger-Congo-type tongue root vowel harmony systems into an overall characterization of pharyngeal phonology is beyond the scope of this chapter.

All in all, it is an exciting time to be researching pharyngeals and, as we will see, there remains much work to keep phonologists, phoneticians, fieldworkers, and dialectologists busy for a long time to come.

2 Pharyngeal typology

Segments for which the pharyngeal articulation is primary have often been referred to as "gutturals" (especially for Semitic). I will refer to them with the more transparent term "primary pharyngeals." They include [q ɢ ʛ ɴ ʀ χ ʁ ħ ʕ ʜ ʢ ʡ]

Table 25.1 Co-occurrence of laryngeal consonants and vowels in Ju|'hoansi (from Miller-Ockhuizen 2003: 142, table XXIII). Copyright © 2003. From *The phonetics and phonology of gutturals: A case study from Ju|'hoansi*. By Amanda Miller-Ockhuizen. The data are here reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc

			V1				
			Non-laryngeal vowels	*Laryngeal vowels*			
			Modal	*Breathy*	*Glottalized*	*Epi-glottalized*	*Total*
C1	Non-laryngeal consonants	Voiceless	265	70	32	84	451
		Voiced	203	71	31	65	370
		Nasal	155	25	23	70	273
	Laryngeal consonants	Aspirated	254	2	1	11	268
		Ejective	169	0	2	0	171
		Uvularized	214	0	0	0	214
		Epiglottalized	132	0	0	0	132
		Total	1392	168	89	230	1879

The Ju|'hoansi glottals are /ɦ/, a set of aspirated consonants, a set of ejective consonants, a set of breathy vowels, and a set of glottalized vowels.[1]

The superordinate class in Ju|'hoansi excludes modally voiced segments. (This is interesting but will not be accounted for here.) In Ju|'hoansi, all the pharyngeals and glottals pattern together in a morphemic Obligatory Contour Principle (OCP) constraint: in root morphemes, none of the pharyngeal or glottal consonants can co-occur with any of the pharyngeal or glottal vowels. As it turns out, this is precisely as predicted by Esling's (2005) clarification of pharyngeal phonetics. On the basis of numerous data from several languages, he clarifies that the structures of the pharynx and the larynx are controlled by one articulatory structure: the "laryngeal constrictor mechanism." Miller-Ockhuizen refers to the superordinate class of pharyngeals and glottals as "gutturals." However, since "gutturals" is already established as a term for primary pharyngeals, I will refer to the class as "laryngeals," in the interest of descriptive transparency. In this terminology, "glottal" and "laryngeal" are not synonymous.

Miller-Ockhuizen's data showing the laryngeal OCP in Ju|'hoansi are presented in Table 25.1. As she explains (2003: 140), the OCP constraint holds over 1,900 distinct monomorphemic root morphemes, with only sixteen exceptions.

2.1 *The pharyngeal class*

The first type of evidence for a pharyngeal class comprising primary and secondary pharyngeals is from Arabic. In some Arabic dialects, including Northern

[1] A [ʔ] in the language is apparently an allophonic or phonetic effect (Miller-Ockhuizen 2003: 115).

Palestinian and Syrian, the feminine suffix occurs as [-i] or [-e], except when it is immediately preceded by a primary or secondary pharyngeal (Herzallah 1990). This is discussed as evidence for the overall pharyngeal class by McCarthy (1994), who illustrates it with Syrian data from Cowell (1964) and Grotzfeld (1965).[2] Some of his examples are:

(2)	ʔɔsːˤa	*ʔəsːˤe	'story'
	mniːħa	*mniːħe	'good'
	tˤabχa	*tˤabχe	'cooking'

A second type of evidence comes from cases of vowel pharyngealization. Maddieson (2009) explains what a pharyngealized vowel is when he states:

> ATR vowels are produced with a wider than expected aperture [between the tongue root and the back wall of the pharynx]. Vowels with a narrower than expected constriction can be labeled pharyngealized.

Thus, "lax" ("−ATR"/"RTR") vowels, written with IPA symbols like [ɪ ɛ ʊ ɔ], are pharyngealized (see also CHAPTER 21: VOWEL HEIGHT). Vowels are typically pharyngealized when they are adjacent to a primary or secondary pharyngeal. (Shahin 2002: 98 takes this as evidence that the feature that represents pharyngealization is optimally vocalic.) This is observed in Arabic (Gairdner 1925; Jakobson 1957; Shahin 2002) and in St'át'imcets (Interior Salish) (van Eijk 1997; Shahin 2002). It is illustrated for Palestinian Arabic in (3) and for St'át'imcets in (4).[3] In St'át'imcets, the immediately preceding vowel is affected. (In both languages, the vowel pharyngealization shows domain effects, and neutrality effects are observed in the Arabic.)

(3)	ħɪlu	*ħilu	'pretty'
	ʕʊlə	*ʕulə	'Ula (girl's name)'
	ðˤorˤə	*ðˤurˤə	'corn'

(4)	ʃʧʊʁʷ	*ʃʧuʁʷ	'stripe'
	mɪxˤæɬ	*mixˤæɬ	'black bear'
	ʧɜʁɘn	*ʧɘʁɘn	'to rip, tear something (TRANS)'

Bellem (2007: 58) suggests that vowel pharyngealization in Arabic is a phonetic effect. However, it is discrete (CHAPTER 89: GRADIENCE AND CATEGORICALITY IN PHONOLOGICAL THEORY). It is also grammatically visible, as it is sensitive to word-internal morphological structure (see Shahin 2002: 114–115). This means it is phonological. Bellem also suggests that it is an effect of vowel length, but this is not so because short vowels are not pharyngealized in forms parallel to those in (3) but without the pharyngeal consonants (see Shahin 2002: 181).

[2] The phonemic vowel inventory of colloquial Levantine Arabics like Syrian and Palestinian (including Bedouin) is /I: E: Æ: O: U: I E Æ O U/, where capital letter symbols encode the fact that the vowel will be realized as non-pharyngealized or pharyngealized ("tense" or "lax"), depending on the nature of the surface form in which it appears; see Shahin (2003) for discussion, as well as evidence that, at least in Palestinian, /E/ is actually /ɛ/.

[3] The phonemic vowel inventory of St'át'imcets is /I Æ U/. Capital letter symbols are used here as explained in fn. 2. The schwa, which has various realizations, including non-pharyngealized [ɘ] and pharyngealized [ɜ], is epenthetic.

(10)	ʔisʕatte	*ʔisʕɛtte	'lice (GEN)'
	t'Arʕ'ætte	*t'Arʕ'ɛtte	'tapeworms (GEN)'

Another pattern is avoidance of primary pharyngeals in coda position. This is observed in various Semitic languages. It is illustrated for Negev Palestinian Arabic and Tiberian Hebrew in (11) and (12), respectively.[10] These data are from McCarthy (1994: 214–215). (See McCarthy 1994 for data showing the avoidance in Hijazi Bedouin Arabic and Tigre.) In both Negev Arabic and Tiberian Hebrew, a vowel is epenthesized so the primary pharyngeal is realized in onset instead of coda position.

(11)	jahardʒ	*jahradʒ	'he speaks'
	ahalam	*ahlam	'I dream'
	bnaʁazil	*bnaʁzil	'we spin'

(12)	jahăpoːk	*jahpoːk	'he will turn'
	jeħĕzaq	*jeħzaq	'he is strong'
	joʕŏmad	*joʕmad	'he is made to stand'

Degemination of primary pharyngeals also occurs, in Tiberian Hebrew and Tigre. This is illustrated in (13) and (14), respectively, with data from McCarthy (1994: 216–217).[11] Compensatory lengthening (CHAPTER 64: COMPENSATORY LENGTHENING) occurs for the forms in (13).

(13)	meːʔeːn	*meʔʔeːn	'he refused'
	jeːħat	*jeħħat	'he marches down'
	raːʕiːm	*raʕʕiːm	'evil ones'

(14)	təbaʔasa	*təbaʔʔasa	'he quarreled'
	ləsˤʕən	*ləsˤaʕʕən	'he loads'

Other effects are also observed. For example, in Jibbāli/Shehri (South Arabian) "Conjugation B" verbs, a syllable with a primary pharyngeal onset attracts stress, and vowels flanking the primary pharyngeal are identical (Hayward *et al.* 1988; Hayward and Hayward 1989). Examples are seen in (15), from Hayward and Hayward (1989: 182).[12]

(15)	j'sɔ.'ʕɔf	*j'sɔ.ʕɔf	*j'se.'ʕɔf	'remove husks (IMPERF)'
	j'rɔ.'χɔs'	*j'rɔ.χɔs'	*j're.'χɔs'	'become cheap (IMPERF)'
	j'ðˤɔ.'hɔr	*j'ðˤɔ.hɔr	*j'ðˤe.'hɔr	'be finished (IMPERF)'

[10] The vowel phonemes of Tiberian Hebrew are /i e ɛ a ɔ o u/ (Malone 1993). In (12), the inserted vowel is schwa, with various colorings. In Negev Arabic, the effect is limited to cases where the primary pharyngeal follows [a].

[11] Rose (1996) states that the Tigre vowel phonemes are the same as Tigrinya's (see fn. 7).

[12] The Jibbāli vowel inventory is /i e ɛ a ə ɔ o u/ (Johnstone 1981).

2.3 *The secondary pharyngeal class*

Some patternings involve secondary pharyngeals and exclude primary pharyngeals. They typically involve domain, directionality, or neutrality effects, which can be quite complex.[13] Due to space limitations, none of those effects will be addressed here. (See CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS.) The most extensively studied of these patternings is "emphasis spread" in Arabic (e.g. Harrell 1957; Younes 1982; Card 1983; Herzallah 1990; Davis 1995; Zawaydeh 1998; Watson 1999; Shahin 2002; Bin-Muqbil 2006). In Arabic, otherwise non-secondary pharyngeal consonants are pharyngealized, and the (short or long) low vowel is realized as [ɑ(ː)], when a secondary pharyngeal consonant occurs elsewhere in the word. Examples, adapted from Davis (1995: 473–474), are seen in (16), in which the underlying secondary pharyngeals are underlined. The underlying secondary pharyngeal set includes /kˤ/ (frequently phonetically realized as [q]) in several Arabic varieties, as seen from forms like [kˤɑːfˤ] 'the letter *qaf*' *vs.* [kæːf] 'the letter *kaf*'.

(16)	bˤɑlːˤɑːsˤ	*bælːæːsˤ	'thief'
	χˤɑjːˤɑːtˤ	*χæjːæːtˤ	'tailor'
	tˤwˤɑːlˤ	*tˤwæːl	'long (PL)'

Interior Salish "retraction harmony" (Bessell 1992, 1998; Doak 1992; van Eijk 1997; Shahin 2002) is similar. (The studies, other than Shahin's, mistakenly conflated this pattern with the vowel pharyngealization illustrated in (4).) St'át'imcets examples are presented in (17). In this language, the only consonant affected is /ʧ/, and the low vowel and epenthetic schwa are realized as [ɑ] and [ʌ], respectively. Affected segments are typically left adjacent to the pharyngealized consonant, although rightward cases seem to exist (Henry Davis, personal communication; Shahin 2010).

(17)	mʌkˤ'	*ɪnɜkˤ'	'to get stuffed, to eat too much'
	ʃkˤjɑxˤ	*ʃkˤjæxˤ	'drunk'
	ʧˤkˤʷænæʔ	*t̠kˤʷænæʔ	'lynx'
	ʔɑlˤʃˤɘm	*ʔælˤʃˤɘm	'sick, ill'

In Tsilhqot'in "flattening," the vowels /i ɪ u ʊ æ ɛ/[14] are realized as [ə̰i/e ə̰ɪo ɔ a ʌ] in the context of secondary pharyngeals. No consonants are affected (see Cook 1993; Hansson 2007). The pattern is illustrated in (18) with data from Cook (1993: 154).[15]

(18)	sˤə̰it	*sˤit	'kingfisher'
	ʧ'ɔxˤ	*ʧ'ʊxˤ	'porcupine'
	hegˤat	*higˤæt	'we shake it out'

[13] See, for example, Hansson (2007) on Tsilhqot'in.

[14] These six vowels are the underlying vowel set in this language (Cook 1993).

[15] I present Cook's data in IPA. Tsilhqot'inists typically transcribe the secondary pharyngeal velars as simple uvulars.

Finally, in Northeast and Northwest Caucasian languages, vowels are reported as lowered, centralized, or fronted in the environment of secondary pharyngeals (Catford 1983; Anderson 1997; Bellem 2005, 2007). No phonological data are available on this except for Kibrik and Kodzasov (1990, cited by Bellem 2005) in Russian. So here we reissue the call of McCarthy (1994) and Anderson (1997) for phonological data for these languages.

2.4 *The dorsal pharyngeal class*

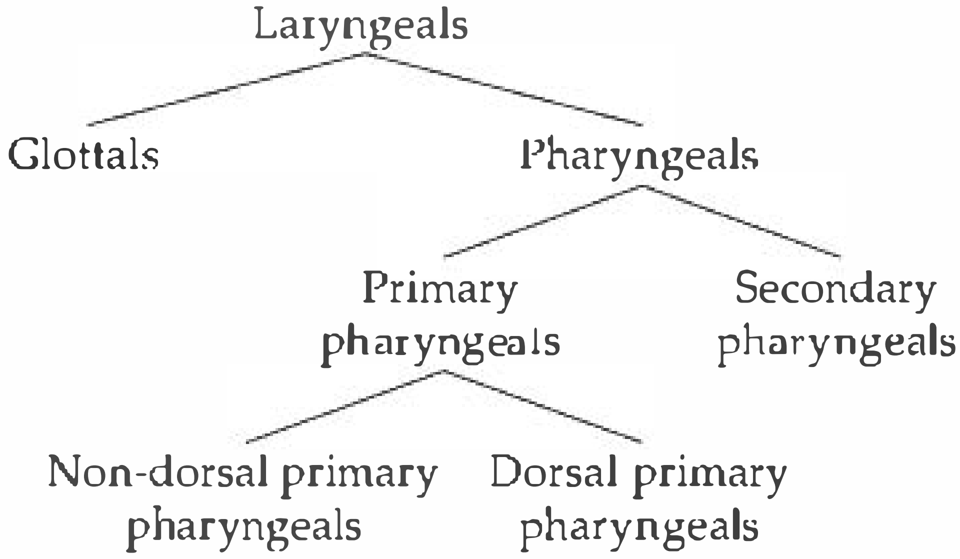
Dorsal primary pharyngeals (i.e. simple uvulars) and secondary pharyngeals can pattern as a class. In Palestinian Arabic, /i/ occurs as backed [u] in certain verb forms containing a root consonant from either group (Herzallah 1990; McCarthy 1994). This occurs most clearly in biliteral roots, as illustrated for typical rural dialects in (19).[16]

(19)	bˤɪsˤʊfːˤ	*bˤɪsˤɪfːˤ	'he lines up'
	bˤɪðˤʊmːˤ	*bˤɪðˤɪmːˤ	'he annexes'
	bɪʃʊχː	*bɪʃɪχː	'he urinates/defecates'

2.5 *Summary, and the issue of glottal primary pharyngeals*

The phonological data reviewed above support the pharyngeal typology in (20). The superordinate laryngeal class is included in the figure.

(20) *The pharyngeal typology, as a subgroup of the laryngeal class*



The superordinate class and its two subclasses are recognized in Halle (1992: 211), which presents a feature geometry with a class of laryngeals comprised of glottal and tongue root subclasses.

We now address the issue of glottal primary pharyngeals. In certain languages, as was illustrated in §2.2, glottal /ʔ h/ pattern as though they are primary pharyngeals. In other languages, for example, St'át'imcets and Tigre (Bessell 1992;

[16] The environment for this backing excludes non-pharyngeal /w/. For example, we have [wˤɪsˤɪl'] 'he arrived'. Interestingly, in Rwali Bedouin Arabic (Parkinson 1992: 114), /w/ is not excluded. In Rwali the word for 'he arrived' is [wʊsˤɪl].

McCarthy 1994; Shahin 2002), they do not. (It has sometimes been thought that glottals that are not primary pharyngeals can condition vowel lowering. This is assumed, for example, by Rose (1996), from supposed cases of lowering in Interior Salish. However, Montler (2004) and Shahin and Blake (2004) have shown that the lowering in those cases is a low-level phonetic effect.) The task has been to explain how glottals can be primary pharyngeals in some languages and not in others (Lindqvist-Gauffin 1969; Trigo 1991; McCarthy 1994; Nolan 1995; Zawaydeh 1999, 2004; Lombardi 2001; Shahin 2002, forthcoming; Borroff 2007). Most proposed solutions presume different representations for the two types of glottal. (Or, within Articulatory Phonology, they presume two types of gesture, e.g. Borroff 2007.) But the phonetic grounding (Archangeli and Pulleyblank 1994: 172) of the different representations has been unclear. Lombardi (2001) proposes that they have the same representation, but the difference in behavior falls out of optimality-theoretic constraint ranking.

No real solution to this problem has been possible without analysis of natural language articulatory data on glottal primary pharyngeals. This is presented by Shahin (forthcoming), using the Laufer and Baer (1988) Hebrew data, and data from Arabic. The analysis indicates that Hebrew and Arabic glottal primary pharyngeals are simply glottal, and lack the aryepiglottic constriction of the other primary pharyngeals. Shahin proposes that the glottals bear the feature of all primary pharyngeals, and so they pattern as primary pharyngeals, but that, at least in Arabic, a paradigmatic constraint forces phonetic non-realization of that feature.

3 Pharyngeal articulation

The first descriptions of pharyngeal articulation, beginning notably with the eighth-century descriptions of Sibawayh (Sibawayh 1966), were based on impressionistic data. Instrumental analysis, including X-ray,[17] MRI, ultrasound, electroglottography, electromagnetic articulography, velopharyngeal port transillumination, and study of jaw movement, has revealed much about the actual vocal tract configurations of these sounds. Likewise, acoustic analysis has shown much about the acoustic effects of the configurations. However, the application of laryngoscopy has permitted us to see behind the epiglottis. This has enabled a critical clarification of the fundamental nature of pharyngeal articulation (Esling 1996, 1999), namely that constriction of the aryepiglottic folds is the articulatory commonality of the overall pharyngeal class.

3.1 *The articulation of primary pharyngeal consonants*

Early laryngoscopic investigations indicated epiglottal activity for pharyngeal consonants (e.g. Laufer and Condax 1979, 1981; Laufer and Baer 1988). Pursuing their line of work, Esling (1996, 1999) has shown that the primary constriction of all consonants traditionally classified as pharyngeal or epiglottal is produced between constricting aryepiglottic folds as active articulator and the epiglottis

[17] Early X-ray studies include Panconcelli-Calzia (1920–1) and Marçais (1948), discussed by Jakobson (1957).

Czaykowska-Higgins 1987; McCarthy 1988; Hess 1998; Zawaydeh 1999, 2003, 2004; Zeroual *et al.* 2009).

However, the findings are rather inconclusive. The aryepiglottic constriction has not always been detected (Zeroual 2004). As stressed by Bellem (2007), extensive cross-linguistic articulatory study is needed. This should include canonical productions of the secondary pharyngeals, such as occur, for Arabic, during Qur'anic recitation.

Secondary pharyngeals in Northeast and Northwest Caucasian languages are described as being produced with pharyngealization (Hess 1992) and also tongue body advancement, or palatalization (Trubetzkoy 1931; Comrie 2005; Bellem 2009). This means they are articulatorily different from the secondary pharyngeal consonants of Arabic and Interior Salish.

Secondary pharyngeal consonants are often labialized (Mitchell 1960; Lehn 1963; Uldall 1992; Watson 1999). They might involve increased muscular tension in the mouth and throat (Lehn 1963; Bonnot 1977).

Secondary pharyngeal consonants have a lowered F2 (Obrecht 1968; Ghazeli 1977; Latimer 1978; Woldu 1981; Younes 1982; Card 1983; El-Dalee 1984; Bessell 1992; Shahin 2002). They also have a raised F1 (Bonnot 1977, 1979; Woldu 1981; Younes 1982; Shahin 2002). In Arabic at least, they have lowered pitch compared to their plain counterparts (Jakobson 1957). Vowels in the environment of secondary pharyngeal consonants in Northeast and Northwest Caucasian have been noted to have a raised F1, lowered F2, and lowered F3 (Bellem 2009).

3.3 *The articulation of pharyngeal vowels*

The vowel [ɑ] is a primary pharyngeal segment, as are [ə ʌ ɔ ɒ] (Esling 2005). This has often been noted, especially for [ɑ] (e.g. Delattre 1971; Whalen *et al.* 1999; Jackson and McGowan 2008). The primary pharyngeal vowels are produced with aryepiglottic constriction (Esling 2005). "Lax" vowels like [ɪ ɛ ʊ ɔ] are secondary pharyngeals, involving aryepiglottic constriction in addition to their primary oral articulation. For other types of secondary pharyngeal vowels, like the constricted tense register vowels of Yi (Tibeto-Burman), secondary aryepiglottic constriction has been observed (Esling 2005). Pharyngeal vowels in Northeast and Northwest Caucasian are often centralized (Bellem 2007).

Primary and secondary pharyngeal vowels have a raised F1; the plain counterparts of the secondary pharyngeal vowels do not have this (Lindau 1978; Hess 1992; Fulop *et al.* 1998; Shank and Wilson 2000b; Gick *et al.* 2006; Kenstowicz 2009). Pharyngeal vowels might also be distinguishable by F1 bandwidth or spectral tilt (see Hess 1992 and discussion in Casali 2008).

3.4 *The articulation of the superordinate laryngeal class*

The aryepiglottic constriction of pharyngeals results from the action of the aryepiglottic constrictor mechanism described by Esling (2005). (This mechanism is also referred to by Esling 2005 as "laryngeal constrictor," "laryngeal constrictor mechanism," and "laryngeal articulator.") Numerous articulatory data reported in Esling (1996, 1999, 2005), among others, have shown that this same mechanism is responsible for glottal gestures, too. This is explained by Esling (2005: 17–18):

For secondary pharyngeals, Card (1983) proposed [F2 drop]. But most phonologists have assumed that, like dorsal primary pharyngeals, they bear the feature [dorsal] (e.g. McCarthy 1994; Shahin 2002; Bin-Muqbil 2006; Gallagher 2007), eschewing the acoustically based feature. For the secondary pharyngeal consonants, [dorsal] and [pharyngeal] are both specified as a secondary articulation feature. A secondary pharyngeal consonant is also specified for some other feature that represents its primary articulation, e.g. [labial], [coronal], or [dorsal]. (Various proposals exist for the representation of secondary articulation features; see Hume 1992; Clements and Hume 1995; Shahin 2002: 16–17; also CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION.) Bellem (2007) points out that the representation of secondary pharyngeals might be different across different dialects for the same language.

Primary pharyngeal vowels are specified like the non-dorsal primary pharyngeals, and also bear specification that identifies them as vowels. Secondary pharyngeal vowels are specified for [pharyngeal] as a secondary articulation.

Let us reconsider [ATR]. In Stewart's (1967) approach to Niger-Congo-type tongue root harmonies, vowels that are tense are [+ATR], and those that are lax are [−ATR]. The fact that [−ATR] vowels and [pharyngeal] vowels are the same indicates that [pharyngeal] is actually a binary feature, and that [+pharyngeal] is specified on consonants and vowels, and [−pharyngeal] is specified only on [+ATR] vowels.

It has been suggested that [−pharyngeal] vowels are specified for [lowered larynx] (Meechan 1992); however, a feature representing laryngeal height does not at present seem necessary in addition to [pharyngeal].

Within a Government Phonology approach (Kaye *et al.* 1985, 1990; Harris 1990, 1994; Harris and Lindsey 1995), the primary and secondary pharyngeals can be understood as all specified for [A] (Bellem 2007), an element that is realized phonetically as "pharyngeality" for consonants and as "*a*-color" for vowels. In Government Phonology, the element [I] is realized as "palatality" for consonants and "*i*-color" for vowels. Bellem uses [I] with [A] to characterize the "emphatic palatalization" of Caucasian secondary pharyngeals.

5 Conclusion

An understanding of pharyngeals in phonology involves, most basically, an understanding of their class and subclasses, as indicated by phonological data. It also involves an understanding of their articulation, and of the representations that are then posited from those representations. This chapter has surveyed what is known about these issues. Pharyngeals and pharyngeal phonology are much better understood than they were twenty years ago, but more work needs to be done. This includes cross-linguistic instrumental and theoretical study of secondary pharyngeal consonants – including, potentially, [r ɹ] – and investigation of all types for the pharyngeals of Northeast and Northwest Caucasian languages. Further work should also include study of Mian (Ok), Iatê (Macro-Ge), Hamer (Omotic), and Even (Tungusic), which are reported to have pharyngealized vowels (Maddieson 2009), but which are poorly represented in the phonology literature. Future research should also identify and test the phonological predictions of Esling's "oral-and-laryngeal" model of speech production.

ACKNOWLEDGMENTS

This chapter has benefited from helpful comments provided by two anonymous reviewers and the *Companion* editors. Language family names used in this chapter are consistent with the *Ethnologue* classification (Lewis 2009).

REFERENCES

Al-Ani, Salman. 1970. *Arabic phonology*. The Hague: Mouton.

Anderson, Gregory. 1997. Lak phonology. In Alan S. Kaye (ed.) *Phonologies of Asia and Africa (including the Caucasus)*, 973–997. Winona Lake, IN: Eisenbrauns.

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Bellem, Alex. 2005. "Pharyngeal" and "pharyngealization" in the Caucasian languages: Phonological correlates vs. phonetic realisations. Paper presented at the 13th Manchester Phonology Meeting.

Bellem, Alex. 2007. Towards a comparative typology of emphatics. Ph.D. dissertation, University of London.

Bellem, Alex. 2009. The "problem" of pharyngealisation and its role in the sound systems of North-East Caucasian languages. Poster presented at the International Workshop on Pharyngeals and Pharyngealisation, Newcastle.

Bessell, Nicola J. 1992. Towards a phonetic and phonological typology of post-velar articulations. Ph.D. dissertation, University of British Columbia.

Bessell, Nicola J. 1998. Local and non-local consonant–vowel interaction in Interior Salish. *Phonology* 15. 1–40.

Bessell, Nicola & Ewa Czaykowska-Higgins. 1991. The phonetics and phonology of postvelar sounds in Moses-Columbia Salish (Nxa'amxcin). Paper presented at the Canadian Linguistics Association Conference, Kingston, Ontario.

Bin-Muqbil, Musaed. 2006. Phonetic and phonological aspects of Arabic emphatics and gutturals. Ph.D. dissertation, University of Wisconsin, Madison.

Bladon, Anthony & Ameen Al-Bamerni. 1982. Nasal coarticulation of pharyngeal and glottal consonants: A deductive account. *Journal of the Acoustical Society of America* 72. 102–104.

Bobaljik, Jonathan D. 1996. Assimilation in the Inuit languages and the place of the uvular nasal. *International Journal of American Linguistics* 62. 323–350.

Bonnot, Jean-François. 1977. Recherche expérimentale sur la nature des consonnes emphatiques de l'arabe classique. *Travaux de l'Institut de Phonétique de Strasbourg* 4. 47–88.

Bonnot, Jean-François. 1979. Étude expérimentale de certains aspects de la gémination et de l'emphase en arabe. *Travaux de l'Institut de Phonétique de Strasbourg* 11. 109–118.

Borroff, Marianne L. 2007. A landmark underspecification account of the patterning of glottal stop. Ph.D. dissertation, Stony Brook University.

Bothorel, André, Péla Simon, François Wioland & Jean-Pierre Zerling. 1986. *Cinéradiographie des voyelles et consonnes du français*. Strasbourg: Publications de l'Institut de Phonétique de Strasbourg.

Brame, Michael K. 1972. On the abstractness of phonology: Maltese ʕ. In Michael K. Brame (ed.) *Contributions to generative phonology*, 22–61. Austin: University of Texas Press.

Bravmann, Max. 1934. Materialen und Untersuchungen zu den phonetischen Lehren der Araber. M.A. thesis, University of Göttingen.

Butcher, Andrew & Kusay Ahmad. 1987. Some acoustic and aerodynamic characteristics of pharyngeals in Iraqi Arabic. *Phonetica* 44. 156–172.

Card, Elizabeth. 1983. A phonetic and phonological study of Arabic emphasis. Ph.D. dissertation, Cornell University.

Carlson, Barry & John H. Esling. 2003. Phonetics and physiology of the historical shift of uvulars to pharyngeals in Nuuchahnulth (Nootka). *Journal of the International Phonetic Association* 33. 183–193.

Carmody, Francis. 1941. An X-ray study of pharyngeal articulation. *University of California Publications in Modern Philology* 21. 377–384.

Casali, Roderic F. 2008. ATR harmony in African languages. *Language and Linguistics Compass* 2/3. 496–549.

Catford, J. C. 1977. Mountain of tongues: The languages of the Caucasus. *Annual Review of Anthropology* 6. 283–314.

Catford, J. C. 1983. Pharyngeal and laryngeal sounds in Caucasian languages. In Diane M. Bless & James H. Abbs (eds.) *Vocal fold physiology: Contemporary research and clinical issues*, 342–350. San Diego: College-Hill Press.

Chambers, J. K. & Peter Trudgill. 1998. *Dialectology*. 2nd edn. Cambridge: Cambridge University Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 245–306. Cambridge, MA & Oxford: Blackwell.

Colarusso, John. 1994. How to describe the sounds of the Northwest Caucasian languages. In Howard Aronson (ed.) *Non-Slavic languages of the USSR*, 61–113. Columbus, OH: Slavica Publishers.

Collis, Dirmid R. F. 1990. *Arctic languages: An awakening*. Paris: United Nations Educational, Scientific and Cultural Organization.

Comrie, Bernard. 2005. Introduction to Caucasian. *Lingua* 115. 1–4.

Connell, Bruce & Amalia Arvaniti (eds.) 1995. *Papers in laboratory phonology IV: Phonology and phonetic evidence*. Cambridge: Cambridge University Press.

Cook, Eung-Do. 1993. Chilcotin flattening and autosegmental phonology. *Lingua* 91. 149–174.

Cowell, Mark W. 1964. *A reference grammar of Syrian Arabic (based on the dialect of Damascus)*. Washington, DC: Georgetown University Press.

Czaykowska-Higgins, Ewa. 1987. Characterizing tongue root behavior. Unpublished ms., MIT.

Davis, Stuart. 1995. Emphasis spread in Arabic and Grounded Phonology. *Linguistic Inquiry* 26. 465–498.

Delattre, Pierre. 1971. Pharyngeal features in the consonants of Arabic, German, Spanish, French, and American English. *Phonetica* 23. 129–155.

Doak, Ivy. 1992. Another look at Coeur d'Alene harmony. *International Journal of American Linguistics* 58. 1–35.

Dolgopolsky, Aharon. 1977. Emphatic consonants in Semitic. *Israeli Oriental Studies* 7. 1–13.

Edmondson, Jerold A., John H. Esling, Jimmy G. Harris & Tsung-chiou Huang. 2005. A laryngoscopic study of glottal and epiglottal/pharyngeal stop and continuant articulations in Amis – an Austronesian language of Taiwan. *Language and Linguistics* 6. 381–396.

Eijk, Jan van. 1997. *The Lillooet language: Phonology, morphology, syntax*. Vancouver: University of British Columbia Press.

El-Dalee, Mohamed. 1984. The feature retraction in Arabic. Ph.D. dissertation, University of Indiana, Bloomington.

Esling, John H. 1996. Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association* 26. 65–88.

Esling, John H. 1999. The IPA categories "pharyngeal" and "epiglottal": Laryngoscopic observations of pharyngeal articulations and larynx height. *Language and Speech* 42. 349–372.

Esling, John H. 2005. There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics* 50. 13–44.

Finch, Roger. 1984. The "emphatic" consonants in the Semitic languages. *Sophia Linguistica* 17. 26–46.

Frisch, Stefan A. & Bushra A. Zawaydeh. 2001. The psychological reality of OCP Place in Arabic. *Language* 77. 91–106.

Fulop, Sean A., Ethelbert Kari & Peter Ladefoged. 1998. An acoustic study of the tongue root contrast in Degema vowels. *Phonetica* 55. 80–98.

Gairdner, W. H. Temple. 1925. *Phonetics of Arabic*. Oxford: Oxford University Press.

Gallagher, Gillian. 2007. Diachronic and synchronic pharyngealization in West Greenlandic. *LSO Working Papers in Linguistics 7: Proceedings of WIGL 2007*, 95–108. Available (October 2010) at http://ling.wisc.edu/lso/wpl/7.

Gerdts, Donna B. & Lisa Matthewson (eds.) 2004. *Studies in Salish linguistics in honour of M. Dale Kinkade*. Missoula: University of Montana.

Ghazeli, Salem. 1977. Back consonants and backing coarticulation in Arabic. Ph.D. dissertation, University of Texas, Austin.

Gick, Bryan, Douglas Pulleyblank, Fiona Campbell & Ngessimo Mutaka. 2006. Low vowels and transparency in Kinande vowel harmony. *Phonology* 23. 1–20.

Gouma, Taoufiik. 2008. L'emphase en arabe marocain, théorie et syllabation. Mémoire de Master II, Université Paris 8.

Gray, Henry. 1918. *Anatomy of the human body*. Available at www.bartleby.com/107/.

Greenberg, Joseph H. 1950. The patterning of root morphemes in Semitic. *Word* 6. 162–181.

Grotzfeld, Heinz. 1965. *Syrisch-Arabische Grammatik (Dialekt von Damaskus)*. Wiesbaden: Harrassowitz.

Halle, Morris. 1988. On the immanent form of phonemes. In William Hirst (ed.) *The making of cognitive science*, 167–183. Cambridge: Cambridge University Press.

Halle, Morris. 1992. Features. In William Bright (ed.) *Oxford international encyclopedia of linguistics*, vol. 3, 207–212. New York: Oxford University Press.

Halle, Morris & Bert Vaux. 1994. Feature spreading and vocalic place. Unpublished ms., MIT & Harvard University.

Hansson, Gunnar Ólafur. 2007. On the evolution of consonant harmony: The case of secondary articulation agreement. *Phonology* 24. 77–120.

Harrell, Richard. 1957. *The phonology of colloquial Egyptian Arabic*. New York: American Council of Learned Societies.

Harris, John. 1990. Segmental complexity and phonological government. *Phonology* 7. 255–300.

Harris, John. 1994. *English sound structure*. Oxford: Blackwell.

Harris, John & Geoff Lindsey. 1995. The elements of phonological representation. In Jacques Durand & Francis Katamba (eds.) *Frontiers of phonology: Atoms, structures, derivations*, 34–79. London: Longman.

Haspelmath, Martin, Matthew S. Dryer, David Gil, Bernard Comrie & Hans-Jörg Bibiko (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press.

Hassan, Zeki. 2009. Some laryngeal and pharyngeal mechanisms observed in the realisation of emphatics and pharyngeal consonants in Iraqi Arabic. Paper presented at the International Workshop on Pharyngeals and Pharyngealisation, Newcastle.

Hassan, Zeki & John H. Esling. 2007. Laryngoscopic (articulatory) and acoustic evidence of a prevailing emphatic feature over the word in Arabic. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 1753–1756. Saarbrücken: Saarland University.

Lewis, M. Paul (ed.) 2009. *Ethnologue: Languages of the world*. 16th edn. Dallas: SIL International. Available at www.ethnologue.com/.

Lindau, Mona. 1978. Vowel features. *Language* 54. 541–563.

Lindqvist-Gauffin, Jan. 1969. Laryngeal mechanisms and speech. *Quarterly Progress and Status Report, Speech Transmission Laboratory, Royal Institute of Technology* 2–3. 26–32.

Lombardi, Linda. 2001. Why Place and Voice are different: Constraint-specific alternations in Optimality Theory. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 13–45. Cambridge: Cambridge University Press.

Maddieson, Ian. 2005a. Uvular consonants. In Haspelmath *et al.* (2005), 30–33.

Maddieson, Ian. 2005b. Presence of uncommon consonants. In Haspelmath *et al.* (2005), 82–85.

Maddieson, Ian. 2009. Typology and occurrence of pharyngeals and pharyngealization around the world. Abstract of paper presented at the International Workshop on Pharyngeals and Pharyngealisation, Newcastle. Available at www.ncl.ac.uk/linguistics/assets/documents/PharyngealsMaddieson.pdf.

Malone, Joseph L. 1993. *Tiberian Hebrew phonology*. Winona Lake, IN: Eisenbrauns.

Marçais, Philippe. 1948. L'articulation de l'emphase dans un parler arabe magrébien. *Annales de l'Institut d'Etudes Orientales d'Alger* 7. 5–28.

Mazaudon, Martine. 1997. A low glide in Marphali. In Roland Bielmeier & Felix Haller (eds.) *Linguistics of the Himalayas and beyond*, 163–188. Berlin & New York: Mouton de Gruyter.

McCarthy, John J. 1985. Features and tiers: The structure of Semitic roots. Paper presented at MIT.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.) *Phonological structure and phonetic form: Papers in laboratory phonology III*, 191–233. Cambridge: Cambridge University Press.

McMahon, April, Paul Foulkes & Laura Tollfree. 1994. Gestural representation and Lexical Phonology. *Phonology* 11. 277–316.

Meechan, Marjory E. 1992. Register in Khmer: The laryngeal specification of pharyngeal expansion. Ph.D. dissertation, University of Ottawa.

Miller, Amanda. 2007. Guttural vowels and guttural co-articulation in Ju|'hoansi. *Journal of Phonetics* 35. 56–84.

Miller-Ockhuizen, Amanda. 2003. *The phonetics and phonology of gutturals: A case study from Ju|'hoansi*. New York: Routledge.

Mitchell, Terence F. 1960. Prominence and syllabification in Arabic. *Bulletin of the School of Oriental and African Studies* 23. 369–389.

Montler, Timothy. 2004. Vowel retraction before glottal stop in Klallam. In Gerdts & Matthewson (2004), 300–310.

Namdaran, Nahal. 2006. Retraction in St'át'imcets: An ultrasonic investigation. M.A. thesis, University of British Columbia.

Nolan, Francis. 1995. The role of the jaw: active or passive? Comments on Lee. In Connell & Arvaniti (1995), 361–367.

Obrecht, Dean. 1968. *Effects of the second formant on the perception of velarization consonants in Arabic*. The Hague: Mouton.

Orser, Brenda. 1985. Stem-initial pharyngeal resonants ʕ, ʕʷ, in Spokane, interior Salish. M.A. thesis, University of Victoria.

Ouni, Slim & Yves Laprie. 2009. Studying pharyngealization using an articulograph. Paper presented at the International Workshop on Pharyngeals and Pharyngealisation, Newcastle.

Panconcelli-Calzia, Giulio. 1920–1. Experimentalphonetische Untersuchungen, *ZES* 11.

Parkinson, Frederick. 1992. The feature pharyngeal in Rwaili Arabic: A case for long distance multiple linking. *OSU Working Papers in Linguistics* 41. 101–134.

Pike, Kenneth L. 1967. Tongue-root position in practical phonetics. *Phonetica* 17. 129–140.
Prunet, Jean-François. 1996. Guttural vowels. In Grover Hudson (ed.) *Essays on Gurage language and culture*, 175–203. Wiesbaden: Harrassowitz.
Pulleyblank, Edwin G. 1998. Pharyngeal glides and zero initials in Chinese. In Benjamin K. Tsou (ed.) *Studia linguistica serica*, 1–26. Hong Kong: City University of Hong Kong.
Pulleyblank, Edwin G. 2003. Non-contrastive features or enhancement by redundant features? *Language and Linguistics* 4. 713–755.
Rischel, Jørgen. 1974. *Topics in West Greenlandic phonology: Regularities underlying the phonetic appearance of wordforms in a polysynthetic language*. Copenhagen: Akademisk Forlag.
Rose, Sharon. 1996. Variable laryngeals and vowel lowering. *Phonology* 13. 73–117.
Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.
Scobbie, James M. 2009. Ultrasound-based tongue root imaging and measurement. Paper presented at the International Workshop on Pharyngeals and Pharyngealisation, Newcastle.
Shahin, Kimary. 2002. *Postvelar harmony*. Amsterdam & Philadelphia: John Benjamins.
Shahin, Kimary. 2003. Vowel innovation in Arabic: inductive grounding and pattern symmetry. In Jacqueline Lecarme (ed.) *Research in Afroasiatic grammar III*, 429–445. Amsterdam & Philadelphia: John Benjamins.
Shahin, Kimary. 2010. Acoustic testing for phonologisation. Unpublished ms., Simon Fraser University.
Shahin, Kimary. Forthcoming. A phonetic study of guttural laryngeals in Palestinian Arabic using laryngoscopic and acoustic analysis. In Barry Heselwood & Zeki Hassan (eds.) *Instrumental studies in Arabic linguistics*. Amsterdam & Philadelphia: John Benjamins.
Shahin, Kimary & Susan J. Blake. 2004. A phonetic study of schwa in St'át'imcets (Lillooet Salish). In Gerdts & Matthewson (2004), 311–327.
Shank, Scott & Ian Wilson. 2000a. Acoustic evidence for /ʕ/ as a glottalized pharyngeal glide in Nuu-chah-nulth. *University of British Columbia Working Papers in Linguistics* 3. 185–197.
Shank, Scott & Ian Wilson. 2000b. An acoustic analysis of vowel formants in pharyngeal and glottal contexts in Nuu-chah-nulth. *University of Washington Working Papers in Linguistics* 19. 75–84.
Sibawayh. 1966. *Al-kitaab*. Beirut: ʕalaam Al-Kutub.
Stewart, John M. 1967. Tongue root position in Akan vowel harmony. *Phonetica* 16. 185–204.
Thalbitzer, William. 1976. *A phonetical study of the Eskimo language*. New York: AMS Press.
Trigo, Loren. 1991. On pharynx–larynx interactions. *Phonology* 8. 113–136.
Trubetzkoy, Nikolai. 1931. Die Konsonantensysteme der ostkaukasischen Sprachen. *Caucasica* 8. 1–39.
Uldall, Elizabeth. 1992. Some observations on the Haskins Laboratories X-ray film of Damascus Arabic. In J. C. Catford, T. Balasubramanian & Vennelakanti Prakasam (eds.) *Sound patterns for the phonetician: Studies in phonetics and phonology in honour of J. C. Catford*, 49–65. Madras: T. R. Publications Private.
Watson, Janet C. E. 1999. The directionality of emphasis spread in Arabic. *Linguistic Inquiry* 30. 289–300.
Whalen, Douglas, A. Min Kang, Harriet S. Magen, Robert K. Fulbright & John C. Gore. 1999. Predicting midsagittal pharynx shape from tongue position during vowel production. *Journal of Speech, Language, and Hearing Research* 42. 592–603.
Wilson, Ian. 2007. The effects of post-velar consonants on vowels in Nuu-chah-nulth: Auditory, acoustic, and articulatory evidence. *Canadian Journal of Linguistics* 52. 43–70.
Woldu, Kiros. 1981. Facts regarding Arabic emphatic consonant production. *Reports from Uppsala University Department of Linguistics* 7. 96–121.

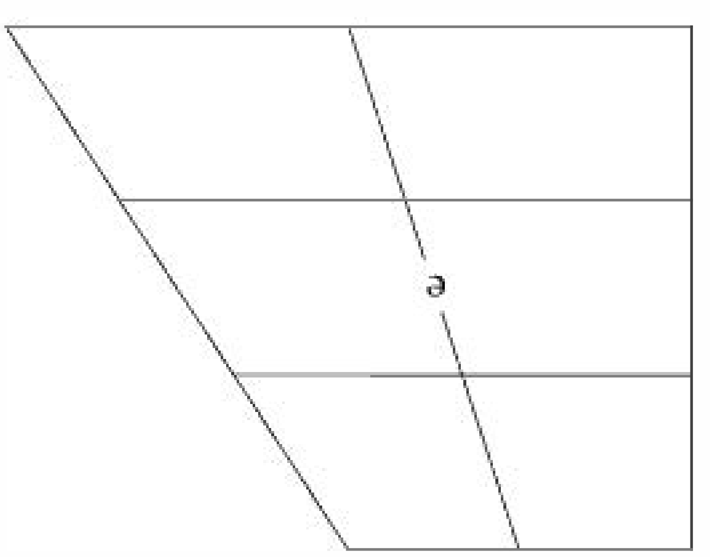
26 Schwa

Daniel Silverman

1 Introduction

Inspection of the chart of the International Phonetic Alphabet (IPA) would suggest that schwa is a vowel like any other; a central open-mid/close-mid unrounded vowel, slightly higher than [ɐ], slightly more central than [ʌ]: i.e. [ə]. Indeed, as the IPA chart necessarily provides idealized phonetic descriptions of its symbols, then, articulatorily speaking, schwa-qua-schwa is just as the IPA chart indicates, as in (1).

(1)



In practice, however, the label "schwa" has been applied to a phonological value that is especially variable in its phonetic properties. In terms of their quality, vowels labeled "schwa" vary to the extent of encompassing a large portion of the vowel space, while tending to gravitate toward the center of this space (see e.g. Browman and Goldstein 1992 for English schwa). This variability is usually a consequence of schwa's context: flanking consonants and vowels may have a significant co-articulatory influence on schwa's phonetic starting and ending postures, typically far more co-articulatory influence than on vowels of other qualities. In terms of duration – a phonetic property that the IPA vowel chart does not indicate – schwa is typically quite short, and this short duration may co-vary with its tendency to be co-articulated.

In acoustic terms, schwa's resonance structure may be modeled (as a first approximation) by a tube that has no significant constrictions anywhere along its length, such that its formants derive not from two tubes (and/or a Helmholtz resonator) as is the case for other vowels, but instead from one long tube (see e.g. Johnson 2003). Given a tube that approximates the length of an adult male vocal tract, this idealized version of schwa possesses formants at 500 Hz, 1500 Hz, 3500 Hz, etc. (Johnson 2003). However, since schwa's articulatory properties are so variable – typically, far more variable than other vowel qualities – its formant values vary in kind.

In phonological terms, schwa has been characterized as a completely underspecified "featureless" vowel by some (e.g. van Oostendorp 1995), and as "weightless" (lacking a mora) by others (e.g. Kager 1989). According to van Oostendorp's approach, for example, Dutch schwa's featureless status may be (i) derived by feature loss, especially when a full vowel reduces to schwa under stresslessness, or (ii) derived by epenthesizing a completely underspecificied vowel. Finally, (iii) its featureless status may be lexical in origin. As discussed by van Oostendorp (1995), Dutch possesses schwas of all three sorts.

Given its short duration and co-articulatory tendencies, schwa bears a phonetic similarity to the mere *audible release* of a consonantal constriction in the context of a following consonant. Indeed, certain epenthetic schwas may have their origin in consonantal release. Although such schwas may play an important functional role by providing acoustic cues to the first consonant in such consonant–consonant sequences, these schwas may, in fact, be "invisible" to the prosodic structure of the language, contributing neither to the syllable structure nor to the metrical structure of the system (Hall 2006).

Moreover, given its short duration and its consequent tendency to camouflage itself to its context through co-articulation, schwa may be confused with its absence, setting up a situation in which schwa–zero alternations may take hold in a system, for example, in Hindi (Ohala 1983).

This chapter is divided into three sections. First, I consider cases of schwa that have their origins in vowel reduction, taking Dutch (e.g. van Oostendorp 1995) and English (Browman and Goldstein 1992; Flemming 2007; Flemming and Johnson 2007) as case studies. Second, I consider cases of schwa that may have their origins in consonantal release, discussing Hall's (2006) findings and taking Indonesian (Cohn 1989) as a case study. Third, I consider two cases of synchronic schwa–zero alternation, one that unambiguously has its origins in (diachronic) schwa deletion (Hindi; Ohala 1983), and one that may be a case of (diachronic) schwa insertion (Chukchee; Kenstowicz 1994).

2 Reduction to schwa

Due to its short duration and its tendency to co-articulate, schwa is a likely outcome of vowel reduction in stressless domains.

2.1 *Reduction to schwa in Dutch*

Schwa in Dutch (Nooteboom 1972; Kager 1989; van Oostendorp 1995; among others) may be characterized as having three sources. Van Oostendorp (1995) calls these e-schwa (epenthetic schwa), r-schwa (reduction schwa), and u-schwa stressless). Most important, schwa's production cannot be accurately characterized as a simple V-to-V interpolation:

> [I]t appears that the tongue position associated with medial schwa cannot be treated simply as an intermediate point on a direct tongue trajectory from V_1 to V_2. Instead, there is evidence that this V_1–V_2 trajectory is warped by an independent schwa component [. . .] a warping of the trajectory toward an overall average or neutral tongue position. (1992: 41, 42)

Thus, the authors conclude that schwa is not genuinely targetless, but rather involves a tendency to gravitate toward the center of the vowel space as the tongue moves from preceding to following context.

In a study involving four American English speakers, Flemming (2007) also investigates the contextual variability of word-medial schwa (as in [səˈpʰoʊz] *suppose* or [ˈpɹ̥abəbl̩] *probable*). His spectrographic analysis shows again that medial schwas are highly variable as a consequence of their context, especially in terms of their F2 properties, the articulatory correlates of which are tongue fronting/backing and lip posture. F1 varies as well – as a consequence of tongue height – but this variation is less pronounced than that found for F2. The limited variation of F1 may be attributable to the fact that flanking consonants necessarily involve a mouth closing/jaw-raising gesture, thus lowering F1.

Overall, Flemming suggests caution in concluding that schwa is completely targetless, even on the F2 dimension, since it is possible that the starting point and endpoint of the F2 trajectory through schwa – despite this trajectory's apparent targetlessness – are influenced by the intervening schwa itself.

Flemming proposes that English schwa's variability is rooted in its short duration. As a consequence of its short duration, vowel quality distinctions are reduced, perhaps to the point of being neutralized. Once neutralized, co-articulation may be engaged in with few limits, as there are no longer any lexical contrasts that might be maintained by inhibiting such co-articulatory tendencies (see especially Öhman 1966 and Manuel 1990, 1999 on the inhibitory role that lexical contrast may play in co-articulation).

Flemming proposes that the origin of American English schwa is rooted in speaker production, invoking the "target undershoot" proposals of Liljencrants and Lindblom (1972). Still, the role of listener perception might be considered as well, since, as vowel duration decreases, the ability to discern quality distinctions is likely to decrease in kind (see, for example, Ohala 1981, Labov 1994, and Silverman 2006).

In sum, stresslessness feeds shortening, shortening feeds contrast loss, and contrast loss feeds co-articulation. Schwa results.

Flemming and Johnson (2007) find that, unlike word-medial schwa, schwa in word-final position (as in [ˈʧʰaɪnə] *china* or [ˈkʰamə] *comma*) displays a relatively consistent mid-central quality, though a certain amount of between-speaker height variation is observed. It is suggested that, since word-final schwa – unlike medial schwa – is in contrast with other vowel qualities (unstressed /i/ and /oʊ/, for example [ˈbeɪɾə̠] *beta*, [ˈpɹɪɾi̠] *pretty*, [ˈmaɾo̠ʊ] *motto*), it tends to be implemented in a more stable fashion so that contrasts here are reliably maintained. Flemming (2007) thus concludes that American English possesses two schwas: word-medial schwa, which is more variable, and word-final schwa, which is more stable.

Another possibility is that there is only one schwa in English, the variability of which is largely a consequence of its lexical context. Within-word motor routines are more frequently produced than are between-word motor routines. As such, they may be more susceptible to fixed co-articulatory effects than are between-word motor routines (see especially Bybee 2001 on the relationship between frequency and co-articulation). Since the context that follows word-final schwa varies in unconstrained ways (depending only on the phonological shape of the following word), its co-articulatory tendencies may be less entrenched, less routinized, than its word-medial counterpart. Consequently, word-final schwa may not as readily possess fixed co-articulatory properties. The result is that word-final schwa may display more stability than its word-medial counterpart, which, in turn, increases the likelihood of schwa maintaining its contrastive status with other vowels here. Thus, whereas Flemming proposes that schwa is more stable in word-final position because it contrasts with other vowel qualities, schwa might instead contrast with more vowel qualities in word-final position because it is more stable.

3 Release into schwa

When lexical or morphological structure brings two consonants together (. . . C_1(+)C_2 . . .), the identity of C_1 may, on occasion, fail to be successfully perceived by a listener. In the absence of consonantal release into a more open gesture – ideally, a vowel – cues that would otherwise be associated with the release of C_1 become jeopardized. Such cues include offset formant transitions and plosive burst frequencies associated with oral place of articulation, and the noise, pop, or low-frequency cues associated with aspiration, ejectivity, or voicing, respectively. In many languages, the loss of such cues sets the stage for merger or neutralization, such that C_1 loses its contrastive oral configuration, its contrastive laryngeal setting, or perhaps both (e.g. Lombardi 1991; Steriade 1997).

However, a very different diachronic route is taken in many other languages: in the context . . . C_1(+)C_2 . . . , C_1 is released into an excrescent vowel before the C_2 constriction is fully achieved, where "excrescent" refers to a particularly short interval of the speech stream that may be an artefactual consequence of repositioning the articulators as they move from one posture to another. Upon the release of C_1, the aforementioned cues that crucially rely on this segment of the speech stream are much more likely to be saliently encoded in the speech signal, and hence C_1 tends to be more resistant to neutralization or merger.

Phonetically speaking, if we assume that the survival of these C_1 cues is dependent in great part upon C_1 release, then this release is likely to become exaggerated over time, culminating in schwa: C_1əC_2. Schwa is short in duration, is subject to significant contextual co-articulation, and is consequently auditorily rather similar to the excrescent vowel of mere release. Hall (2006), for example, finds that inserted schwa and other so-called "intrusive" vowels – vowels that may have their origin in excrescence, but come to bear phonetic similarity to full-fledged vocalism – are more often encountered between heterorganic consonants, as opposed to homorganic consonants (see also CHAPTER 67: VOWEL EPENTHESIS). This is not an unexpected result: C_1C_2 homorganicity decreases the likelihood of C_1 release, since the articulators are likely to maintain their positions as C_1 is followed by C_2 in such homorganic clusters; employing Goldsmith's (1976) consequence of C_1 neutralization, then it might more likely possess C_1 "release," which might diachronically culminate in schwa or a vowel copy that may or may not be prosodically "invisible" (Martinet 1952; Hoenigswald 1960).

Indonesian is a language that seems to possess an "intrusive" schwa in the sense of Hall. As reported by Cohn (1989), monomorphemic Indonesian words possess right-to-left syllabic trochees, end-rule right, with initial dactyls in words with an odd number of syllables, excluding three- and (by necessity) one-syllable words. However, schwas are completely invisible to stress. Examples presented in (6) (Cohn 1989: 174) include both native Indonesian words and assimilated loans. Whether the loans should be characterized as possessing genuinely epenthetic schwa, or perhaps instead involve underlying schwa, remains an open question for now (see especially Silverman 1992 for proposals regarding the interpretation of apparently epenthetic vowels in loan phonology).

(6)	*All full vowels*	*Schwas and full vowels*		
a.	ˈσσ	əˈσ	bəˈri	'give'
			kərˈja	'work'
b.	σˈσσ	əəˈσ	sətəˈlah	'after'
c.		ˈσəσ	ˈgamələn	'Indonesian orchestra'
d.	ˌσσˈσσ	σˈσəσ	aˈpartəmen	'apartment'
e.		əˈσəσ	ʧəˈritəra	'story'
f.		əəˈσσ	pərəmˈpuan	'woman'
g.		σəˈσσ	kopəˈrasi	'cooperation'
h.	ˌσσˌσσˈσσ	ˌσəσσˈσσ	ˌdifərensiˈasi	'differentiation'
			ˌdivərsifiˈkasi	'diversification'

Significantly, Cohn reports that the distribution of schwa is largely predictable, and may thus be viewed as a consequence of epenthesis. In the parlance of Hall, Indonesian schwa is intrusive in nature: it is invisible to stress, and its presence is largely predictable. Note in particular that almost all schwas in (6) are found between heterorganic consonants, at least one of which is a sonorant.

Given its apparent predictability, the origin of Indonesian schwa might indeed lie in consonantal release into an excrescent vowel.

4 Schwa–zero alternations

As noted, schwa is typically short in duration, is subject to significant co-articulation, and, correspondingly, possesses few acoustic features that render it auditorily distinct from its surrounding context. As a consequence of schwa's auditory indistinctness, its *presence* in a given phonetic context may be susceptible to confusion with its *absence* in an otherwise identical phonetic context. When two or more auditory events are confusable with each other, a condition is set up that may lead to diachronic change. One plausible outcome of this particular situation may be the introduction of schwa–zero alternations: in certain acoustic contexts and/or under certain functional conditions schwa may diachronically survive, while in other acoustic contexts and/or under other functional conditions schwa may disappear. The question is: in what situations is the schwa likely to survive, and in what situations is the schwa likely to be lost?

4.1 *Schwa–zero alternation in Hindi*

The Hindi schwa–zero alternation is discussed extensively by Ohala (1983). Consider the forms in (7) (excerpted and modified from her tables 6.1 and 6.3). According to Ohala (1999), the vowel in alternation with zero is actually slightly lower than schwa: [ɐ]. It is transcribed [ə] in Ohala (1983), and herein as well, for the sake of internal consistency. Schwas that alternate and/or vary with zero are underlined.

(7)	a.	pitʃka	'squeezed'	pitʃək	'squeeze'
		pɪgʰla	'melted'	pɪgʰəl	'melt'
		dewrani	'brother-in-law's wife'	dewər	'brother-in-law'
		nəmkɪn	'salty'	nəmək	'salt'
		sɪski	'a sob'	sɪsək	'sob'
		hɪrni	'doe'	hɪrən	'deer'
		təɽpa	'cause to be restless'	təɽəp	'restlessness'
		wapsi	'on return'	wapəs	'return'
		upri	'pertaining to the top'	upər	'top'
		ubʈən	'an unguent'	ubəʈna	'to anoint'
		gərdʒila	'thunderous'	gərədʒna	'to thunder'
	b.	a+səməj	asəməj	'inopportune'	
		ə+ʃərir	əʃərir	'without body'	
		ɔ+kələŋk	ɔkələŋk	'spotless'	
		ku+ʃəkun	kuʃəkun	'bad omen'	

The generalization is that schwa alternates with zero in would-be VCəCV contexts (7a), provided that it is not the first vowel of the morpheme (7b). Ohala's synchronic characterization of this alternation is expressed with the linear rewrite rule conventions of Chomsky and Halle (1968: 121): "ə → Ø / VC __ CV. *Condition*: There may be no morpheme boundary to the left of the /ə/."

The first question to ask is whether the schwa–zero alternation in Hindi is a refllex of historic schwa *insertion* or historic schwa *deletion*. Fortunately, the historical record is very clear on this matter. Hindi schwa derives from Sanskrit [ə] and short [a]. In Old Hindi, this vowel, and some instances of other short vowels ([ɪ] and [ʊ]), alternated with zero in VCVCV: contexts (Misra 1967).

The route from short [a] to the schwa–zero alternation found today may be related to the fact that contrastively short low vowels, due exactly to their contrastively short duration, are quite likely to gradually rise: as a consequence of their attendant jaw lowering, it takes longer to implement low vowels than non-low vowels, and so *contrastively* short low vowels are thus especially susceptible to rising. Upon rising, they become more schwa-like, and are thus susceptible to confusion with their absence in certain contexts. One such context is VCəCV. Provided that phonetic confusion between VCØCV and pre-existing VCCV sequences does not induce undue semantic confusion (by inducing a significant amount of homophony), it is quite possible that the sound pattern may ultimately change from VCəCV to VCCV. This is the Hindi pattern.

Schwa is susceptible to confusion with its absence in many other acoustic contexts as well, such as when it finds itself flanked by more than one consonant on one side, for example VCCəCV and VCəCCV. Nonetheless, schwa usually does not delete in such contexts in Hindi. This may be related to the further possibility of medial C loss here. Were schwa to delete in a VC<u>C</u>əCV or VCə<u>C</u>CV context, the medial consonant might suffer a significant cue loss, as it would lack both approach cues and release cues. Thus, were VCCəCV or VCəCCV to become VC<u>C</u>CV, the sequence might be confused with VCCV. That is, the loss of schwa in these contexts may lead to a percept involving only two – not three – consonants. At this point, the chances of inducing homophony – hence confusion on the part of listeners – increase considerably, since in theory these lost medial consonants may possess many different values, and thus many words may rely on them to maintain their acoustic distinctness. Since speech signals that confuse listeners (as opposed to those that do not confuse listeners) are less likely to be reproduced as these listeners become speakers, the presence of confusing signals as part of the conventionalized speech repertoire may be passively curtailed (Martinet 1952; Labov 1994; Silverman 2006). This may have influenced the present-day Hindi pattern: VCCəCV and VCəCCV do not alternate with VCCV.

There are, however, patterned exceptions to the absence of schwa deletion in VCCəCV and VCəCCV contexts. Consider the forms in (8) (excerpted and modified from Ohala's (1983) table 6.6; underlined schwas vary with their absence).

(8)	kadəmbri ~ kadəmb<u>ə</u>ri	'a novel, name for a girl'
	ustra ~ ust<u>ə</u>ra	'razor'
	puɳɖrik ~ puɳɖ<u>ə</u>rik	'white lotus'
	məɲʤri ~ məɲʤ<u>ə</u>ri	'tiny cluster of flowers, name for a girl'

Hindi possesses variable schwa deletion in these forms, and thus indeed possesses sequences of three consonants that are a consequence of schwa loss. As stated, however, these exceptions are patterned, for as Ohala observes, schwa deletion here results in triconsonantal sequences that are also found elsewhere, and are usually of the form nasal–homorganic stop–sonorant. The development of schwa deletion in such forms is thus not completely unexpected: the phonetic properties of these particular triconsonantal sequences are readily recoverable from the speech signal, since the medial consonant here does not possess place features that are distinct from the preceding nasal, and thus it does not contribute place cues of its own. Consequently, not only are such sequences more likely to be found elsewhere, but also – and perhaps even as a partial consequence of their presence elsewhere – such sequences may more readily enter the language via developments such as schwa deletion. Indeed, as Ohala notes, "[T]hree-consonant clusters in native words are rather few, and in general, most two- and three-consonant clusters that we find at the phonetic level [. . .] occur due to the ə-deletion rule" (1983: 135).

4.2 *Schwa–zero alternation in Chukchee*

Another case of schwa–zero alternation is reportedly found in Chukchee. According to Kenstowicz (1994), Chukchee possesses a synchronic process of schwa insertion under varying phonological and morphological conditions. Consider the forms in (9). In (9a) the schwa precedes the morpheme boundary; in (9b) it follows. (Data

Schwa is often the result of vowel reduction. Its short duration and its consequent tendency to co-articulate make schwa a likely candidate for the vocalism of stressless domains.

Some schwas may have their origins in the audible release of a consonant, when this consonant is immediately followed by another consonant: cues to the phonetic content of consonants are more reliably communicated upon their audible release, ideally into a vowel. These schwas may or may not be visible to the prosodic structure of the language.

Perhaps as a consequence of its tendency to camouflage itself, schwa is especially susceptible to deletion, and thus may alternate with zero under varying conditions.

REFERENCES

Browman, Catherine P. & Louis Goldstein. 1992. "Targetless" schwa: An articulatory analysis. In Gerard J. Docherty & D. Robert Ladd (eds.) *Papers in laboratory phonology II: Gesture, segment, prosody*, 26–56. Cambridge: Cambridge University Press.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Cohn, Abigail C. 1989. Stress in Indonesian and bracketing paradoxes. *Natural Language and Linguistic Theory* 7. 167–216.

Firth, J. R. 1948. Sounds and prosodies. *Transactions of the Philological Society*. 127–152.

Flemming, Edward. 2007. The phonetics of schwa vowels. Unpublished ms., MIT.

Flemming, Edward & Stephanie Johnson. 2007. *Rosa's roses*: Reduced vowels in American English. *Journal of the International Phonetic Association* 37. 83–96.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.

Hall, Nancy. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23. 387–429.

Hayes, Bruce. 1986. Inalterability in CV phonology. *Language* 62. 321–351.

Hoenigswald, Henry M. 1960. *Language change and linguistic reconstruction*. Chicago, IL: University of Chicago Press.

Hooper, Joan B. 1978. Constraints on schwa-deletion in American English. In Jacek Fisiak (ed.) *Recent developments in historical linguistics*. 4th edn., 183–207. The Hague: Mouton.

Johnson, Keith. 2003. *Acoustic and auditory phonetics*. 2nd edn. Malden, MA: Blackwell.

Kager, René. 1989. *A metrical theory of stress and destressing in English and Dutch*. Dordrecht: Foris.

Kenstowicz, Michael. 1994. Syllabification in Chukchee: A constraints-based analysis. *Proceedings of the Formal Linguistics Society of Mid-america* 4. 160–181.

Labov, William. 1994. *Principles of linguistic change*, vol. 1: *Internal factors*. Oxford: Blackwell.

Levin, Juliette. 1987. Between epenthesis and excrescence: What happens after redundancy rules. *Proceedings of the West Coast Conference on Formal Linguistics* 6. 187–201.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48. 839–862.

Lombardi, Linda. 1991. Laryngeal features and laryngeal neutralization. Ph.D. dissertation, University of Massachusetts, Amherst.

Manuel, Sharon Y. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America* 88. 1286–1298.

Manuel, Sharon Y. 1999. Cross-language studies: Relating language-particular coarticulation patterns to other language-particular facts. In W. J. Hardcastle & Nigel Hewlett (eds.) *Coarticulation: Theory, data and techniques*, 179–198. Cambridge: Cambridge University Press.

Martinet, André. 1952. Function, structure, and sound change. *Word* 8. 1–32.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

Miller, J. E. & Osamu Fujimura. 1982. Graphic displays of combined presentations of acoustic and articulatory information. *The Bell System Technical Journal* 61. 799–810.

Misra, B. G. 1967. Historical phonology of Modern Standard Hindi: Proto-Indo-European to the present. Ph.D. dissertation, Cornell University.

Nooteboom, Sieb G. 1972. Production and perception of vowel duration: A study of the durational properties of vowels in Dutch. Ph.D. dissertation, University of Utrecht.

Ohala, John J. 1981. The listener as a source of sound change. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 17(2). 178–203.

Ohala, Manjari. 1983. *Aspects of Hindi phonology*. New Delhi: Motilal Banarsidass.

Ohala, Manjari. 1999. Hindi. *Handbook of the International Phonetic Association*, 100–103. Cambridge: Cambridge University Press.

Öhman, Sven E. G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39. 151–168.

Oostendorp, Marc van. 1995. Vowel quality and syllable projection. Ph.D. dissertation, University of Tilburg.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Schein, Barry & Donca Steriade. 1986. On geminates. *Linguistic Inquiry* 17. 691–744.

Silverman, Daniel. 1992. Multiple scansions in loanword phonology: Evidence from Cantonese. *Phonology* 9. 289–328.

Silverman, Daniel. 2006. *A critical introduction to phonology: Of sound, mind, and body*. London & New York: Continuum.

Skorik, Petr J. 1961. *Grammatika chukotskogo jazyka*, vol. 1. Leningrad: Nauka.

Steriade, Donca. 1997. Phonetics in phonology: The case of laryngeal neutralization. Unpublished ms., University of California, Los Angeles.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.

27 The Organization of Features

Christian Uffmann

1 Introduction

Early generative theory assumed that segments were specified by a bundle of distinctive features, but that there was no intrinsic order to these features, which were unordered lists. This position was not held for any principled reasons; it was born out of necessity, made explicit in *The sound pattern of English* (*SPE*; Chomsky and Halle 1968). In their discussion of distinctive features in chapter 7, Chomsky and Halle group them into sets of articulatorily related features but go on to state that

> This subdivision of features is made primarily for purposes of exposition and has little theoretical basis at present. It seems likely, however, that ultimately the features themselves will be seen to be organized in a hierarchical structure which may resemble the structure that we have imposed on them for purely expository reasons. (Chomsky and Halle 1968: 300)

It was the advent of Autosegmental Phonology (Goldsmith 1976, 1990; see also chapter 14: autosegments) that allowed the imposition of a hierarchical order on distinctive features by grouping features into sets that are characterized by similar articulatory properties and by common participation in phonological processes. These new models of segment-internal feature organization became known as feature geometries.

This chapter will outline the main models of feature geometry that emerged from the mid-1980s onwards and trace the relevant debates regarding specific proposals of feature organization. It will also look at the decline of these models in the wake of Optimality Theory in the mid-1990s and look at alternative proposals regarding how the generalizations that first motivated feature geometry could be captured differently in a constraint-based model of phonology. For reasons of space, theories of feature organization which mark a more radical departure from classical feature theory, such as Element Theory (Harris and Lindsey 1995), will not be discussed here.

The chapter is organized as follows: §2 will motivate the original feature geometry proposal by showing that phonological processes may target not just individual features but sets of features, and that these are cross-linguistically recurring sets. §3 will briefly outline the major models of feature geometry that were proposed and point out the major controversies surrounding them, showing that the early hopes for establishing a uniform and universal feature hierarchy had been overly optimistic. §4 will then outline how phonological processes are captured in feature geometry and how feature geometry provides a principled approach to non-local segment interaction, a problem which had dogged the classic *SPE*-type rule formalism. §5 turns to alternative proposals made within Optimality Theory on how the tendency of features to behave as sets can be re-formalized in non-autosegmental terms, as constraint interaction. §6 will point out future directions of research. First, however, we will motivate the idea that features are ordered into hierarchical sets.

2 Motivating feature sets

Sometimes, a process does not target just a single feature, but several at the same time. This set of features is not random, however; it recurs cross-linguistically, and also has a phonetic or articulatory basis, in that all members of the set share some phonetic property. A well-known and common process is that of nasal place assimilation, in which a nasal stop assimilates in place to a following consonant. The examples in (1) illustrate a case of (post-lexical) nasal place assimilation in English (see CHAPTER 81: LOCAL ASSIMILATION).

(1) *Nasal place assimilation in English*

in Sussex	[ɪn sʌsɪks]	*in Britain*	[ɪm bɹɪtn̩]
in Canada	[ɪŋ kænədə]	*in Europe*	[ɪɲ juːɹəp]
in France	[ɪɱ fɹɑːns]	*in theory*	[ɪn̪ θɪəɹi]

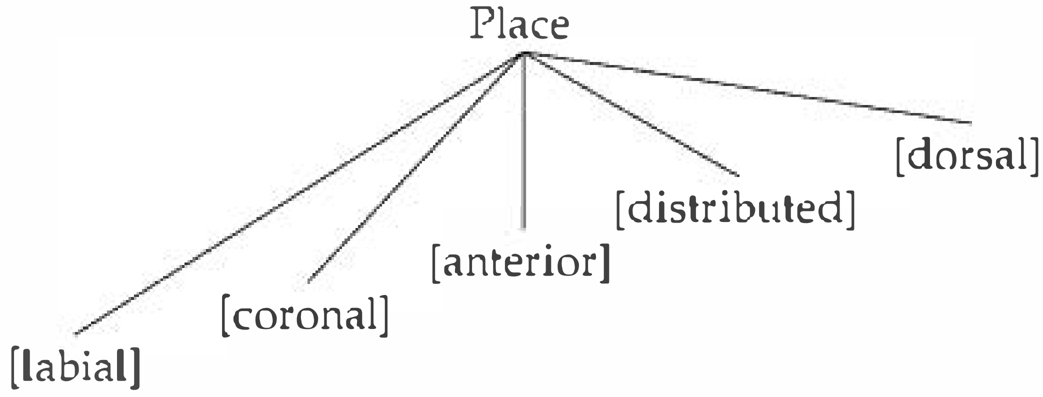
This process is structure-changing; the nasal assimilates to the exact point of articulation of the following consonant (CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION). In traditional feature theory, this process can be expressed as the rule in (2) (see also McCarthy 1988: 86):

(2) *Nasal place assimilation:* SPE-*type rule*

$$\begin{bmatrix}+\text{nas}\\+\text{cor}\\+\text{ant}\end{bmatrix} \rightarrow \begin{bmatrix}\alpha\text{cor}\\\beta\text{ant}\\\gamma\text{distr}\\\delta\text{back}\end{bmatrix} \;/\_\_\; \begin{bmatrix}\alpha\text{cor}\\\beta\text{ant}\\\gamma\text{distr}\\\delta\text{back}\end{bmatrix}$$

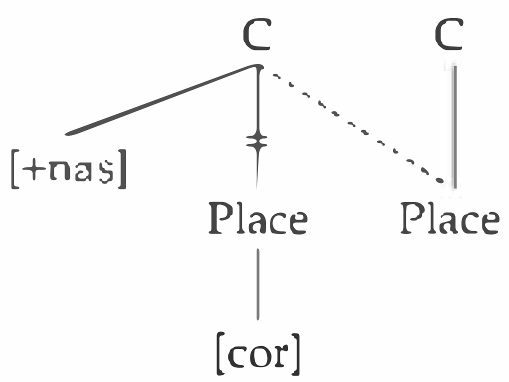
The problem with this rule is that the set of features in the structural change and the environment does not have any special status in the theory. It could easily be replaced by any random set of features, say [lateral, voice, distributed, continuant]. Yet such feature groupings are unattested cross-linguistically, while the above grouping is cross-linguistically common, and probably universal. In addition, as pointed out by McCarthy (1988), assimilation itself does not have a special status. Why do the features in the target of the rule acquire the values of an adjacent

(5) *The Place node*



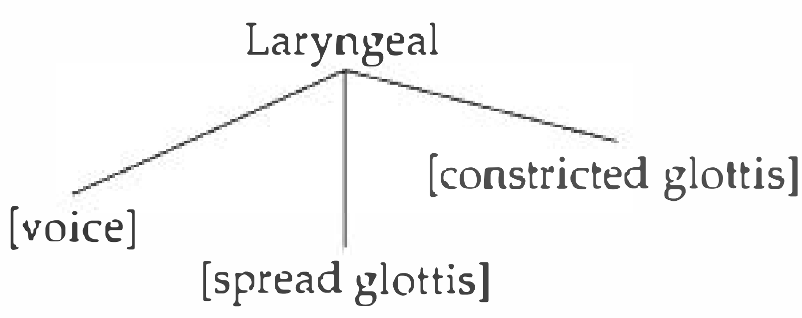
Example (6) shows how place assimilation can now be re-analyzed as a single operation, regressive spreading of the Place node. Features that are daughters of this node will spread along.

(6) *Place assimilation as Place node spreading*



A similar analysis is available for cases of laryngeal spreading or neutralization. Here, all laryngeal features are subsumed under a Laryngeal node, which can spread in assimilation or delink in cases of laryngeal neutralization.

(7) *The Laryngeal node*



Models of feature geometry generalize this approach: all features are contained within a tree-like structure. The next subsection will look at this approach more generally, and present a basic geometry which became a standard model in the second half of the 1980s. §3 will then introduce controversies regarding aspects of this structure.

2.2 *Feature geometry: Outline*

To summarize this discussion, we can formulate the following four axioms, which also characterize the research project that was embarked upon in the second half of the 1980s, the goal of finding a cross-linguistically valid model of feature geometry:

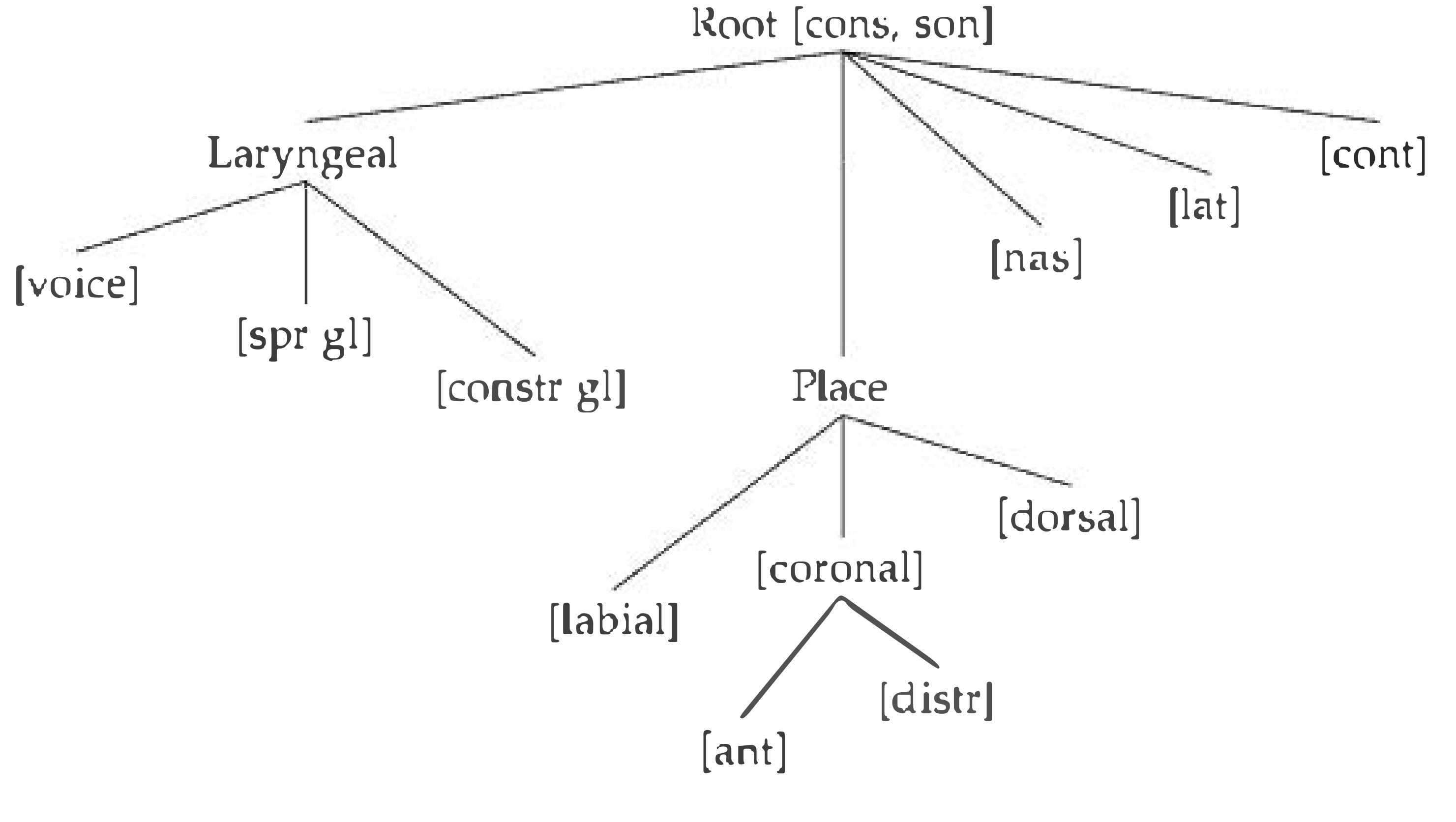
(8) a. Features are autosegments, residing on individual tiers.
b. Features can be grouped into sets or classes.
c. These sets are organized in a hierarchical structure, via intervening class nodes.
d. This structure, called a feature geometry, is universal.

The conceptualization of features as autosegments has important consequences for possible phonological operations, which are reduced to linking and delinking features and nodes. In addition, phonological processes perform individual operations only (Clements and Hume 1995: 250). A phonological rule can thus target an individual feature or a class node. Operations involving several features must consist of several rules (which can be motivated independently); alternatively, the features involved in the process form a single constituent, as dependents of a shared class node. One important motivation for positing feature-geometric structure therefore lies in constraining the notion of what a possible rule in phonology is. As was stated famously by McCarthy in an early article on feature geometry, "simply put, if the representations are right, then the rules will follow" (1988: 84).

The main goal of this research program is, however, to establish a universal geometry. The crucial task is therefore to establish exactly which features form sets. We have already seen that there are two types of evidence for grouping features under a class node. Most importantly, there is class behavior. If several features are targeted by the same process, as we have seen for place features and laryngeal features, this is interpreted as evidence that they are sisters of a superordinate node, which is targeted by this process. Secondly, there is also phonetic evidence: features which share a common articulatory property are grouped under a superordinate node. In the examples discussed so far, the evidence converges, but this is not necessarily the case, resulting in alternative proposals, which will be discussed in greater detail in §3.

At this point, it is pertinent to look at a widely assumed basic model of this geometry, to motivate it briefly, and then to proceed to open questions and controversial issues in §3.

(9) *A basic feature geometry*



The Laryngeal node has already been introduced; it is uncontroversial within the model. Different laryngeal features may be assumed (e.g. [stiff vocal folds], [slack vocal folds] replacing [voice], e.g. Halle 1995; Halle *et al.* 2000), but the general existence of a Laryngeal class node is undisputed.

The Place node has been modified slightly *vis-à-vis* (5), by grouping [anterior] and [distributed] under [coronal]. The main motivation for this move is that only coronals can be sub-specified for these features (distinguishing between alveolar/postalveolar and apical/laminal places of articulation). There is also evidence that when [coronal] spreads, it takes [anterior] and [distributed] with it; see §4 for a case of coronal harmony. Further subdivisions of the Place node are controversial, e.g. a Lingual node grouping [coronal] and [dorsal] together (see Keyser and Stevens 1994; Clements and Hume 1995; Hume 1996) or a Peripheral node as the mother of [labial] and [dorsal], corresponding to the traditional feature [grave] (Rice 1993; Rice and Avery 1993). See CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION for a detailed discussion of place features.

A quick remark should be made regarding gutturals (uvular, pharyngeal, and laryngeal segments), which form a natural class in some languages, e.g. Arabic (see also CHAPTER 25: PHARYNGEALS). There is no uniformly accepted proposal as to how to represent them in the geometry. It has been proposed to include them under the Place node, through inclusion of one or several features like [pharyngeal] and [radical] (see Clements 1991; for an extended model, see also McCarthy 1994; Rose 1994), or to expand the Laryngeal node into a Guttural node, adding features like [ATR] (Halle 1992, 1995; see also §3). We cannot discuss evidence for either approach here, but simply state that the position of gutturals in the feature tree is an unresolved issue (see also CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION).

The tree in (9) also shows that manner features are commonly not subsumed under one shared node, contrary to Clements's (1985) original proposal, which did recognize a Manner node. The reason, as discussed in detail in McCarthy (1988), is that manner features demonstrably never behave as a single constituent in phonological processes. For example, [nasal, continuant, lateral, strident] never spread as a group. It is therefore assumed that manner features associate directly with the Root node (e.g. McCarthy 1988; Rice 1993; Halle 1995; Halle *et al.* 2000, among others), or via an intermediate Oral or Supralaryngeal node (e.g. Clements 1985; Clements and Hume 1995). Processes targeting a manner feature are thus expected to target only one feature (there is no mother node that could be targeted). Two manner features stand out as special: [consonantal, sonorant] are properties of the root node itself rather than features with autosegmental properties (McCarthy 1988). The reason is that they never seem to spread or delink, unlike other features, and are therefore treated as separate.

3 Feature geometry: Controversial issues

For space reasons, this section cannot aim at a comprehensive overview of all the controversial issues regarding the shape of the feature tree. Rather, we will outline some of the debates, to give an overview of the major proposals that exist and to present the kinds of arguments used to motivate feature-geometric structure. We will focus on vowel features, perhaps the main dividing issue in the field, and add some remarks on manner features as well.

We have already seen that there are two main arguments for grouping features under a class node. First, phonological behavior: features that spread or delink together are grouped under a common node. Second, grounding in articulatory phonetics: features are grouped together that share an articulatory property, e.g. place of articulation or laryngeal setting. Strict application of the first principle can present conflicting evidence, though: groupings may be non-universal (found in some languages, but not others) or variable. Class behavior alone thus does not yield a universal geometry. We will see evidence for this in the discussion of manner features in §3.2. On the other hand, relying on the articulatory properties of features alone may not capture all groups of features which pattern consistently as a class. We turn to an example of this now in the discussion of vowel features.

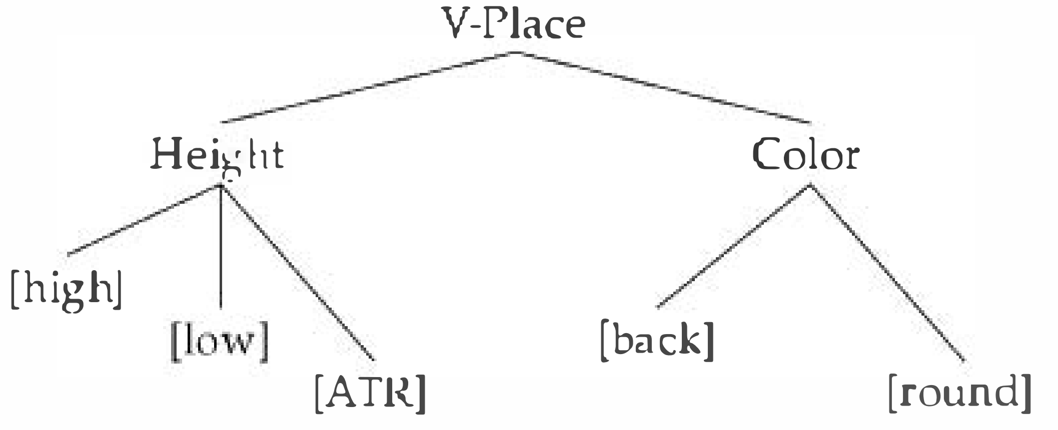
3.1 *Vowel features*

A major schism in feature geometry concerns the representation of vowels. This schism is an effect of the two opposing principles for establishing feature classes, articulation and phonological behavior. Articulator Theory (e.g. Sagey 1986; Halle 1992, 1995; Halle *et al.* 2000; Avery and Idsardi 2001) grounds the feature tree in articulation, while a different school maintains the primacy of phonological patterning (e.g. Hume 1990, 1992, 1996; Clements 1991; Odden 1991, 1994; Clements and Hume 1995), from which a different model, known as Unified Feature Theory, emerged. This section will outline the main differences between the two models and the reasoning behind them.

3.1.1 *Articulator Theory*

In Articulator Theory (Sagey 1986; Halle 1992, 1995), priority is given to articulatory considerations in the grouping of features in the geometry. "Articulator-bound" features (features that can be articulated by one specific articulator only, such as [voice], which is bound to the larynx, or [round], which can only be executed by the lips) are daughters of articulator nodes. This also provides a reason for why manner features are not subsumed under a class node but associate directly with the root node instead: they are articulator-free, that is, they can be executed by different articulators (features like [continuant] or [strident], for example). Only features that are articulator-bound, then, can be part of a set dominated by a class node, which corresponds to an articulator. The tree in (10) provides the basic geometry assumed in Articulator Theory. Note that the features [coronal, labial, dorsal] have been promoted to class node status. Features that are executed by one of these articulators are daughters of these nodes.

(12) *Vowel place features* (Odden 1991; adapted)



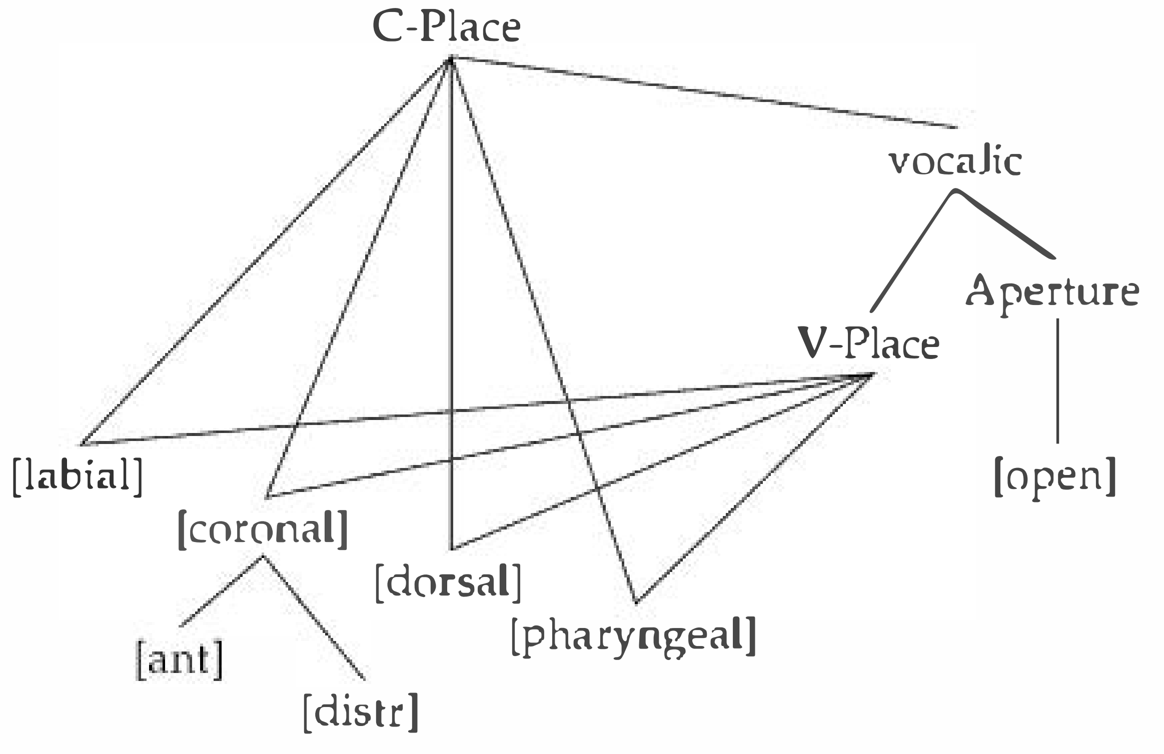
Odden argues that Articulator Theory makes the wrong predictions with regard to which vowel features can spread together. It predicts that [high, low, back] could spread together, to the exclusion of [round], by spreading the Dorsal node, although such a process is unattested. In addition, it cannot spread [back] and [round] in one single process, because they are daughters of different class nodes. He concludes that phonological patterning should guide decisions about feature grouping as much as phonetic (articulatory) considerations. He also offers a phonetic explanation for his proposal, but this explanation is grounded in acoustics and perception, rather than articulation. Both lip rounding and tongue backing have an effect on the second formant of a vowel, while tongue height has an effect on the first formant. The basic distinction between a Height and a Color node can thus be motivated perceptually.

Parallel to Odden's work, a more radical departure from traditional feature theory was developed by Clements and Hume (Hume 1990; Clements 1991; Hume 1992; Clements and Hume 1995; Hume 1996). They, too, assume a V-Place node, which is distinct from the C-Place node, but in addition they propose that vowel features be replaced by consonantal place features. They redefine the traditional features [back, round, low] in the following terms:

(13) [+round] = [labial]
[+back] = [dorsal]
[–back] = [coronal]
[+low] = [pharyngeal]

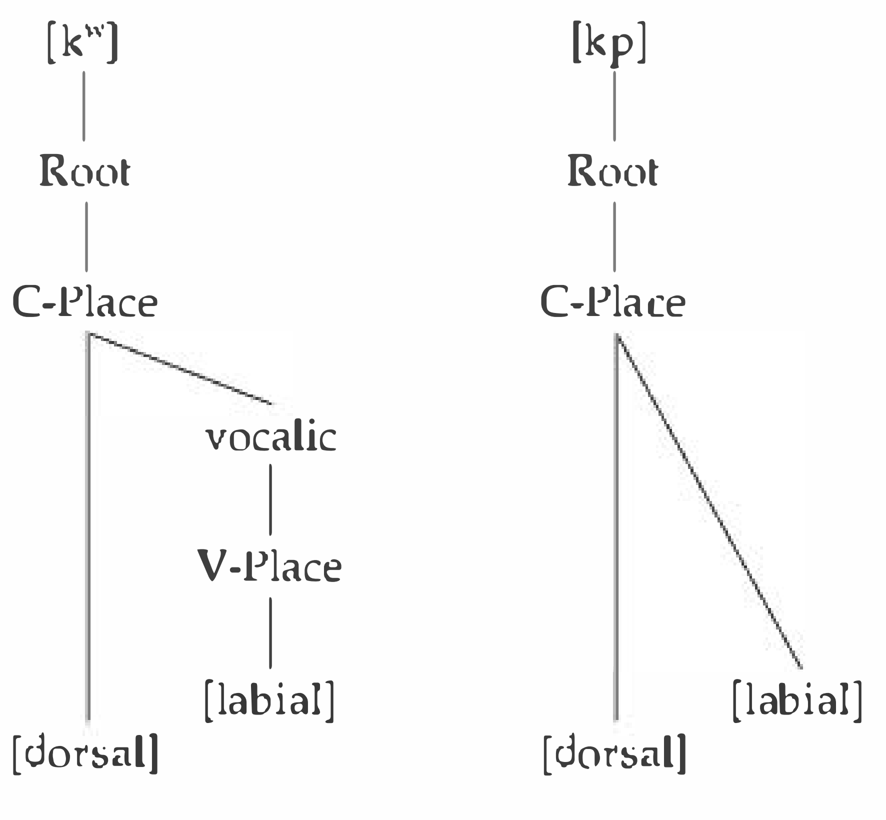
This redefinition is based on observed consonant–vowel interactions where, for example, round vowels and labial consonants, or low vowels and gutturals, form a natural class. Hume (1996) discusses evidence from Maltese where the imperfective prefix vowel changes, depending on the stem-initial consonant. Normally, the prefix vowel is determined by the stem vocalism; however, it is [i] before coronal obstruents (as in [ji-dlam] 'to grow dark', cf. perfective [dalam]), and /i/ is lowered to [a] before gutturals (as in [ja-ħdem] 'to work', cf. UR /ħidim/). With the new vowel place features, these can now be straightforwardly explained as the spreading of [coronal] and [pharyngeal], respectively. The geometry which Clements and Hume propose for vowel features is given in (14): consonants and vowels share the same place features; how the feature is realized phonetically depends on whether it is a daughter of the C-Place node or of the V-Place node. The Aperture node is a simplified version of Odden's (1991) Height node, containing recursive [open] features. For details on this node, see Clements (1991), Clements and Hume (1995), and Salting (1998). The resulting geometry is known as Unified Feature Theory (Clements and Hume 1995).

(14) *Feature geometry with unified features*



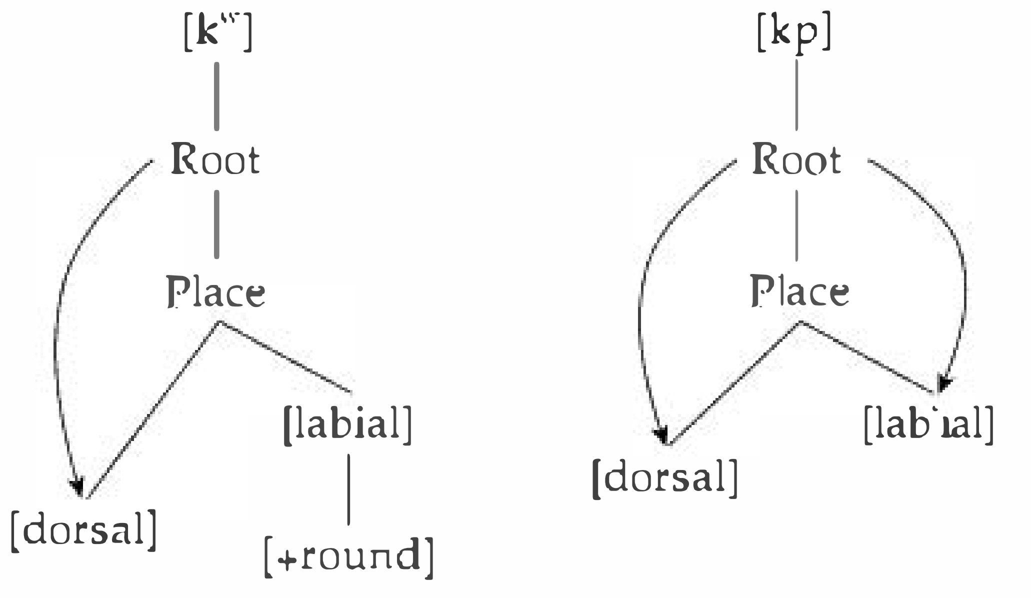
An additional advantage of the Unified Feature Theory model is that it can straightforwardly capture secondary articulations, such as labialization, palatalization, and pharyngealization, and distinguish them from double articulations, such as co-articulated stops. In secondary articulation, the major articulator is a feature dependent on the C-Place node, while the secondary articulation (such as palatalization or labialization) is expressed by a feature dependent on the V-Place node. The two representations in (15) illustrate this difference by comparing a co-articulated stop [kp] (no subordination of articulators) with labialized [kʷ] (secondary labialization). This representation also captures the observation that secondary articulations can function as blockers in vowel harmony; since they are vocalic articulations, they are expected to interact with processes targeting vocalic segments. See also CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS; CHAPTER 71: PALATALIZATION; and CHAPTER 29: SECONDARY AND DOUBLE ARTICULATION for more detailed discussions of secondary articulations.

(15) *Secondary* vs. *double articulations* (Unified Feature Theory)



Articulator Theory does not offer a comparably straightforward way of distinguishing between primary and secondary articulations, but has to resort to a new type of phonological primitive, a pointer device, to indicate which of the two articulators is the main (or designated) articulator, as in (16).[1]

(16) *Secondary* vs. *double articulations* (Articulator Theory)



To summarize, there are thus two different proposals with respect to how vowel features are represented in the feature tree, based on two different organizing principles of the tree: articulatory grounding *vs.* constituency in phonological processes (providing evidence for a separate V-Place node and a unified set of features). This conflict has still not been resolved – on the contrary, it is exacerbated when manner features are taken into consideration as well.

3.2 *More on manner features*

To recapitulate from §2, manner features are not grouped under a shared class node, at least since McCarthy (1988), motivated in Articulator Theory by the observation that they are not articulator-bound.[2] This section will briefly look at some proposals that group individual manner features with other features, based on the same kind of argument that motivated the reconceptualization of vowel features in the previous section. We will see, however, that in the case of manner features there is conflicting evidence which has dealt a serious blow to the endeavor of finding a universal feature geometry.

Recall first one major distinction. While most manner features are analyzed as autosegments that associate directly with the root, without an intervening class node, the major class features [consonantal] and [sonorant] are granted special status; rather than being autosegments, they are viewed as part of the root node itself. The reason for this is that these two features never seem to spread or delink, as other features would. Their function instead is entirely classificatory, to distinguish vowels from consonants, and obstruents from sonorants. This, however,

[1] A solution to this problem (of adding a new type of representational primitive, the pointer) that stays true to the principles of Articulator Theory is suggested in Halle *et al.* (2000), where the set of features itself is expanded to distinguish between primary and secondary articulators (each articulator coming with two features).

[2] See Clements (1985) for the original proposal for a Manner node, and Morén (2003, 2007) for the reintroduction of a Manner node (on which more in §6).

which are consistent with an analysis that views [nasal] as variably associated with either the SV node or a Soft Palate node, with individual languages taking one of the two options. See CHAPTER 78: NASAL HARMONY for a review of these nasal harmony types and a summary of Piggott's arguments. Piggott's idea that feature dependencies may be variable, aiming at a "softer" approach to universals in feature organization, has not been pursued systematically, however. We will return to this point in the concluding section of the chapter.

3.3 *Interim summary*

By the mid-1990s it had become increasingly clear that the goal of finding a universal (and universally agreed upon) hierarchy of features could not be attained. The two major organizing principles of feature geometry, articulatory similarity and phonological patterning, do not always match up perfectly. While the basic structure of the geometry is fairly uncontroversial, there are diverging proposals regarding feature organization outside of the core sets of place and laryngeal features. Vowel features are a case in point: articulatorily diverse groups of vowel features nevertheless form phonological constituents (such as [back] and [round]). This does not mean that phonological patterning should always trump articulatory considerations, however: evidence from patterning can yield conflicting evidence as well, as we have seen in the case of manner features, which associate variably with other features in phonological processes. While Piggott (1992) explicitly allows such variability to account for the variable patterning of nasals, the idea of parameterizing the principles of feature geometry was never pursued systematically.

It may therefore not be surprising that the advent of a major new theory in the 1990s, Optimality Theory, stimulated a new research program to see how the feature-geometric generalizations could be captured differently and potentially more accurately within a theory of violable constraints, rather than a theory of (universal) representations. We will turn to these proposals in §5 and evaluate them critically, also in the light of advances in representational theory, in §6. First, however, we will turn to another aspect of feature geometry, which will be important for the overall evaluation of this research program, namely the way in which this theory has contributed to the understanding of phonological operations, besides proposing a new theory of phonological representations.

4 Non-local feature interaction

We have seen how phonological processes are formalized autosegmentally in feature geometry, as the spreading or delinking of a node, either a feature or a class node. §2 showed how place assimilation can be understood as the spreading of the Place node, taking all place features with it. More radical changes can also be formalized as a single operation; for example, debuccalization, i.e. the loss of oral features (as in /t/ → [ʔ]), can be formalized as the delinking of the Oral node, leaving only the Laryngeal node to specify a segment.

Feature geometry also allowed a breakthrough in the understanding of non-local interactions, that is, processes in which trigger and target are not directly adjacent. Traditional rule-based theory had a problem with capturing such non-local interactions: under what conditions can trigger and target be separated? How

Similar generalizations can be made for dissimilatory processes, which can also occur over a distance. The Obligatory Contour Principle (OCP), which prohibits identical adjacent feature specifications, is the trigger for dissimilation, but it does not necessarily hold over linearly adjacent segments but can also refer to two autosegments adjacent on their respective tier. See Odden (1994) for a typology of long-distance dissimilation. In essence, the same feature that can assimilate long-distance can also dissimilate. Laryngeal dissimilation is fairly common and can occur over a distance. For example, Dahl's Law in Eastern Bantu languages changes a voiceless prefix-initial stop to voiced if the stem-initial consonant is voiceless as well. Lyman's Law in Japanese prohibits more than one [+voice] obstruent in a word. Grassmann's Law in Indo-European dissimilates aspirated segments. Also common is liquid dissimilation, as in the following Latin examples:

(21) *Latin liquid dissimilation*

nav-alis	*fat-alis*	*crimin-alis*	*radic-alis*
sol-aris	*milit-aris*	*peculi-aris*	*line-aris*
flor-alis	*later-alis*	*plur-alis*	*sepulchr-alis*

These examples show that the denominal adjectival suffix *-alis* becomes *-aris* if preceded by an [l] in the stem which is not directly adjacent. Intervening [r] (as in *pluralis*) blocks the process. This can be understood as an OCP effect on the lateral tier: in **sol-alis* there would be two tier-adjacent instances of [+lateral]. Again, note the crucial role of underspecification: since [±lateral] is only contrastive for liquids, only they carry this feature. Hence, [r], being the only segment which is specified as [−lateral], is the only potential blocker of the dissimilatory process.

In sum, feature geometry does not just constrain possible operations in terms of which features spread together; it also provides an elegant analysis of non-local interactions for assimilation and dissimilation. At their relevant tier, the interactions are in fact local, involving adjacent autosegments. In addition, the proposal is falsifiable: any non-local interaction that cannot be made local on some tier constitutes counterevidence to the model. The question of whether there exist such counterexamples is still unresolved and a possible direction for future research. Conversely, any theory of phonology that wishes to abandon feature geometry must have an account of non-local interactions. It is these alternative proposals, which arose as a consequence of the emergence of Optimality Theory as a new model of grammar, to which we now turn.

5 Feature organization in Optimality Theory

The advent of Optimality Theory (OT; Prince and Smolensky 1993) allowed a fresh look at the generalizations that feature geometry is trying to capture by opening up the possibility that these generalizations could also emerge from a theory of constraint interaction (CHAPTER 63: MARKEDNESS AND FAITHFULNESS CONSTRAINTS). Consequently, there was a shift in interest away from feature geometry in the second half of the 1990s. This section will review and motivate that shift, and by default). Example (22) gives some examples, with the [−high] plural suffix [-ler, -lar] and the [+high] genitive suffix [-in, -yn, -ɨn, -un].

(22) *Turkish vowel harmony*

nom sg	*gen sg*	*nom pl*	
ip	ipin	ipler	'rope'
gøz	gøzyn	gøzler	'eye'
pul	pulun	pullar	'stamp'
at	atɨn	atlar	'horse'

Traditionally, Turkish harmony is analyzed as two separate processes. In feature-geometric terms, spreading is either of a single feature [back], or of the Color node, comprising [back] and [round] (adopting the proposal in Odden 1991, discussed above). Padgett suggests that there is only one process, which spreads Color features. This process always targets both features, although only [back] spreads if spreading of [round] would create a marked segment, i.e. [o] or [ø]. He achieves this by ranking a markedness constraint against non-high round vowels above a constraint SPREAD(Color), where the class Color comprises [back, round]. No spreading yields two constraint violations (one for each feature); spreading of [back] only yields one violation. Violation of SPREAD(Color) is therefore minimal and only licensed to avoid violation of the markedness constraint against non-high round vowels.

A feature-geometric analysis, in comparison, would predict what Padgett calls "sour grapes spreading": either all features spread (because the Color node spreads), or none do. The gradient behavior of spreading (spreading as much as possible while satisfying higher-ranked markedness constraints) can therefore not be addressed straightforwardly in a feature-geometric model (although a feature-geometric implementation is a possible option in OT, by spreading the Color node if possible, or only [back] if spreading of the Color node were to create a violation of the markedness constraint against non-high round vowels). There is an additional property to Padgett's analysis: the absence of feature-geometric structure means that non-local interactions (such as vowel harmony) can no longer be analyzed as local at the level of some tier. In vowel harmony, spreading therefore cannot skip intervening consonants, which have to participate in the harmony. All spreading is strictly local, although the effect of vowel harmony on a consonant may be imperceptible (see also Ní Chiosáin and Padgett 1997, 2001). Locality issues will be discussed separately in the next subsection.

Yip (2004, 2005) goes a step further than Padgett in showing that class behavior can be modeled solely as constraint interaction, without the need for feature classes as indices on features. She looks at the feature [lateral], whose status in the geometry has always been ambiguous, with conflicting proposals to group it either under the Coronal node or the SV node (cf. §3.2). She links this ambiguity to typological observations regarding the markedness of different types of laterals, finding that coronal laterals are less marked than dorsal laterals and that sonorant laterals are less marked than obstruent laterals, from which she derives two presumably universal markedness hierarchies. She then goes on to show that the observations that led to the proposals about the position of [lateral] in the feature tree (spreading of [lateral] with [coronal] or only to [coronal] targets;

spreading of [lateral] with SV or only to targets having an SV node) also fall out from interleaving these hierarchies with faithfulness constraints. Essentially, in Yip's approach a pro-spreading constraint like Spread or Share wants to spread as many features as possible, unless this creates an illicit segment (which is militated against by a highly ranked markedness constraint). The observation of variable class membership is thus re-analyzed as an epiphenomenon of the interaction of markedness and faithfulness constraints generally, and an effect of markedness scales on feature co-occurrence more specifically. (See chapter 31: lateral consonants for a more detailed sketch of this analysis.) Unfortunately, this approach has to my knowledge never been applied more generally, and its consequences have not been explored in depth. Future research should shed more light on whether Yip's approach is indeed a viable alternative to representational devices that encode class membership.

5.2 *Strict locality in spreading*

In sum, then, generalizations about class behavior can be made in OT without having to resort to tree structures that encode such behavior representationally, although some of the details may still warrant further research. The third point of the list of advances in feature geometry, non-local interactions, is a more controversial issue. We saw that Padgett (1995, 2002) has to analyze all spreading as strictly local, in the absence of complex autosegmental representations, a view that is applied and developed further, for example, in Ní Chiosáin and Padgett (1997, 2001), Walker (1998), Gafos (1999), and Baković (2000). In the case of vowel harmony, this implies that consonants also participate in vowel harmony. They argue that this is supported phonetically by co-articulation effects on the consonants (such as lip rounding in round harmony). While this may be a welcome simplification of the system (obviating the need for representational devices that introduce relativized notions of locality), it also introduces some complications. Ní Chiosáin and Padgett (1997, 2001) note that vowel harmony is cross-linguistically frequent, while consonant harmony is non-existent. It is thus necessary to have a mechanism that allows the propagation of vowel features through consonants while blocking the propagation of consonant features through vowels. They therefore co-index place features for degree of aperture. A [labial$_V$] articulation (rounding on vowels) can be co-articulated on a consonant, but a consonantal place feature like [labial$_C$] could not be co-articulated on a vowel, because it is stronger and would deprive the vowel of its vocalicness, a principle they ground aerodynamically in what they call the "bottleneck effect." See chapter 75: consonant–vowel place feature interactions for an elaboration of this idea of "inherent V-Place."

One problem with this approach is that distinctive features are commonly used to encode possible contrasts, but there are no languages that would contrast consonants with or without a vocalic co-articulation (which is different from secondary articulations like labialization, palatalization, etc.). Ní Chiosáin and Padgett therefore need an additional mechanism to rule out such unattested contrasts. Here they turn to Dispersion Theory (Flemming 1995), arguing that such potential contrasts are not well dispersed enough to survive into actual systems. This final point is probably most controversial, since Dispersion Theory has been criticized independently (see e.g. McCarthy 2002). Nevertheless, the idea of strictly local specified, providing a principled algorithm for underspecification (Dresher *et al.* 1994; Dresher 2009). See also CHAPTER 2: CONTRAST for an introduction. The application of this algorithm generates a hierarchical tree structure itself. If learners posit new features on the basis of contrast they observe within a set of segments, it is quite possible that the contrastive hierarchy will to some extent be reflected in an emergent geometry. For example, in a language in which all liquids are [coronal], the feature [lateral] may be used to contrastively subdivide the set of coronals. Consequently, learners may posit that [lateral] is a dependent of [coronal] in the geometry as well. In a language in which liquids are phonologically placeless, however, learners may posit that [lateral] subdivides the set of sonorants, as a dependent of the SV node. The ambiguity of [lateral], discussed in CHAPTER 31: LATERAL CONSONANTS, may thus reflect the system of contrasts in a language more generally. The concept of emergent features thus shows the potential for a new understanding of feature organization. Future research could shed more light on which aspects of feature organization are provided by UG, which are emergent via the phonetic content of features and which are emergent via the contrast encoded by a feature.

Finally, one additional direction for potential future developments should be mentioned, a recent challenge to the standard opinion that the anchor of features is the segment. In a detailed typological survey, Kehrein (2002) and Kehrein and Golston (2004) conclude that the anchor of laryngeal features is not the segment but rather the syllable constituent (onset, nucleus, coda), since they find no evidence for contrastive laryngeal specifications within syllable constituents in any language. The Laryngeal node is hence not formalized as a dependent of the segmental root node but as a sister of segmental root nodes, as a dependent of a syllable constituent like the onset. Non-contrastive differences across languages are relegated to phonetic implementation and the phasing of laryngeal gestures. If Golston and Kehrein are correct, this opens up the possibility for more research in the area. Are there other features that are not properties of segments but of higher prosodic categories, either universally or in individual languages? The once strict division between tree structure above the segment (prosody) and below the segment (feature geometry) would then become permeable, which in turn would have consequences for our understanding of locality in feature interaction. Seemingly long-distance interactions may be local not by virtue of tier locality but by virtue of the feature itself spreading higher up in the prosodic hierarchy. For now, however, this idea is mostly speculative.

In sum, the research program that was embarked upon in the mid-1980s, to investigate the possibility of segment-internal structure which organizes distinctive features into sets or classes, is far from finished. After peaking in interest in the early 1990s and a slow but steady demise when mainstream phonological interest shifted to the role of constraints in phonology instead, the program has seen some renewed interest more recently. With some original problems of orthodox feature geometry still unresolved and the promise of Optimality Theory to provide a more elegant and empirically superior alternative not having been fully realized, the issue of feature organization is still as alive as it was 20 years ago. New directions in research, as outlined in the final section of this chapter, may shed new light on the issue and bring us closer to a better understanding of feature organization.

REFERENCES

Avery, Peter & William J. Idsardi. 2001. Laryngeal dimensions, completion and enhancement. In T. A. Hall (ed.) *Distinctive feature theory*, 41–70. Berlin & New York: Mouton de Gruyter.
Avery, Peter & Keren Rice. 1989. Segment structure and coronal underspecification. *Phonology* 6. 179–200.
Baković, Eric. 2000. Harmony, dominance, and control. Ph.D. dissertation, Rutgers University.
Beeler, Madison S. 1970. Sibilant harmony in Chumash. *International Journal of American Linguistics* 36. 14–17.
Blaho, Sylvia, Patrik Bye & Martin Krämer (eds.) 2007. *Freedom of analysis?* Berlin & New York: Mouton de Gruyter.
Blevins, Juliette. 1994. A place for lateral in the feature geometry. *Journal of Linguistics* 30. 301–348.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.
Clements, G. N. 1991. Place of articulation in consonants and vowels: A unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5. 77–123.
Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 245–306. Cambridge, MA & Oxford: Blackwell.
Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.
Dresher, B. Elan, Glyne L. Piggott & Keren Rice. 1994. Contrast in phonology: Overview. *Toronto Working Papers in Linguistics* 13. iii–xvii.
Flemming, Edward. 1995. Auditory representations in phonology. Ph.D. dissertation, University of California, Los Angeles.
Gafos, Adamantios I. 1999. *The articulatory basis of locality in phonology*. New York: Garland.
Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.
Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Oxford & Cambridge, MA: Blackwell.
Halle, Morris. 1992. Features. In William Bright (ed.) *Oxford international encyclopedia of linguistics*, vol. 3, 207–212. New York: Oxford University Press.
Halle, Morris. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26. 1–46.
Halle, Morris, Bert Vaux & Andrew Wolfe. 2000. On feature spreading and the representation of place of articulation. *Linguistic Inquiry* 31. 387–444.
Hansson, Gunnar Ólafur. 2001. Theoretical and typological issues in consonant harmony. Ph.D. dissertation, University of California, Berkeley.
Harris, John & Geoff Lindsey. 1995. The elements of phonological representation. In Jacques Durand & Francis Katamba (eds.) *Frontiers of phonology: Atoms, structures, derivations*, 34–79. London & New York: Longman.
Hume, Elizabeth. 1990. Front vowels, palatal consonants, and the rule of umlaut in Korean. *Papers from the Annual Meeting of the North East Linguistic Society* 20. 230–243.
Hume, Elizabeth. 1992. Front vowels, coronal consonants and their interaction in non-linear phonology. Ph.D. dissertation, Cornell University.
Hume, Elizabeth. 1996. Coronal consonant, front vowel parallels in Maltese. *Natural Language and Linguistic Theory* 14. 163–203.
Hume, Elizabeth & David Odden. 1996. Reconsidering [consonantal]. *Phonology* 13. 345–376.
Kaisse, Ellen M. 1992. Can [consonantal] spread? *Language* 68. 313–332.
Kehrein, Wolfgang. 2002. *Phonological representation and phonetic phasing: Affricates and laryngeals*. Tübingen: Niemeyer.

Kehrein, Wolfgang & Chris Golston. 2004. A prosodic theory of laryngeal contrasts. *Phonology* 21. 325–357.

Keyser, Samuel J. & Kenneth N. Stevens. 1994. Feature geometry and the vocal tract. *Phonology* 11. 207–236.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.) *Phonological structure and phonetic form: Papers in laboratory phonology III*, 191–233. Cambridge: Cambridge University Press.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John J. 2007a. Slouching towards optimality: Coda reduction in OT-CC. In Phonological Society of Japan (ed.) *Phonological studies 10*, 89–104. Tokyo: Kaitakusha.

McCarthy, John J. 2007b. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.

McCarthy, John J. 2008a. The gradual path to cluster simplification. *Phonology* 25. 271–319.

McCarthy, John J. 2008b. Harmony in harmonic serialism. Unpublished ms., University of Massachusetts, Amherst.

Mielke, Jeff. 2005. Ambivalence and ambiguity in laterals and nasals. *Phonology* 22. 169–203.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Morén, Bruce. 2003. The parallel structures model of feature geometry. *Working Papers of the Cornell Phonetics Laboratory* 15. 194–270.

Morén, Bruce. 2007. The division of labor between segment-internal structure and violable constraints. In Blaho *et al.* (2007), 313–344.

Ní Chiosáin, Máire & Jaye Padgett. 1997. Markedness, segment realization, and locality in spreading. Technical Report LRC-97-01, Linguistics Research Center, University of California, Santa Cruz.

Ní Chiosáin, Máire & Jaye Padgett. 2001. Markedness, segment realization, and locality in spreading. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 118–156. Cambridge: Cambridge University Press.

Odden, David. 1991. Vowel geometry. *Phonology* 8. 261–289.

Odden, David. 1994. Adjacency parameters in phonology. *Language* 70. 289–330.

Oostendorp, Marc van & Jeroen van de Weijer (eds.) 2005. The internal organization of phonological segments. Berlin & New York: Mouton de Gruyter.

Padgett, Jaye. 1991. Stricture in feature geometry. Ph.D. dissertation, University of Massachusetts, Amherst.

Padgett, Jaye. 1994. Stricture and nasal place assimilation. *Natural Language and Linguistic Theory* 12. 465–513.

Padgett, Jaye. 1995. Feature classes. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 385–420. Amherst: CLSA.

Padgett, Jaye. 2002. Feature classes in phonology. *Language* 78. 81–110.

Piggott, Glyne L. 1992. Variability in feature dependency: The case of nasality. *Natural Language and Linguistic Theory* 10. 33–77.

Poser, William J. 1993. Are strict cycle effects derivable? In Sharon Hargus & Ellen M. Kaisse (eds.) *Studies in Lexical Phonology*, 315–321. San Diego: Academic Press.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Rice, Keren. 1993. A re-examination of the feature [sonorant]: The status of "sonorant obstruents." *Language* 69. 308–344.

Rice, Keren. 1994. Peripheral in consonants. *Canadian Journal of Linguistics* 39. 191–216.

Rice, Keren & Peter Avery. 1989. On the interaction between sonorancy and voicing. *Toronto Working Papers in Linguistics* 10. 65–82.

Rice, Keren & Peter Avery. 1993. Segmental complexity and the structure of inventories. *Toronto Working Papers in Linguistics* 12. 131–153.
Rose, Sharon. 1994. Guttural contrasts. *Toronto Working Papers in Linguistics* 13. 147–174.
Rose, Sharon & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80. 475–531.
Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.
Salting, Don. 1998. The nested subregister model of vowel height. Ph.D. dissertation, Indiana University.
Steriade, Donca. 1993. Segments, contours and clusters. In André Crochetière, Jean-Claude Boulanger & Conrad Ouellon (eds.) *Proceedings of the 15th International Congress of Linguistics*, 71–82. Quebec: Université Laval.
Steriade, Donca. 1994. Complex onsets as single segments: The Mazateco pattern. In Jennifer Cole & Charles W. Kisseberth (eds.) *Perspectives in phonology*, 203–291. Stanford: CSLI.
Uffmann, Christian. 2005. Optimal geometries. In van Oostendorp & van de Weijer (2005), 27–62.
Uffmann, Christian. 2007. Restricting GEN. In Blaho *et al.* (2007), 281–312.
Walker, Rachel. 1998. Nasalization, neutral segments, and opacity effects. Ph.D. dissertation, University of California, Santa Cruz.
Walker, Rachel. 2000. Long-distance consonantal identity effects. *Proceedings of the West Coast Conference on Formal Linguistics* 19. 532–545.
Yip, Moira. 2004. Lateral survival: An OT account. *International Journal of English Studies* 4(2). 25–51.
Yip, Moira. 2005. Variability in feature affiliations through violable constraints: The case of [lateral]. In van Oostendorp & van de Weijer (2005), 63–91.

28 The Representation of Fricatives

Bert Vaux
Brett Miller

1 Introduction

The phonetic properties of fricatives have recently received a great deal of attention (Shadle 1985; Ladefoged and Maddieson 1996: ch. 5; Johnson 1997: ch. 6; Stevens 1998: ch. 8; Maniwa *et al.* 2009; Ramsay 2009; among others). The *phonological* properties of this class of sounds, on the other hand, have, with a few notable exceptions, remained largely undisputed since the publication of Chomsky and Halle (1968), which itself essentially carries on the featural analysis of fricatives in Trubetzkoy (1939) and Jakobson *et al.* (1952). In this analysis, the class of fricatives is characterized by the distinctive features [−sonorant, +continuant].

Once one scratches the surface of the subject, though, a number of challenging questions appear:

(1) a. Do fricatives actually behave as a distinct phonological class?
b. Are all fricatives [−sonorant]?
c. Are all fricatives [+consonantal]?
d. Are all or any fricatives [+spread glottis] (or its equivalent)?

This chapter addresses these and other challenges by synthesizing what we have learned from traditional and contemporary descriptive and theoretical studies involving fricatives, with an eye towards determining what properties (if any) consistently characterize this phonological class and why. We shall see that the exact membership and feature characterization of the fricative class depends on how one defines the features involved, but that there are good phonological reasons for assuming the existence of a coherent fricative class defined by the features [−sonorant, +continuant] and including not only the relatively uncontroversial suspects {f v ɸ β θ ð x ɣ χ ʁ} but also what we can call the strident or sibilant fricatives {s z ʃ ʒ ʂ ʐ ɕ ʑ}. The laryngeals {h ʜ ɦ ħ ɧ} behave less uniformly: some of them pattern with fricatives in some languages and with sonorants in others. There is also variation within the fricative class for the features [spread glottis], [strident], [ATR], and possibly [consonantal], so these should not be included in the definition of the fricative class. Since the class appears adequately defined with the independently needed features [+continuant, −sonorant], we conclude on grounds of parsimony that this definition is preferable to one requiring the introduction of additional features such as [fricative].

We do not investigate stridency in this chapter, both because it does not characterize the class of fricatives as a whole and because it has been well dealt with by Clements (2006). Nor do we deal with tongue root advancement, which Vaux (1992, 1996) has suggested is a common attribute of voiced fricatives only by virtue of their membership in the class of voiced obstruents.

1.1 *Are fricatives a bona fide phonological natural class?*

Fricatives present a challenge to the ontology of distinctive feature theory, in that they possess a clear and unique phonetic identity resulting from turbulent airflow through a narrow constriction that is reflected in neither of the cornerstones of distinctive feature theory (CHAPTER 17: DISTINCTIVE FEATURES), (i) properties that are phonologically active, and (ii) properties that are necessary to distinguish a phoneme (or in this case a class of phonemes) from all others. The problem is this: if we were to postulate a phonological feature such as [fricative] to capture the distinctive phonetics of the fricative class, we would run afoul of criterion (i) by virtue of the fact that there is no evidence for such a feature being phonologically active; the attested range of phonological behaviors of fricatives can be captured by other independently required members of their feature complement, as we will see throughout the rest of this chapter. These same features suffice to distinguish the class of fricatives from stops, vowels, and so on; the feature [fricative] therefore is also not required by criterion (ii).

Before examining the phonological representation of fricatives, we should therefore consider first whether there is sufficient evidence for considering them to form a natural class. Is there anything that is truly distinctive phonologically about the set of fricatives, or are they simply what is left of the class of obstruents after one subtracts the stops? Do fricatives ever behave as a class to the exclusion of all other phonemes? The answer to this question is a qualified yes: many phonological phenomena specifically target or are triggered by the fricative set of the language in question, but (unsurprisingly), for a variety of historical and accidental reasons, none of the phenomena in question suffice to delineate the entire cross-linguistic class of fricatives.

Tiberian Hebrew spirantization, for instance, produces alternations between plain oral stops [p b t d k g] and fricatives [f v θ ð x ɣ] respectively, as in (1) (Idsardi 1998: 39), but does not produce alternations for the emphatic stop [tˤ] or for the guttural fricatives [ħ h]. The Tiberian Hebrew surface consonant inventory is given in (2), based on Rendsberg (1997) and Green (2004).

(2) *Tiberian Hebrew surface consonant inventory*

	labial	dental	alveolar	palatal	velar	pharyngeal	glottal
stops	p b	t d			k g		ʔ
fricatives	f v	θ ð	s	ʃ	x ɣ	ħ ʕ	h
emphatics		tˤ	sˤ		q		
nasals	m	n					
liquids		l r					
glides	w			j			

The interpretation of the emphatics is conjectural, and the positioning of the pharyngeals and glottals should be understood as arbitrary with respect to their [sonorant] values, which we will not investigate here (though we discuss the sonorance of laryngeals in various other languages in §2.2; see also CHAPTER 25: PHARYNGEALS).

(3) *Examples of Hebrew spirantization alternations*

a.	[t ~ θ]	√ktb	kaːˈθav	'he wrote'
	[k ~ x]		jixˈtoːv	'he was writing'
b.	[d ~ ð]	√gdl	gaːˈðluː	'they were great'
	[g ~ ɣ]		jiɣˈdaːluː	'he was being great'

The facts in (3) are not a problem for the notion that fricatives form a phonological class, because the phenomenon can be plausibly analyzed as involving spreading of [+continuant] from vowels to a following singleton segment specified with [+consonantal] and perhaps [−constricted glottis]. This is shown in the simplified spirantization rule in (4), which abstracts away from certain morphological and lexical exceptions (on which see Coetzee 1999; Green 2004).

(4) *Spirantization rule*[1]

[−cons, −constr gl][+cons, (−constr gl)]$_i$ X$_j$
|
[+cont]

If this analysis is correct (see further §2.1.2), we do not expect the guttural fricatives to have [−continuant] counterparts in Tiberian Hebrew; the rule only serves to produce fricatives from underlying stops, and has no effect on underlying fricatives such as the gutturals. Tiberian Hebrew spirantization therefore provides evidence neither for nor against the gutturals being fricatives rather than glides (i.e. obstruents rather than sonorants).

Nor can anything clear be inferred about the lack of spirantized allophones for the emphatic stops, though it seems likely that the emphatics had some trait such that their spirantized outcomes would have been fairly unusual, complex, or difficult sounds. No emphatics, and no other comparable class of non-participants, are to be found when one considers the spirantization processes reconstructed in Old Irish (Celtic: Thurneysen 1946) and observed synchronically in Shoshone and Southern Paiute (Numic: Charney 1993). These cases of spirantization could also be handled by (4) or by a very similar rule.

With these examples in mind, let us briefly consider additional phonological phenomena that can be reasonably said to target or be triggered by the class of fricatives. Perhaps the best-known case arguably stems from the difficulties language learners have in producing fricatives, which leads to a host of avoidance and mutation strategies by individual learners (CHAPTER 101: THE INTERPRETATION

[1] If the target is not specified [−constricted glottis], then emphatics will spirantize and need to be repaired. The analysis of emphatic stops as [+constricted glottis] (laryngeally constricted) appears reasonable because the range of articulations typically reconstructed for them (Rendsburg 1997: 73) is within the laryngeal tract (CHAPTER 25: PHARYNGEALS).

2 Feature specifications

Silbert and de Jong (2008) state that there is a problem with finding a parsimonious feature definition of fricatives: they maintain that because fricatives consist, in large part, of random turbulent noise, they present a challenge to attempts to ground our theory of phonological features in specific phonetic correlates. As far as we can tell, though, the phonological features that are generally considered to delineate the fricative class, [+continuant] and [−sonorant], do not encounter this problem (see also CHAPTER 8: SONORANTS; CHAPTER 13: THE STRICTURE FEATURES). Continuants can be defined straightforwardly as sounds with oral airflow egress, and obstruents as sounds with positive oral pressure buildup. Fricatives conform to both of these definitions, so there is no difficulty in correlating the feature specification of fricatives with well-established phonetic cues.

But do fricatives show evidence of being *phonologically* [+continuant] and [−sonorant]? And do they possess any other invariant feature specifications? In this section we consider phonological evidence bearing on the specifications of fricatives for [continuant] (§2.1), [sonorant] (§2.2), [consonantal] (§2.3), and [spread glottis] (§2.4).

2.1 *Continuance*

Continuance is perhaps the quintessential feature of the fricative class; in fact, Jakobson *et al.* (1952: 43) specified the feature value [+continuant] only for fricatives and not for vowels (or *h*, interestingly), making it effectively equivalent phonologically to Ladefoged's [fricative] feature. Chomsky and Halle (1968: 177) add {r l h} to the continuant set, but still exclude vowels (though their definition of [+continuant] in terms of not having enough constriction to stop airflow might lead us to expect otherwise).

Once we assume that vowels are [+continuant] and incorporate the feature [sonorant] (as Chomsky and Halle 1968: ch. 8 did), we are led to revise our classification of the fricatives to [+continuant, −sonorant]. But are either of these features phonologically active? In this subsection we provide evidence that [continuant] is active in fricatives, drawing on delinking and spreading of [+continuant] (stopping and spirantization, respectively), delinking and spreading of [−continuant] (deaffrication and intrusive stop formation, respectively), and delinking of [αcontinuant] (manner dissimilation). The interested reader can consult Cser (1999) for discussion of additional phenomena bearing on the phonological status of the feature [continuant]. Evidence involving [sonorant] is presented in §2.2.

2.1.1 *Stopping*

Let us first consider the cross-linguistically common process of stopping, which involves changing fricatives to corresponding stops and can be reasonably analyzed as delinking of [+continuant] with subsequent replacement by its complement, [−continuant]. This is common after nasals, as we already saw in Nivkh (and as occurs in Spanish, according to Baković 1994 and Kenstowicz 1994). It is also common as a positional and absolute neutralization process in first

(9) *Word-initial stopping deaffrication*

[dɛli]	*jelly*
[dʌmp]	*jump*
[dus]	*juice*
[dɪp]	*chip*
[diːʒ]	*cheese*
[dɛːə]	*chair*

2.1.2 Spirantization

Besides delinking in the patterns just described, the feature [+continuant] can spread to stop consonants, producing fricatives. In addition to the Tiberian Hebrew case mentioned in §1.1 we arguably find spirantization of this type in Spanish (Harris 1969), producing alternations of the sort in (10):

(10) *Spanish spirantization* (data from Baković 1994)

a.	[beso]	'kiss'	[ese βeso]	'that kiss'
			[el βeso]	'the kiss'
			[dar βesos]	'give kisses'
b.	[dato]	'date'	[ese ðato]	'that date'
			[el dato]	'the date'
			[dar ðatos]	'give dates'
c.	[gato]	'cat'	[ese ɣato]	'that cat'
			[el ɣato]	'the cat'
			[dar ɣatos]	'give cats'

According to Harris (1969: 39), the alternants "appear as continuants except initially and after homorganic noncontinuant sonorants." Scholars generally (e.g. Goldsmith 1981) analyze the system in terms of the feature [+continuant] spreading to underlying voiced stops from preceding [+continuant] segments, with the [l–d] cases being a bit of a problem. (We have encountered speakers who have [lð] in such cases, but the secondary literature appears to be unanimous on [ld] being the only option. See also CHAPTER 13: THE STRICTURE FEATURES.) Lozano (1979) and Baković (1994) invert the analysis, proposing underlying voiced fricatives that undergo fortition in syllable onsets under certain conditions. If they are correct, Spanish still presents evidence for [continuant] being phonologically active in fricatives; the only difference is that the Spanish facts are then an example of stopping (see §2.1.1) rather than spirantization. (But see Mascaró 1991 for a critique of these phonological solutions.)

Because the Spanish continuant allophones are often highly sonorous and even realized as glides (CHAPTER 15: GLIDES), it could be asked whether fricatives are phonologically involved in the alternation at all, or whether the fricative allophones are merely phonetically fortified variants of the glides. But the considerably different behavior of what are conventionally analyzed as underlying glides indicates that this is not the case. Although we have heard the Spanish glide /j/ realized with fortified variants [ʝ] and even [ʤ], the distribution of fricative realizations for /j/ and the strength of their frication appear to be much rarer and less salient than with the fricative realizations of /b d g/, suggesting that phonological [continuant] alternation in obstruents does indeed occur in Spanish.

data requiring it (Fallon 2002; Samuels 2009). This quandary raises the question: when if ever is it acceptable to analyze systematic and categorical (as opposed to variable or gradient) spirantization as [+continuant] spreading from vowels?

The assumption that spreading of otherwise active features must have a certain status in terms of contrast structure seems reasonable, since features can only spread if they are specified, and specification is widely assumed to relate somehow to contrastiveness (CHAPTER 2: CONTRAST). But to claim that vowels cannot spread [+continuant] because they never contrast for it is problematic. This assumption presupposes that there is an algorithm capable of exhaustively defining which features are contrastive in which environments in a given language by specifying them in a particular *order*, leaving the remaining features predictable from these and therefore (at least in some cases) never phonologically specified. Yet no successful algorithm has been developed for this task, *pace* Dresher (2009). The two-best known candidates, the Pairwise Algorithm and the Successive Division Algorithm, are critiqued in Samuels (2009: 77–94), and found to have serious problems. Samuels (2009) essentially abandons the search for such algorithms, while Parker's more limited discussion (CHAPTER 49: SONORITY) favorably highlights an application of the Successive Division Algorithm in Anywa and Dholuo. Crucially, this application makes order of specification depend on language-particular feature activity rather than on contrast as measured by any independent criterion.

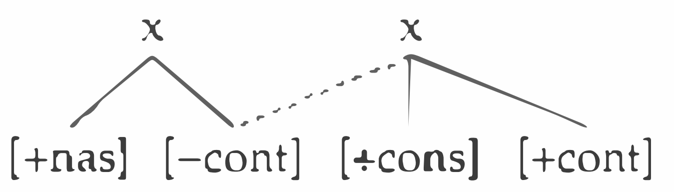
Hence, the argument that a feature cannot be active in a certain environment because it is not contrastive there must be suspended until a proper specification algorithm itself is established. This much is essentially acknowledged by Dresher (2009: 9, 209). After investigating possible algorithms for implementing the "Contrastivist Hypothesis" ("that phonological computation operates only on contrastive features"), Dresher concludes that phonological computation does apparently require non-contrastive features in some situations. Dresher recommends retaining the Contrastivist Hypothesis and a form of the Successive Division Algorithm (requiring a serial grammar theory) because they cover much or possibly most of the data, but once we allow exceptions, any data that are more elegantly treated in contravention of the hypothesis can also demand exceptional status, including spirantizations which can be analyzed as [+continuant] spreading from vowels.

It may eventually prove relevant to distinguish between two kinds of non-contrastive features, if only one of them seems needed in phonological representations. One kind is not predictable from context; an example would be archiphonemically underspecified features (on which see Inkelas 1995; Samuels 2009). The other kind is necessarily predictable from contexts: all vowels, for example, are necessarily continuant. We will not pursue the matter further here. For now it simply seems reasonable to accept cases of spirantization like those mentioned in this section as examples of fricatives produced by [+continuant] spread.

2.1.3 *Deaffrication*

Further evidence that [continuant] is active in fricatives comes from phenomena that delink [−continuant] specifications. Perhaps the most widespread process of this type is the more common variety of deaffrication, which deletes the [−continuant] closure phase of the segment, producing a fricative. We find examples of this in the Aslanbeg dialect of Armenian, which deletes the stop component of affricates in coda position (13).

(16) *Intrusive stop formation as* [−continuant] *spreading*



Here again, the analysis hinges in part on fricatives being [+continuant]; if they were not, the reasons for a preceding nasal changing a fricative into what is essentially an affricate would be unclear.

2.1.5 *Manner dissimilation*

Thus far we have seen evidence for spreading and delinking of [+continuant] and [−continuant], each of which implicates fricatives as being [+continuant]. The [+continuant] specification of fricatives is also revealed in dissimilation processes (CHAPTER 60: DISSIMILATION), such as we find in Modern Greek, where it can either trigger or undergo [continuant] delinking. In this language, voiceless stop + stop and (non-*s*) fricative + fricative clusters optionally dissimilate to stop + fricative (Newton 1972; Kaisse 1988), as in (17) (Tserdanelis 2001) (see also CHAPTER 13: THE STRICTURE FEATURES):

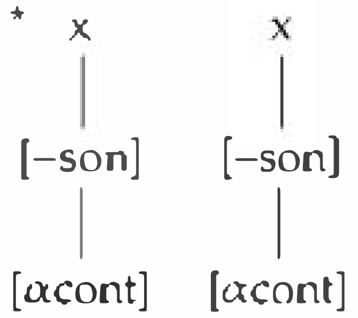
(17) *Modern Greek manner dissimilation*

a.	ptero	~	ftero	'feather'
	ktena	~	xtena	'comb'
	epta	~	efta	'seven'
	okto	~	oxto	'eight'
	ekpiisi	~	expiisi	'sale'
b.	xθes	~	xtes	'yesterday'
	fθinos	~	ftinos	'cheap'
	skefθika	~	skeftika	'I thought'
	anixθika	~	anixtika	'I was opened'
	fxaristo	~	fkaristo	'I thank'
c.	sxini	~	skini	'rope'
	pisθika	~	pistika	'I was convinced'
	sfoŋgos	~	spoŋgos	'sponge'
d.	trex-o	:	e-trek-sa	'I run (PRES : PAST)'
	kafsimo	~	kapsimo	'burning'
	kaθ-izo	:	e-kat-sa	'I sit (PRES : PAST)'

When the underlying cluster contains two stops, the first member changes into a fricative (17a), whereas in fricative clusters the second member becomes a stop (17b). Interestingly, obstruent clusters containing an /s/ invariably delink the [continuant] specification of the other segment (17c, d).

If we assume that stops and fricatives differ in being [−continuant] and [+continuant] respectively, we can interpret the dissimilation facts as OCP-driven delinking of [continuant] values, where the relativized version of the OCP here would look something like (18).

(18) *The OCP manner constraint for Modern Greek* (modified from Tserdanelis 2001)



How exactly violations of this constraint are repaired is a matter of some debate, but the specifics are not relevant here. The key for our purposes is that dissimilation processes of this sort only make sense if stops and fricatives are polar opposites delimited by a single binary feature, in this case [continuant]; other representational options such as [fricative] simply do not capture this sort of interaction insightfully.

In this subsection we have seen that the evidence from spreading and delinking of [±continuant], which surfaces in a broad range of phenomena from spirantization to deaffrication to intrusive stop formation to manner dissimilation, dovetails nicely with the assumption that fricatives are specified as [+continuant]. Competing theories that replace [+continuant] with [fricative], A_f, and the like tend to deal less well with these facts.

2.2 *Sonorance*

We next turn to the other feature value most commonly attributed to fricatives, [−sonorant]. This specification is presumably invoked to distinguish fricatives from the rest of the continuants, e.g. vowels and (in some languages at least) [l] and [r], which are [+sonorant] (CHAPTER 8: SONORANTS). Treating all fricatives as obstruents raises a couple of questions, however:

(19) a. Are the glottal continuants /h ɦ/ obstruents? If not, are they still fricatives?
b. Russian *v* patterns partly with obstruents and partly with sonorants; what is its [sonorant] value?
c. What is the relative sonority ranking of fricatives and stops?
d. Why is the boundary between voiced fricatives and homorganic glides sometimes unclear?

In this subsection we consider each of these issues in turn.

2.2.1 *Glottals*

Trubetzkoy (1939) classed all of the fricatives as obstruents. If the glottal continuants /h ɦ/ are obstruent, then they belong straightforwardly to the class of fricatives defined as [+continuant, −sonorant] segments. If, however, these segments are sonorant, as suggested by Chomsky and Halle (1968) and some of the ancient Indian phonetic treatises, such as the Taittirīya Prātiśākhya (1.13; Whitney 1871), then we must either exclude them from the fricative class or adopt one of the proposals in the literature for encoding fricativity directly, such as Ladefoged's

(1989) [fricative] feature, Articulatory Phonology's fricative gesture (Browman and Goldstein 1986), Steriade's (1993) fricative aperture node A_f, or Moren's (2003) consonantal manner feature [open]. Below we consider phonological evidence bearing on the sonorance of laryngeals, specifically the plain glottals (on the distinction see CHAPTER 25: PHARYNGEALS).

Glottals pattern with obstruents to the exclusion of sonorants in some languages and vice versa in others. The following examples and more like them can be found in the P-Base database (Mielke 2007). Sonorant examples include Kickapoo (Algic) and Supyire Senoufo (Niger-Congo). In Kickapoo, after any sonorant, including /h/ (the only glottal), the second member of a glide plus unaccented vowel sequence in either order is glottalized. In Supyire Senoufo, nasalized vowels can be preceded by only /Ṽ m n ɲ ŋ l ʔ/, and there are no other glottals or consonantal sonorants in the inventory. Obstruent examples include Jordanian Arabic, where only /ʔ h/ and other obstruents can be C_2 and only /m n r l/ can be C_1 in -C_1C_2# and -$C_1C_2C_3$-; Maltese, where only /ʔ h/ and other obstruents can be C_1 in #C_1C_2-; and Balangao (Austronesian), where obstruents including /ʔ h/ delete after the common prefix /maŋ-/. There is also evidence for [−sonorant] spreading from fricatives to the palatal nasal [ɲ] in the Nilo-Saharan language Bilaala (Olson and Schultz 2002).

Similar ambivalence is shown by /ɦ/. In Czech (Slavic: Mielke 2007; Dankovičová 1999), this segment is the only laryngeal in the language and participates in a regressive obstruent voicing assimilation to the exclusion of sonorants. This suggests that /ɦ/ is [−sonorant] in Czech. At the same time, in Oowekyala (Wakashan: Howe 2000), where /ɦ/ is again the only laryngeal in the inventory, the segment appears to be [+sonorant]. This is indicated not only by the fact that /ɦ/ is voiced like all of the sonorants and unlike any of the obstruents in the language, but more importantly, by the fact that /ɦ/ takes [+constricted glottis] root-initially before a reduplicant in the plural, like sonorant consonants but unlike obstruents. Furthermore, the laryngeal continuants [h] and [ɦ] appear to systematically avoid participating in the stopping and deletion phenomena in child phonologies discussed in §2.1.

If obstruence is cued by increased pressure in the vocal tract relative to outside air pressure, then cross-linguistic variation in the classification of glottals as obstruents or sonorants may be due to variation in whether this pressure buildup involves supraglottal impedance or not. In most sounds this is automatically the case, but with glottals it is not; hence, glottals will be sonorants if obstruents are defined by pressure buildup involving supralaryngeal impedance, and obstruents otherwise. Cross-linguistic variation in selection among phonetic cues for the definition of phonological classes is increasingly well established (Mielke 2008: 74–76; Samuels 2009: 70). On the application of this perspective to ambivalent glottal sonorance in particular see also Mielke (2009: 11–12).

In sum, the simplest solution of ambivalent glottal sonorance for the purpose of defining fricatives appears to be to retain the definition of fricatives as continuant obstruents, including glottal continuants in this class when they pattern with other obstruents and not otherwise. When glottal continuants are classed as fricatives on phonological grounds, an available phonetic correlate for [−sonorant] that is consistent with all members of the class is pressure buildup in the vocal tract. When glottal continuants are classed as sonorants on phonological grounds, Subsequent research has identified many additional phenomena consistent with the theory that voiceless fricatives are [+spread glottis] and voiced fricatives are [−spread glottis], but the above examples should suffice to make the point. Neither the traditional Hallean nor the Government Phonology representations of fricatives are able to account for facts of this type.

Beckman and Ringen (2009) and Nicolae and Nevins (2009) suggest a modification to the generalization proposed by Vaux (1998), namely that it holds only for languages where [spread glottis] is active; in languages where obstruents contrast for [voice] rather than [spread glottis], they say, voiceless fricatives are [−voice] (or unspecified for [voice] if it is privative). Both Beckman and Ringen and Nicolae and Nevins base their argument on the fact that in the [voice] languages they have examined phonetically (Finnish for Beckman and Ringen, Russian for both sets of authors), there is no devoicing of following sonorants by voiceless fricatives.

This suggestion (henceforth BRNN, after the surnames of the authors) is intriguing, but unfortunately relies on a single tenuous assumption, namely that if a language contrasts [spread glottis] in obstruents it must spread [spread glottis] from obstruents to following sonorants. The fact that two languages (Russian and Finnish) happen to conform to BRNN is hardly proof of its cross-linguistic validity. If BRNN is true, moreover, it makes an interesting prediction when taken in conjunction with the aforementioned finding that post-nasal voicing (PNV) avoids creating *[+nasal, +spread glottis] configurations (Vaux 1998): unlike in [spread glottis] languages such as New Julfa Armenian, where PNV does not target fricatives, it should freely target voiceless fricatives in languages possessing a [voice] rather than a [spread glottis] contrast in their obstruent system (i.e. systems like those of French, Japanese, or Modern Greek, which oppose unaspirated voiced and voiceless stops). This prediction is not borne out: though we know of one so-called [voice] language where PNV applies to fricatives (Nande; Hyman 2003), in all other cases known to us fricatives differ from stops in *not* undergoing PNV, exactly as we find in [spread glottis] languages and counter to the prediction of BRNN. In Modern Greek, for example, sequences of nasal consonant + voiceless fricative either delete the nasal (optional but preferred across word boundaries, as in /ton θe'o/ 'the god (ACC)' → [toθe'o]) or assimilate the nasal to the fricative in place of articulation, with no voicing (word-internally, as in /sin-xo'ro/ 'forgive' → [siŋxo'ro]) (Holton *et al.* 2004).

The BRNN typology has a phonetic dimension as well. Instead of showing that voiceless fricatives (or some of them) are [+spread glottis] only in languages where stops are [+spread glottis], their typology could be reinterpreted as showing that vocal fold abduction overshoots fricative release into a following sonorant consonant in languages where it also overshoots stop release into a following vowel. It is worth asking why we should use gestural overshoot from one segment into a neighboring segment as a diagnostic for that gesture's specification being phonological.

Van Oostendorp (2007) proposes a further modification for Dutch, wherein stops are opposed for [voice], but fricatives are opposed for [spread glottis]. This system is designed to account for the peculiar properties of voicing assimilation in Dutch, which seem to implicate length contrasts as well. If van Oostendorp's proposal is on the right track, we should be able to investigate whether Dutch voiceless fricatives trigger devoicing of following sonorants as a way to explore the BRNN theory further.

What emerges from the research documented in this subsection is that there is strong evidence for a distinction in [spread glottis] specifications between voiced and voiceless fricatives, at least in languages where [spread glottis] is phonologically active. This being the case, we cannot include [spread glottis] values in the set of invariant specifications for the class of fricative consonants.

3 Conclusions

We have argued that the apparently straightforward characterization of fricatives as [+continuant, −sonorant, +consonantal] and perhaps some value of [spread glottis] turns out on closer inspection to be problematic, but the simpler specification [+continuant, −sonorant] appears adequate if glottal continuants are allowed cross-linguistically variable and primarily phonologically determined membership in the fricative class. Fricatives are generally also [+consonantal], with some peripheral gray areas where the evidence is sparse (§2.3). Finally, it appears that voiceless but not voiced fricatives are generally [+spread glottis], at least in languages where the feature is contrastive, but on this point too the evidence is problematic.

ACKNOWLEDGMENTS

Thanks to Kevin Hervig, Petros Karatsareas, Andrew Nevins, Keren Rice, and Patrick Taylor for comments on drafts of this chapter.

REFERENCES

Austerlitz, Robert. 1956. Gilyak nursery words. *Word* 12. 260–279.

Baković, Eric. 1994. Strong onsets and Spanish fortition. *MIT Working Papers in Linguistics* 23. 21–39.

Ball, Martin J., Joan Rahilly & Paul Tench. 1996. *The phonetic transcription of disordered speech.* San Diego: Singular Publications.

Beckman, Jill N. & Catherine Ringen. 2009. A typological investigation of evidence for [sg] in fricatives. Paper presented at the 17th Manchester Phonology Meeting. Handout available (July 2010) at www.siue.edu/~khildeb/BeckmanRingen_17.pdf.

Bermúdez-Otero, Ricardo. 2001. Voicing and continuancy in Catalan: A nonvacuous Duke-of-York gambit and a Richness-of-the-Base paradox. Unpublished ms., University of Manchester. Available (July 2010) at www.bermudez-otero.com/Catalan.pdf.

Blevins, Juliette. 1993. Klamath laryngeal phonology. *International Journal of American Linguistics* 59. 237–279.

Booij, Geert. 1995. *The phonology of Dutch.* Oxford: Clarendon Press.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Buckley, Eugene. 1994. *Theoretical aspects of Kashaya phonology and morphology.* Stanford: CSLI.

Celce-Murcia, Marianne. 1977. Phonological factors in vocabulary acquisition: A case study of a two-year-old, English-French bilingual. *Working Papers in Bilingualism* 13. 27–41.

Charney, Jean Ormsbee. 1993. *A grammar of Comanche.* Lincoln: University of Nebraska Press.
Chiat, Shulamuth. 1989. The relation between prosodic structure, syllabification and segmental realization: Evidence from a child with fricative stopping. *Clinical Linguistics and Phonetics* 3. 223–242.
Chin, Steven & Daniel Dinnsen. 1991. Feature geometry in disordered phonologies. *Clinical Linguistics and Phonetics* 5. 329–337.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.
Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.
Clements, G. N. 1987. Phonological feature representation and the description of intrusive stops. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 29–50.
Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 283–333. Cambridge: Cambridge University Press.
Clements, G. N. 2006. The feature [strident]. Unpublished ms., CNRS. Available (July 2010) at www.ucalgary.ca/dflynn/files/dflynn/Clements06.pdf.
Clements, G. N. & Sylvester Osu. 2002. Explosives, implosives and nonexplosives: The linguistic function of air pressure differences in stops. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology 7*, 299–350. Berlin & New York: Mouton de Gruyter.
Coats, Herbert S. & Alex P. Harshenin. 1971. On the phonological properties of Russian *v*. *Slavic and East European Journal* 15. 466–478.
Coetzee, Andries W. 1999. *Tiberian Hebrew phonology: Focusing on consonant clusters.* Assen: van Gorcum.
Connell, Bruce. 2007. Mambila fricative vowels and Bantu spirantization. *Africana Linguistica* 13. 7–31.
Cser, András. 1999. On the feature [continuant]. *Theoretical Linguistics* 25. 215–234.
Cser, András. 2003. *The typology and modelling of obstruent lenition and fortition processes.* Budapest: Akadémiai Kiadó.
Dankovičová, Jana. 1999. Czech. In *Handbook of the International Phonetic Association*, 70–73. Cambridge: Cambridge University Press.
Davidson, Lisa, Peter W. Jusczyk & Paul Smolensky. 2006. Optimality in language acquisition I: The initial and final states of the phonological grammar. In Paul Smolensky & Géraldine Legendre (eds.) *The harmonic mind*, vol. 2, 231–278. Cambridge: MIT Press.
de Lacy, Paul (ed.) 2007. *The Cambridge handbook of phonology.* Cambridge: Cambridge University Press.
Dell, François & Mohamed Elmedlaoui. 1985. Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7. 105–130.
Dell, François & Mohamed Elmedlaoui. 2002. *Syllables in Tashlhiyt Berber and in Moroccan Arabic.* Dordrecht: Kluwer.
Dinnsen, Daniel A. 2001. New insights from Optimality Theory for acquisition. *Clinical Linguistics and Phonetics* 15. 15-18.
Dinnsen, Daniel A. & Ashley W. Farris-Trimble. 2008. An opacity-tolerant conspiracy in phonological acquisition. *Indiana University Working Papers in Linguistics* 6. 99–118.
Dinnsen, Daniel A., Kathleen O'Connor & Judith Gierut. 2001. The puzzle-puddle-pickle problem and the Duke-of-York gambit in acquisition. *Journal of Linguistics* 37. 503–525.
Dodd, Barbara. 1995. Procedures for classification of subgroups of speech disorder. In Barbara Dodd (ed.) *Differential diagnosis and treatment of children with speech disorder*, 49–64. London: Whurr.
Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology.* Cambridge: Cambridge University Press.

Jakobson, Roman. 1957. Notes on Gilyak. *Bulletin of the Institute of History and Philology, Academia Sinica* 29. 255–281. Reprinted 1971 as Roman Jakobson. *Selected writings II: Word and language*, 72–97. The Hague & Paris: Mouton.

Jakobson, Roman. 1978. Mutual assimilation of Russian voiced and voiceless consonants. *Studia Linguistica* 32. 107–110.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Johnson, Keith. 1997. *Acoustic and auditory phonetics*. Cambridge, MA: Blackwell.

Jones, Mark & Carmen Llamas. 2008. Fricated realizations of /t/ in Dublin and Middlesbrough English: An acoustic analysis of plosive frication and surface fricative contrasts. *English Language and Linguistics* 12. 419–443.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kaisse, Ellen M. 1988. Modern Greek continuant dissimilation and the OCP. Unpublished ms., University of Washington.

Kaisse, Ellen M. 1992. Can [consonantal] spread? *Language* 68. 313–332.

Kaye, Jonathan. 2000. A users' guide to Government Phonology. Unpublished ms., University of Ulster.

Kelly, J. 1974. Close vowels in Fang. *Bulletin of the School of Oriental and African Studies* 37. 119–123.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Kloster-Jensen, Martin. 1991. Letter to the editor. *Journal of the International Phonetics Association* 21. 42.

Ladefoged, Peter. 1962. Sub-glottal activity during speech. In A. Sovijärvi & P. Aalto (eds.) *Proceedings of the 4th International Congress of Phonetic Sciences*, 73–91. The Hague: Mouton.

Ladefoged, Peter. 1989. Representing phonetic structure. *UCLA Working Papers in Phonetics* 73.

Ladefoged, Peter. 2001. *A course in phonetics*. 4th edn. Boston: Heinle & Heinle.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford & Malden, MA: Blackwell.

Laufer, Asher. 1991. Phonetic representation: Glottal fricatives. *Journal of the International Phonetic Association* 21. 91–93.

Lavoie, Lisa. 2001. *Consonant strength: Phonological patterns and phonetic manifestations*. New York: Garland.

Locke, John L. 1983. *Phonological acquisition and change*. New York: Academic Press.

Lozano, María del Carmen. 1979. Stop and spirant alternations: Fortition and spirantization processes in Spanish phonology. Ph.D. dissertation, Ohio State University.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Maniwa, Kazumi, Allard Jongman & Travis Wade. 2009. Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America* 125. 3962–3973.

Marlett, Stephen A. 2001. An introduction to phonological analysis. Unpublished ms., Summer Institute of Linguistics. Available (July 2010) at www.sil.org/mexico/ling/phonology-marlett/e004-phonall.pdf.

Martínez-Celdrán, Eugenio. 2008. Some chimeras of traditional Spanish phonetics. In Laura Colantonni & Jeffrey Steele (eds.) *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, 32–46. Somerville, MA: Cascadilla Press.

Mascaró, Joan. 1991. Iberian spirantization and continuant spreading. *Catalan Working Papers in Linguistics* 1991. 167–179.

McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.

McCarthy, John J. 2008. *Doing Optimality Theory*. Malden, MA: Wiley-Blackwell.

Mielke, Jeff. 2007. P-base 1.92. Available (July 2010) at http://aixl.uottawa.ca/~jmielke/pbase/index.html.

Mielke, Jeff. 2008. *The emergence of distinctive features.* Oxford: Oxford University Press.

Mielke, Jeff. 2009. A phonetically-based phonetic similarity metric. Paper presented at the 40th Annual Meeting of the North East Linguistic Society, MIT.

Mołczanow, Janina. 2007. The problem of the Russian labial fricatives: Why V is different. *Poznań Studies in Contemporary Linguistics* 43. 49–74.

Morén, Bruce. 2003. The parallel structures model of feature geometry. *Working Papers of the Cornell Phonetics Laboratory* 15. 194–270.

Naidoo, Shamila. 2005. Intrusive stop formation in Zulu: An application of feature geometry theory. D.Litt. dissertation, Stellenbosch University.

Newton, Brian. 1972. *The generative interpretation of dialect: A study of Modern Greek phonology.* Cambridge: Cambridge University Press.

Nicolae, Andreea & Andrew Nevins. 2009. The phonetics and phonology of fricative non-neutralization in Turkish. Paper presented at the 40th Annual Meeting of the North East Linguistic Society, MIT.

O'Grady, William & Michael Dobrovolsky. 1992. *Contemporary linguistic analysis: An introduction.* Toronto: Copp Clark Pitman Ltd.

Olson, Kenneth S. & Paul H. Schultz. 2002. Can [sonorant] spread? In J. Albert Bickford (ed.) *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 46. Available (July 2010) at www.und.edu/dept/linguistics/wp/2002.htm.

Oostendorp, Marc van. 2007. Exceptions to final devoicing. In Jeroen van de Weijer & Erik Jan van der Torre (eds.) *Voicing in Dutch: (De)voicing – phonology, phonetics, and psycholinguistics*, 81–98. Amsterdam & Philadelphia: John Benjamins.

Padgett, Jaye. 2002. Russian voicing assimilation, final devoicing, and the problem of [v]. Unpublished ms., University of California, Santa Cruz (ROA-528).

Padgett, Jaye. 2008. Glides, vowels, and features. *Lingua* 118. 1937–1955.

Paterson, Marietta. 1994. Articulation and phonological disorders in hearing-impaired school-aged children with severe and profound sensorineural losses. In Nicholas Bankson & John Bernthal (eds.) *Child phonology: Characteristics, assessment, and intervention with special populations*, 199–226. New York: Thieme.

Petrova, Olga. 1997. Russian voice assimilation in Optimality Theory. Unpublished ms., University of Iowa.

Pike, Kenneth L. 1943. *Phonetics: A critical analysis of phonetic theory and a technic for the practical description of sounds.* Ann Arbor: University of Michigan Press.

Plapp, Rosemary K. 1999. Lexical Phonology and Optimality Theory: Analysis of Russian. Ph.D. dissertation, University of Iowa.

Poliquin, Gabriel. 2006. Canadian French vowel harmony. Ph.D. dissertation, Harvard University.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar.* Malden, MA & Oxford: Blackwell.

Pyatt, Elizabeth J. 1997. An integrated model of the syntax and phonology of Celtic mutation. Ph.D. dissertation, Harvard University.

Ramsay, Gordon. 2009. The influence of dynamic flow conditions on sound generation in fricative consonants. *Journal of the Acoustical Society of America* 125. 2499.

Rendsburg, Gary A. 1997. Ancient Hebrew phonology. In Alan S. Kaye (ed.) *Phonologies of Asia and Africa*, 65–83. Winona Lake: Eisenbrauns.

Riggle, Jason. 2008. Phonological features. Unpublished ms., University of Chicago.

Salus, Peter & Mary Salus. 1974. Developmental neurophysiology and phonological acquisition order. *Language* 50. 151–160.

Samuels, Bridget. 2009. The structure of phonological theory. Ph.D. dissertation, Harvard University.

Shadle, Christine. 1985. The acoustics of fricative consonants. Ph.D. dissertation, MIT.

Shiraishi, Hidetoshi. 2006. Topics in Nivkh phonology. Ph.D. dissertation, University of Groningen.

Silbert, Noah Haskell & Kenneth J. de Jong. 2008. Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustic Society of America* 123. 2769–2779.

Smith, Neil V. 1973. *The acquisition of phonology: A case study.* Cambridge: Cambridge University Press.

So, Lydia K. H. & Barbara J. Dodd. 1994. Phonologically disordered Cantonese-speaking children. *Clinical Linguistics and Phonetics* 8. 235–255.

Steriade, Donca. 1993. Segments, contours and clusters. In André Crochetière, Jean-Claude Boulanger & Conrad Ouellon (eds.) *Proceedings of the 15th International Congress of Linguistics*, 71-82. Quebec: Université Laval.

Stevens, Kenneth N. 1998. *Acoustic phonetics.* Cambridge, MA: MIT Press.

Svantesson, Jan-Olof, Anna Tsendina, Anastasia Karlsson & Vivan Franzén. 2005. *The phonology of Mongolian.* Oxford: Oxford University Press.

Thurneysen, Rudolf. 1946. *A grammar of Old Irish.* Dublin: Dublin Institute for Advanced Studies.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie.* Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology.* Berkeley & Los Angeles: University of California Press.

Tserdanelis, Georgios. 2001. A perceptual account of manner dissimilation in Greek. *OSU Working Papers in Linguistics* 55. 172–199.

Tsuchida, Ayako. 1997. Phonetics and phonology of Japanese vowel devoicing. Ph.D. dissertation, Cornell University.

Vaux, Bert. 1992. Adjarian's Law and Consonantal ATR in Armenian. In John Greppin (ed.) *Proceedings of the 4th International Conference on Armenian Linguistics*, 271–293. Delmar, NY: Caravan.

Vaux, Bert. 1993. The phonology of voiced aspirates in the New Julfa dialect of Armenian. Paper presented at the 10th Anniversary Conference of the *Association Internationale des Études Arméniennes*, School of Oriental and African Studies, London. Published 1997 in Nicholas Awde (ed.) *Armenian perspectives*, 231–248. Richmond, Surrey: Curzon.

Vaux, Bert. 1996. The status of ATR in feature geometry. *Linguistic Inquiry* 27. 175–182.

Vaux, Bert. 1998. The laryngeal specifications of fricatives. *Linguistic Inquiry* 29. 497–511.

Vaux, Bert. 2001. The Armenian dialect of Aslanbeg. *Annual of Armenian Linguistics* 21. 31–64.

Vaux, Bert & Andrew Wolfe. 2009. The appendix. In Eric Raimy & Charles Cairns (eds.) *Contemporary views on architecture and representations in phonology*, 101–143. Cambridge, MA: MIT Press.

Verner, Karl. 1875. Eine Ausnahme der ersten Lautverschiebung. *Zeitschrift für Vergleichende Sprachforschung* 23. 97–130.

Whitney, William Dwight. 1862. The atharva-veda-prātiçākhya, or Çāunikīyā caturādhyāyikā: Text, translation, and notes. *Journal of the American Oriental Society* 7.

Whitney, William Dwight. 1871. Taittirīya-prātiśākhya. *Journal of the American Oriental Society* 9.

Zamma, Hideki. 2004. Allomorphy, spirantization and palatalization in English: With or without geometry. *Tsukuba English Studies* 22. 371–384.

Zec, Draga. 2007. The syllable. In de Lacy (2007), 161–194.

29 Secondary and Double Articulation

Jeroen van de Weijer

This chapter deals with two categories of "complex segments," viz. consonants with secondary articulation, such as /kʷ/ or /lˠ/ (= /ɫ/ 'dark *l*'), and consonants that involve two articulations of equal status, such as /k͡p/. The main difference between these two types of sounds is that in the former there is a major ("consonantal") articulatory stricture on which a vowel-like minor articulation is superimposed, while in the latter the two articulations have a stricture type of equal status (typically, stop or nasal). While consonants with secondary articulation are very common in the languages of the world, consonants with double articulation are much rarer.

I will discuss these segment types in turn below: secondary articulation in §1 and double articulation in §2. I will focus on the phonological representation of both types of segments, as brought out by their function in phonemic inventories and their behavior in assimilation rules and syllable structure restrictions. Note, finally, that two other types of modifiication, which do not involve place of articulation but laryngeal setting, viz. aspiration and glottalization (or laryngealization), will not be covered in the present chapter (see CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION and CHAPTER 111: LARYNGEAL CONTRAST IN KOREAN for discussion). §3 offers a brief conclusion.

1 Secondary articulation

This section discusses segments with secondary articulation. After a short introduction to the phonetic phenomenon of co-articulation, which often gives rise to segments with secondary articulation historically, we will discuss and illustrate the four main types (labialization, velarization, palatalization, and pharyngealization). Then we turn to different ways in which such segments have been represented in the literature, starting from Chomsky and Halle (1968), through Clements and Hume (1995) and recent proposals in Dependency Phonology.

1.1 *Differentiating co-articulation and secondary articulation*

A common phonetic process is "co-articulation" (see e.g. Kühnert and Nolan 1999): the anticipatory shaping of the speech organs to accommodate sounds that will be articulated in the very near future. Co-articulation may also be perseverative, in which case it refers to the inertia of the speech organs after articulating certain sounds. For example, round vowels will typically impose their rounding on neighboring consonants, e.g. the /t/ and /n/ in a word like English *cartoon*, which has medial [t^w], while front vowels may impose a palatal quality on such consonants, as in *canteen*, with [t^j] (CHAPTER 71: PALATALIZATION; CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). We can refer to such processes as involving allophonic secondary articulation. Other types of secondary articulation are velarization ([$t^ɣ$]), which is similarly induced by back vowels, and pharyngealization ([$t^ʕ$]), which is associated with low vowels. If the secondary articulation on the consonant becomes independent of its triggering environment, its status may become phonemic (see Padgett 2003 for a description of this historical process in Russian; see also CHAPTER 121: SLAVIC PALATALIZATION). In this chapter we focus on phonemic types of secondary articulation. For instance, while the difference in velarization between clear *l* and dark *ł* is allophonic in English (which is, of course, determined not by vowel quality but (mostly) by syllabic position in this language, with clear *l* occurring in the onset and dark *ł* in the coda), in Marshallese, a Malayo-Polynesian language of the Marshall Islands, these two sounds are phonemic (Ladefoged and Maddieson 1996: 363); see also CHAPTER 31: LATERAL CONSONANTS.

(1) *Marshallese phonemic velarization contrast in laterals*

łał 'knock'
lał 'earth'

Note that in this language this particular type of secondary articulation occurs on a range of segments, viz. all the bilabials and all the coronals in the language (rather than just on the lateral illustrated in (1)), something which is quite common in languages (see Hinskens and van de Weijer 2003 for more examples and discussion). Other languages that have segments with phonemic secondary articulation are Russian (e.g. Kenstowicz 1994: 41; Padgett 2003) and Irish (Ní Chiosáin 1991, 1994, among others). In both languages, palatalized consonants contrast with velarized ones. Examples from Russian are given in (2):

(2) m^jat^j 'to rumble'
$m^ɣat^j$ 'mother'
$m^ɣat^ɣ$ 'checkmate'
$m^jat^ɣ$ 'rumpled'

Many dialects of Arabic have a contrast between plain and pharyngealized consonants, traditionally also referred to as "emphatic" consonants (see e.g. Watson 2002; CHAPTER 25: PHARYNGEALS). In Arabic these consonants vary between velarized and fully pharyngealized. Examples (taken from Sykes 2008) are given in (3):

(3) *Arabic minimal pairs*

pharyngealized		*plain*	
sˤafiːr	'whistle'	safir	'ambassador'
dˤam	'gloss'	dam	'blood'
tˤiːn	'mud'	tiːn	'fig'

The final common type of secondary articulation is labialization or rounding. This is especially common in dorsal (i.e. velar or uvular) consonants, although labial or coronal consonants may also be phonemically labialized. A language illustrating phonemic labialization is Kwakw'ala, a Wakashan language spoken on Vancouver Island (Grubb 1977; Ladefoged and Maddieson 1996):[1]

(4)

velar	*labialized velar*	*uvular*	*labialized uvular*
kasa	kʷesa	qesa	qʷesa
'beat soft'	'splashing'	'coiling'	'peeling'
gisgas	gʷesu	ɢaɢas	ɢʷalas
'incest'	'pig'	'grandparent'	'lizard'
xesa	xʷasa	χasa	χʷat'a
'lost'	a dance	'rotten'	'sparrow'
k'ata	kʷ'esa	q'asa	qʷ'asa
'writing'	'light (weight)'	'sea otter'	'crying'

The most common types of secondary articulation are thus labialization, velarization, palatalization, and pharyngealization. Since labialization does not involve the tongue body, it can be combined with other types of secondary articulation. Marshallese, for instance, is described as having, besides a range of velarized consonants, labio-velarized segments such as /nˠʷ/.[2] Labio-palatalization can appear as an allophonic variant of labialization in front vowel contexts (Ladefoged and Maddieson 1996: 356), and labio-dentalization has been described for Kom and Kuteb by Ladefoged (1968). Finally, Ladefoged and Maddieson mention sulcalization as a possible type of secondary articulation, involving deep grooving of the back of the tongue (Catford 1977). This occurs in English [s], but is not known to be a contrastive type of secondary articulation in any language.

Finally in this section, a note on terminology. The term "palatalization" can be used in at least two ways: (i) as a historical process in which coronal or velar consonants shift to palatal or palato-alveolar place of articulation under the influence of non-low front vowels, and (ii) as a synchronic or diachronic process in which consonants acquire a secondary palatal articulation (and sometimes also affrication). The latter type of change leads, or may lead, to a synchronic state in which some consonants have this type of secondary articulation. This is what is investigated here; see CHAPTER 71: PALATALIZATION for a description of the first type of palatalization (i.e. as a process), and CHAPTER 121: SLAVIC PALATALIZATION for an analysis of different palatalization rules in the Slavic languages.

[1] Ladefoged and Maddieson (1996) note that other sources report the plain velars and uvulars as palatalized.

[2] See for further discussion Hale (2000, 2007: ch. 5); Hale and Reiss (2008: ch. 6). The Marshallese pattern would appear to form a counterexample to Trubetzkoy's (1939) supposition that labialization and velarization never contrast on the same segment.

1.2 Phonological representation

This section deals with the question of how segments with secondary articulation should be represented in terms of features. First, we will investigate their representation in Chomsky and Halle (1968; *SPE*), from which we will move on to feature-geometrical approaches, and finally end with a discussion of secondary articulation in frameworks embracing the dependency relation.

1.2.1 *Preliminaries and* SPE

Pre-theoretically, segments with secondary articulation (usually consonants) would appear to involve some combination of a consonantal and a vocalic articulation. Recall that they are often formed as a result of phonetic co-articulation, involving some kind of "compression" of two segments into one: a front vowel may leave a "trace" on a neighboring consonant and, after the vowel disappears, secondary articulation remains. This suggests that they belong to a class of "complex segments," to which affricates (CHAPTER 16: AFFRICATES), prenasalized stops (CHAPTER 23: PARTIALLY NASAL SEGMENTS) and perhaps other segments, such as clicks (CHAPTER 18: THE REPRESENTATION OF CLICKS), also belong.[3] Such segments are usually analyzed as involving branching somewhere in their segmental representation:

(5) 

Although not all authors agree even on the existence of complex segments in phonological representations (cf. e.g. Kehrein 2002 for discussion), in the present section we will approach secondary articulation from a traditional segmental perspective, and discuss its representation in a number of theories.

Let us start with the representation of secondary articulation in *SPE*. Just like "syllable," the term "secondary articulation" does not appear in the subject index, although the topic is discussed in some detail (*SPE*: 305ff.). In *SPE*, the features [high], [low], and [back] play a dual role: for tongue body consonants (often referred to as dorsals) these features are used to distinguish between palatals, velars, uvulars, and pharyngeals, in the following manner (*SPE*: 305):

(6) *Dorsals in* SPE

	palatals	velars	uvulars	pharyngeals
high	+	+	−	−
low	−	−	−	+
back	−	+	+	+

All of the dorsal consonants are [−anterior]. Chomsky and Halle go on to investigate the role of the features [high], [low], and [back] for other classes of sounds (i.e. non-dorsals) and propose that they:

> may be used in a natural manner to characterize subsidiary consonantal articulations such as palatalization, velarization, and pharyngealization. These subsidiary

[3] And perhaps /s/ plus stop clusters, as a kind of "reversed affricates"; see van de Weijer (1996) and CHAPTER 38: THE REPRESENTATION OF SC CLUSTERS for discussion.

> articulations consist in the superimposition of vowel-like articulations on the basic consonantal articulation. In palatalization the superimposed subsidiary articulation is [i]-like; in velarization, [i]-like; and in pharyngealization, [a]-like. The most straightforward procedure is, therefore, to express these superimposed vowel-like articulations with the help of the features "high," "low," and "back," which are used to characterize the same articulations when they appear in the vowels. (*SPE*: 305–306)

A number of examples are given in (7) (from *SPE*: 307):

(7) *Secondary articulation in* SPE

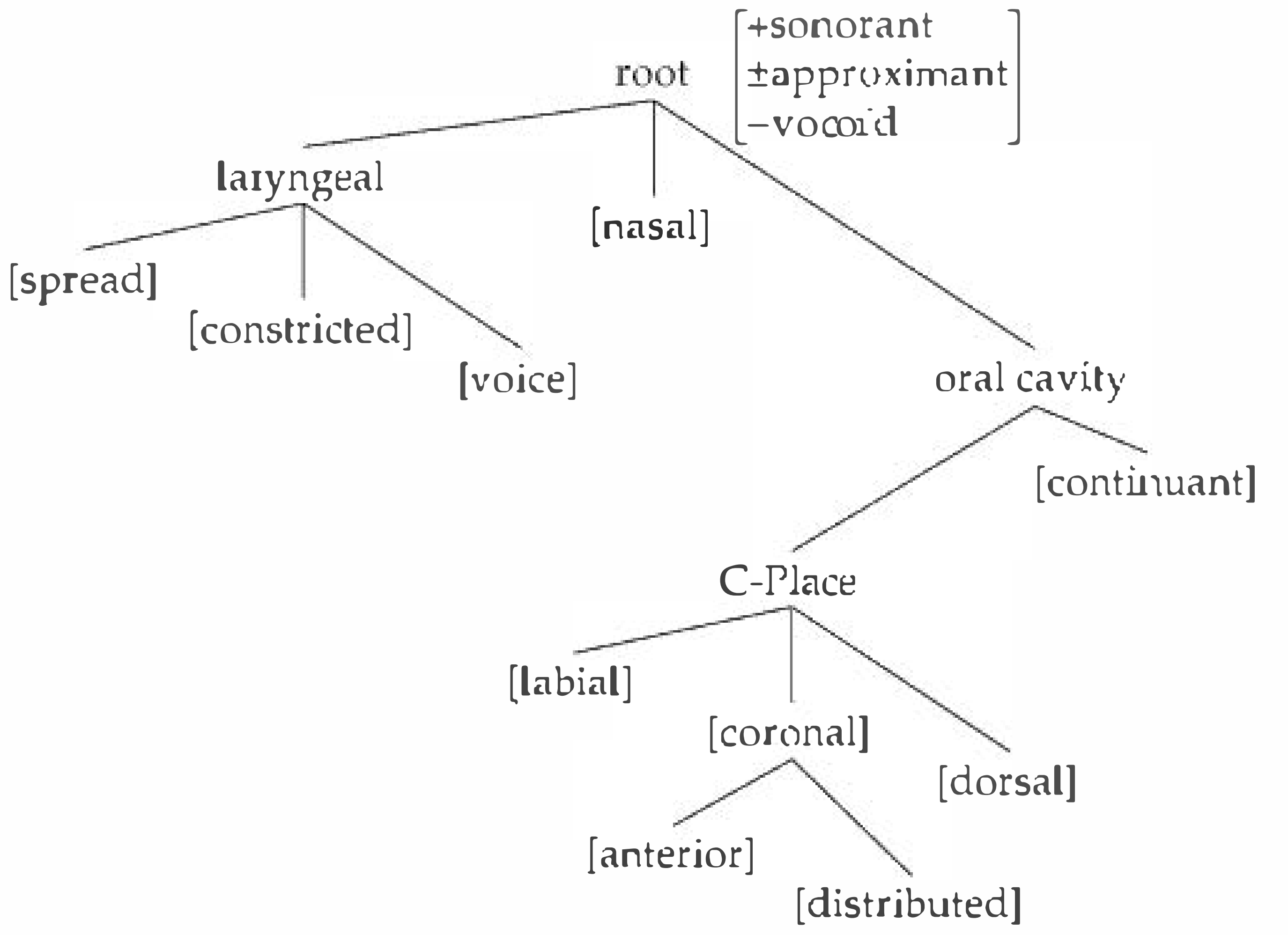
	anterior	coronal	high	low	back
palatalized labials	+	–	+	–	–
palatalized dentals	+	+	+	–	–
velarized labials	+	–	+	–	+
velarized dentals	+	+	+	–	+
velarized palato-alveolars	–	+	+	–	+
uvularized labials	+	–	–	–	+
pharyngealized dentals	+	+	–	+	+

Thus segments with secondary articulation are *not* analyzed as involving segmental branching in *SPE*: rather, the presence of certain vowel features with specific values will result in such segments. The rigidity of *SPE*'s feature matrix was criticized in Campbell (1974), who proposed a "complex symbol" representation for segments with secondary articulation (and for affricates). This and similar critiques led to the development of autosegmental phonology (Leben 1973; Goldsmith 1976; see also CHAPTER 14: AUTOSEGMENTS; CHAPTER 45: THE REPRESENTATION OF TONE), in which features could have scope over more than one segment, and in which opposite feature values could occur in the same segment (first illustrated for tonal features, and later extended to other features).

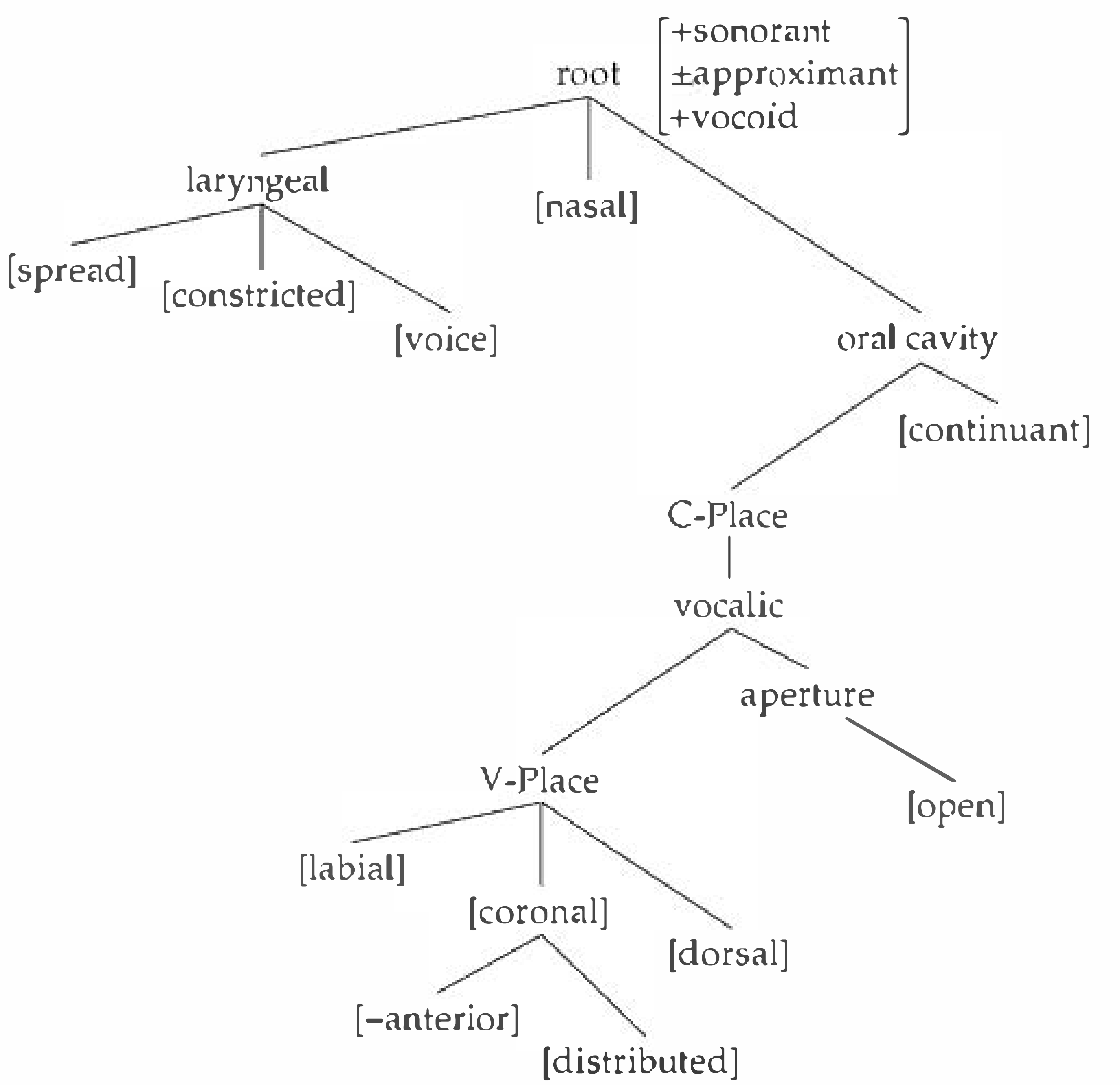
The representation of primary places of articulation and of secondary articulation in *SPE* raises obvious questions (CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION). Velars are represented as [+back], i.e. non-anterior, non-coronal, consonants (*SPE*: 307).[4] Palatalized consonants are represented as [–back] consonants. At first sight, this would entail that palatalized velars are represented as [+back, –back], a combination of opposite values that is ruled out for other features as a simple articulatory impossibility. The same problem is pointed out by Campbell (1974), who argues that the problem is compounded by the impossibility, in *SPE*, of ruling out, for instance, a contrast between palatalized palatals and non-palatalized palatals without giving up the articulatory underpinning of the phonetic feature framework (cf. also McCarthy 1991: 82 in the context of the representation of guttural consonants). Campbell (1974) argues that such contrasts are found, e.g. in Livonian (which contrasts /ʃ/ and /ʃ^{j}/) and Mordvin, which contrasts /c/ and /c^{j}/.

[4] Note that the representation of velars as [+back] makes *prima facie* false predictions with respect to their behavior in vowel harmony: e.g. in a front–back harmony system all velars (plain or palatalized) would be predicted to act as blockers in a standard, autosegmental treatment of harmony. See below for the role of consonants with secondary articulation in vowel harmony processes.

(8) *Consonantal feature tree* (Clements and Hume 1995: 292)

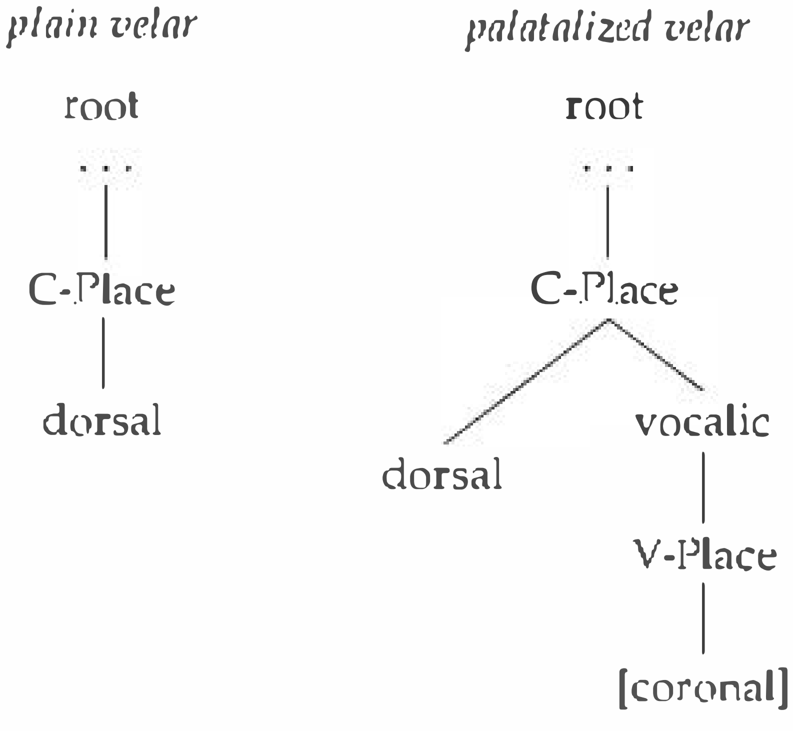


(9) *Vocoid feature tree* (Clements and Hume 1995: 292)



The representation of consonants with secondary articulation is an important point for Clements and Hume. In (10), I give the representation of a plain velar and a palatalized velar, on the basis of Clements and Hume (1995: 288):

(10) *Secondary articulation*



As (10) shows, consonants with secondary articulation are represented as branching structures (cf. (5) above): one part is labeled as relatively consonant-like and the other part as relatively vowel-like: the highest place node receives the interpretation of the major class features (C-Place in (10)). This gives us a welcome perspective on possible types of secondary articulation, because the set of features below the V-Place node is predicted to correspond to possible types of secondary articulation. This prediction is borne out: the presence of the features [labial], [coronal], and [dorsal] in such structures corresponds straightforwardly to labialization, palatalization, and velarization, respectively. Clements and Hume (1995: 287) note that the fourth common type of secondary articulation, that of pharyngealization (cf. (3) above), can be expressed by a feature [pharyngeal], which is not included in the tree model in (9): they discuss its arboreal affiliation (1995: 273ff.), and remain uncommitted as to its final locus (cf. also Hayward and Hayward 1989 and McCarthy 1991 on the feature [guttural] and the representation of this class of segments). Other, less common, types of secondary articulation may be captured by combinations of features under a V-Place node (such as concomitant labialization and velarization) or by the supplementary use of features below the aperture node (which regulates vowel height by way of $[open]_x$ features (which represent degree of openness in their model) in the absence of features like [high] and [low]), such as uvularization.

This model expresses in a straightforward way the phenomenon that co-articulation of a consonant and vowel in time may lead to phonemic secondary articulation: in such a case the vocalic place features will have become permanently part of the consonant in question. It also accounts directly for the role of consonants with secondary articulation in vowel harmony processes: for instance, consonants which are palatalized will introduce a [−back] feature on the harmony process, while velarized consonants will favor back vowels (see e.g. Clements and Sezer 1982 for a discussion of Turkish, and van der Hulst and van de Weijer 1995 for general discussion of this point). This is expressed by the fact that the vocalic features on a consonant with secondary palatalization will be on the same autosegmental tier as the vocalic features of vowels, and therefore predicted to interact directly.

Some further remarks can be made with respect to the representation of consonants with secondary articulation, which we will briefly discuss in turn. First, note that (8) and (9) present *two* distinct feature trees, with partly disjunctive feature sets, rather than one segmental tree for all segments. This has a number of formal consequences that we will not go into here (but cf. Casali 1995; CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). Although the proposal nicely captures the interaction between consonants with secondary articulation with vowels in vowel harmony, it does not technically predict interaction between e.g. primary labials and round vowels. Moreover, some stipulation may be needed as to the level(s) at which branching may or may not take place.

Secondly, as has been pointed out before in the literature (e.g. van der Hulst and van de Weijer 1995), the C-Place node in vowels seems to have only a diacritic function, in that it prevents consonant-to-consonant place spreading, while permitting vowel-to-vowel assimilation (as in vowel harmony, which spreads from truly adjacent vocalic node to vocalic node). The C-Place node in vowels is therefore the price that is paid for capturing the vowel harmony effect described above.

Finally, consider the fact that the two parts of a consonant with secondary articulation, i.e. the consonantal part and the vocalic part, seem to have an unequal status: the segment as a whole acts as a consonant in terms of its function in syllable structure and must be specified as such by way of major class features or otherwise. An alternative way of specifying the "inequality" of the two parts is explored below.

To conclude this subsection, feature-geometrical approaches are much better equipped than earlier theories to capture complexity in segmental representation in general, and of consonants with secondary articulation in particular.

1.2.3 *Dependency Phonology*

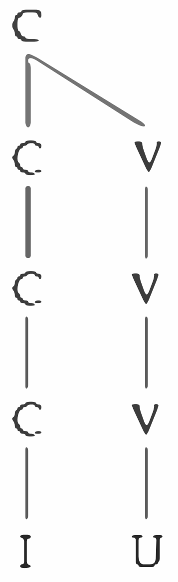
Central to frameworks such as Dependency and Government Phonology (see references below) is the insight that most or perhaps all linguistic structure is headed: syntactic phrases consist of specifiers, complements and heads, morphological constructions are (often) right-headed (cf. Williams 1981), a nucleus is obligatory in the syllable and there is a strong–weak relation between syllables in a foot. Headship is manifested in different ways (e.g. obligatoriness, phonetic prominence). It is only natural to extend this approach to intrasegmental structure, especially in the context of segments that appear to consist of two parts (relatively unequal in terms of degree of stricture in the case of segments with secondary articulation, relatively equal in the case of segments with double articulation; see §2).

In dependency- and government-inspired approaches (see Anderson and Ewen 1987; Harris 1990; Smith 2000; and many others), secondary articulation consists of the adjunction of a dependent Place feature onto an otherwise well-formed consonantal tree structure. This brings out the "secondary" nature of the secondary articulation rather well: the head of the structure is the consonant, so that this articulation will be primary in the traditional sense. An example from Smith (2000: 258) is given in (11):

(11) *Secondary articulation* (Smith 2000)

labialized coronal

/t^w/



In this display, the C components, dominating the place element I (= coronal), form the head of the segment and the V components, dominating U (= labial), form the dependent, in this case a vowel-like superimposition of rounding. The part on the left is a prototypical consonant, i.e. a voiceless stop (a sonorant, a fricative, or a voiced segment would have one or more Vs in its structure), and the part on the right is a prototypical vowel (a glide would again be an intermediate category, with one or more Cs in its structure). Such representations of major class have also been explored in the context of lenition processes, for instance (see e.g. Anderson and Ewen 1987; also CHAPTER 66: LENITION).

The main difference between feature-geometrical models and dependency representations like that in (11) lies in the use of the dependency relation itself. Besides being independently motivated – as pointed out above, headship plays an important role in many realms of linguistics – there are four clear advantages of putting this relation to use, which the feature-based approach lacks: first, it is immediately clear which articulation is primary and which is secondary: since the C branch is the head, the segment as a whole is a consonant – no stipulations of interpretation are necessary. A second advantage of making use of the dependency relation is that constraints are expected to hold on the dependent, where fewer contrasts are expected to hold than in heads (Dresher and van der Hulst 1998) (cf. the lack of frequency of vowel contrasts in weak position of a foot compared to that in stressed positions (CHAPTER 40: THE FOOT), or the fewer tonal contrasts in the same context; Yip 1999). A possible natural condition on dependent branches, for instance, is that they cannot be segmentally complex: this would express the fact that phonetically simple types of secondary articulation such as labialization, palatalization, and velarization are common and that combined secondary articulation types such as labio-velarization are very rare (cf. above), or might be phonologically derived. Finally, the head–dependent relation is binary. This means that a primary articulation can only be combined with one secondary articulation in a segment, while in the feature-geometrical model there is a priori no reason why a single C-Place node could not be combined with two or more V-Place nodes, each potentially dominating multiple features.

On the other hand, it is unclear if, or to what extent, representations like (11) can be interpreted autosegmentally, i.e. whether the different levels on which Cs and Vs appear are different, independent autosegmental tiers. If they are not, this framework loses some predictive power compared to feature-geometry approaches, for instance in the analysis of the behavior of consonants with secondary articulation in vowel harmony (cf. Ewen 1995). It remains a goal to combine both approaches into a single overarching theory.

1.3 *Conclusion*

In this section we have discussed the representation of segments with secondary articulation. After a discussion of such segments in traditional feature theory, we have seen that feature geometry models or dependency models, or a combination of these, have shed further light on the phonological behavior of such segments, with respect to both their occurrence across languages, and their behavior diachronically and synchronically. With these insights, let us turn to segments that combine two articulations of equal stricture, double articulations.

2 Doubly articulated segments

While in segments with secondary articulation the vowel-like superimposed articulation is really "secondary," that is, of lesser structure than the primary articulation, there is another class of consonants which involve articulation types of equal structure. This is the topic of the present section. We will discuss their occurrence in the languages of the world and their representation in modern feature theories.

2.1 *Labial-velars and other double articulations*

Double articulations which involve articulation types of equal stricture could, in principle, involve two stops, two fricatives, two nasals, or two other types of articulations. In practice, however, producing two fricatives at different places of articulation is very difficult (CHAPTER 28: THE REPRESENTATION OF FRICATIVES), and this option is not systematically employed by any known language.[6] Only a stop articulation (whether oral or nasal) seems to be robust enough to allow for execution at two different places at the same time. As regards place of articulation, labial-velar stops are relatively common, and most of this section will be concerned with these segments. Other double articulations are labial-coronals, which have only been demonstrated beyond doubt for Yeletnye (Papua New Guinea). Examples from Ladefoged and Maddieson (1996: 344) are given in (12):

(12) *Labial-coronals and labial-velars in Yeletnye*

labial-alveolar		*labial-postalveolar*		*labial-velar*	
t͡pənə	'lung'	t̠͡pənə	'horn'	k͡pene	'coconut bag'
n͡mdboo	'pulp'	n̠͡md̠boo	'many'	ŋ͡mgbo	'fog'
n͡mo	'bird'	n̠͡mo	'we'	ŋ͡mo	'breast'

[6] Cf. Ladefoged and Maddieson (1996: 329–330), who also discuss the well-known case of Swedish /ɧ/, which has been suggested to contain two simultaneous fricatives, but is probably better analyzed as a segment with secondary articulation; cf. Lindblad (1980); Ladefoged and Maddieson (1996: 171f.).

The fact that labial-coronals and labial-velars (see below) have been reported raises the question of whether coronal-velars might also exist. Ladefoged and Maddieson (1996: 345) are not aware of any language with segments that would fit this description, so that in effect all double articulations involve the labial articulator.[7]

The most typical examples of doubly articulated stops are the labial-velar stops /k͡p/ and /g͡b/, which are especially common in west African and northern central African languages (Ladefoged and Maddieson 1996: 333), and are also relatively common at the eastern end of New Guinea. Note that labial-velars may contrast with clusters of labial followed by velar, as in the following examples from Eggon, a Benue-Congo language spoken in Nigeria (Ladefoged and Maddieson 1996: 334):

(13) *Labials, velars, labial-velars, and clusters in Eggon*

single segments		*clusters*	
pom	'pound (VB)'	kba	'dig'
abu	'dog'	g͡bga	'grind'
aku	'room'	akpki	'stomach'
gom	'break'	bga	'beat, kill'
k͡pu	'die'	kpu	'kneel'
g͡bu	'arrive'	gba	'divide'

There are clear phonetic differences between the clusters and the individual stops, and individual consonants in the clusters may be affected by a lenition process (Maddieson 1981; Ladefoged and Maddieson 1996: 334).

Nasals preceding labial-velar stops are also typically labial-velar, i.e. they usually assimilate to both places of articulation. Note that in a number of cases the nasal only appears to assimilate to the dorsal place of articulation in the labial-velar (never, apparently, to the labial articulation alone, according to Cahill 1995). This is the situation reported for Kpelle (Kreidler 2001) and Gã (Padgett 2004: 387 and references cited there): in the latter language, assimilation of nasals to labial-velars within a morpheme is total (resulting in labial-velar nasals), while across morpheme boundaries only the velar place of articulation assimilates. Padgett takes this as evidence against class nodes (cf. also Padgett 2002). Although the proposal to abolish class nodes would in itself form a welcome restriction on the representational limits of the theory, it could also be construed as evidence in favor of class nodes: in the former process the whole class node spreads, and in the latter only the feature (as argued by Halle 1995). We will not pursue this here.

An interesting case of a potentially labial-velar segment that is much more common is that of the labial-velar median approximant [w], which occurs in English and in many other languages, while a labial-palatal glide [ɥ] is reported in languages such as French and Mandarin Chinese (CHAPTER 15: GLIDES). Note that there are very few languages that contrast a non-velarized labial approximant [β̞] with

[7] One type of articulation that could be considered as coronal-velar are clicks, which involve a coronal (or labial) articulation combined with a velar closure. Since clicks involve a non-pulmonic airstream mechanism, they will not be considered here (see CHAPTER 18: THE REPRESENTATION OF CLICKS for discussion). Another possibility is that simultaneous coronal and dorsal articulation might give rise to an articulation at a place in between these two primary places, such as palatal, retroflex, or palato-alveolar. See e.g. Keating (1988) for discussion from a phonetic perspective.

[w], or a non-labialized velar glide [ɰ] with [w] (see Maddieson 1984 for a handful of such cases). In most cases, [w] can be analyzed as underlyingly just labial with velarization added as a phonetic enhancement effect. That labial is primary in English [w] is borne out, for instance, by the fact that coronal nasals can become labial before [w] (e.g. [m] in *sandwich*; see Wells 2008) and phonotactic restrictions against initial *[bw], *[pw], *[fw] (cf. Kenstowicz 1994: 541 for similar reasoning).

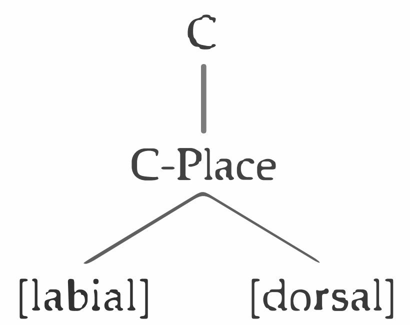
2.2 *The representation of segments with double articulation*

In this section we will briefly discuss a number of approaches to the representation of segments with double articulation, again starting with *SPE* and ending with dependency theories.

The representation of doubly articulated consonants is regarded as equivalent to secondary articulation in *SPE*, leaving the degree of rounding to phonetic implementation or perhaps contextual determining (*SPE*: 310). The implication is that there is no representational difference between consonants with a moderate degree of lip rounding, such as [k^w] and [g^w], and consonants with complete closure at the lips, such as [k͡p] and [g͡b]. This implication is clearly incorrect, however, since there are a number of languages that contrast labialized velars with labial-velars: such as Efik (Cook 1985) and Kpelle (Welmers 1962; Hyman 1973; Maddieson 1984). The inadequacy of the *SPE* proposal was also pointed out by Campbell (1974).

Thus the later feature-geometrical models and dependency-inspired proposals also improve on *SPE* in this respect, proposing that, instead of a vowel-like superimposition, two consonants of equal stricture are imposed on each other in cases like these. In Clements and Hume's model, this would entail representations like the following (cf. Clements and Hume 1995: 253):

(14) *Labial-velars*



The labial-palatals of Yeletnye would be represented with [coronal] and [dorsal] articulator nodes.[8]

Much as with secondary articulation, the question can be raised whether the two parts of the articulation have equal status in a doubly articulated segment. Is the representation in (14) primarily a labial or primarily a dorsal, or are both articulations truly on an equal footing? Recall that a nasal before a labial-velar sometimes assimilates to both parts, but sometimes only to the dorsal part,

[8] Note that if palatals are complex segments (cf. fn. 7), coronal-dorsals and palatals would have to have distinct representations.

Harris, John. 1990. Segmental complexity and phonological government. *Phonology* 7. 255–300.
Hayward, Katrina & Richard Hayward. 1989. "Guttural": Arguments for a new distinctive feature. *Transactions of the Philological Society* 87. 179–193.
Hinskens, Frans & Jeroen van de Weijer. 2003. Patterns of segmental modification in consonant inventories: A cross-linguistic study. *Linguistics* 41. 1041–1084.
Hulst, Harry van der & Jeroen van de Weijer. 1995. Vowel harmony. In Goldsmith (1995), 495–534.
Hume, Elizabeth. 1992. Front vowels, coronal consonants and their interaction in non-linear phonology. Ph.D. dissertation, Cornell University.
Hyman, Larry M. 1973. Notes on the history of Southwestern Mande. *Studies in African Linguistics* 4. 183–196.
Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge, MA: MIT Press.
Keating, Patricia. 1988. Palatals as complex segments: X-ray evidence. *UCLA Working Papers in Phonetics* 69. 77–91.
Kehrein, Wolfgang. 2002. *Phonological representation and phonetic phasing: Affricates and laryngeals.* Tübingen: Niemeyer.
Kenstowicz, Michael. 1994. *Phonology in generative grammar.* Cambridge, MA & Oxford: Blackwell.
Kreidler, Charles W. 2001. *Phonology: Critical concepts in linguistics.* London: Routledge.
Kühnert, Barbara & Francis Nolan. 1999. The origin of co-articulation. In W. J. Hardcastle & Nigel Hewlett (eds.) *Co-articulation: Theory, data and techniques*, 7–30. Cambridge: Cambridge University Press.
Ladefoged, Peter. 1968. *A phonetic study of West African languages: An auditory-instrumental survey.* 2nd edn. Cambridge: Cambridge University Press.
Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages.* Oxford & Malden, MA: Blackwell.
Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT.
Lindblad, Per. 1980. *Svenskans sje- och tje-ljud i ett allmänfonetisk perspektiv.* Lund: Gleerup.
Maddieson, Ian. 1981. Unusual consonant clusters and complex segments in Eggon. *Studies in African Linguistics* 8. 89–92.
Maddieson, Ian. 1984. *Patterns of sounds.* Cambridge: Cambridge University Press.
McCarthy, John J. 1991. Semitic gutturals and distinctive feature theory. In Bernard Comrie & Mushira Eid (eds.) *Perspectives on Arabic linguistics III*, 63–91. Amsterdam & Philadelphia: John Benjamins.
Ní Chiosáin, Máire. 1991. Topics in the phonology of Irish. Ph.D. dissertation, University of Massachusetts, Amherst.
Ní Chiosáin, Máire. 1994. Irish palatalisation and the representation of place features. *Phonology* 11. 89–106.
Ohala, John J. & Manjari Ohala. 1993. The phonetics of nasal phonology: Theorems and data. In Marie K. Huffman & Rena A. Krakow (eds.) *Nasals, nasalization, and the velum*, 225–249. San Diego: Academic Press.
Padgett, Jaye. 2002. Feature classes in phonology. *Language* 78. 81–110.
Padgett, Jaye. 2003. The emergence of contrastive palatalization in Russian. In D. Eric Holt (ed.) *Optimality Theory and language change*, 307–335. Dordrecht: Kluwer.
Padgett, Jaye. 2004. Partial class behavior and nasal place assimilation. In John J. McCarthy (ed.) *Optimality Theory in phonology: A reader*, 379–393. Malden, MA & Oxford: Blackwell. Originally published in Keiichiro Suzuki & Dirk Elzinga (eds.) 1995, *Proceedings of the 1995 Southwestern Workshop on Optimality Theory (SWOT)*, 145–183. Tucson: Department of Linguistics, University of Arizona.
Roberts, John R. 1987. *Amele.* London: Croom Helm.
Ryder, Mary Ellen. 1987. An autosegmental treatment of nasal assimilation and labio-velars. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23. 253–265.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT. Published 1990, New York: Garland.

Smith, Norval. 2000. Dependency theory meets OT: A proposal for a new approach to segmental structure. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: Phonology, syntax and acquisition*, 234–276. Oxford: Oxford University Press.

Sykes, Robert D. 2008. L2 perception of Arabic pharyngealization contrasts. Unpublished ms., University of Utah.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.

Watson, Janet C. E. 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Weijer, Jeroen van de. 1996. *Segmental structure and complex segments*. Tübingen: Niemeyer.

Wells, J. C. (ed.) 2008. *Longman pronunciation dictionary*. 3rd edn. Harlow: Pearson.

Welmers, William E. 1962. The phonology of Kpelle. *Journal of African Languages* 1. 69–93.

Williams, Edwin S. 1981. On the notions "lexically related" and "head of a word". *Linguistic Inquiry* 12. 245–274.

Yip, Moira. 1999. Feet, tonal reduction and speech rate at the word and phrase level in Chinese. In René Kager & Wim Zonneveld (eds.) *Phrasal phonology*, 171–194. Nijmegen: Nijmegen University Press.

30 The Representation of Rhotics

Richard Wiese

1 Introductory definition and overview

Rho is the name of the Greek letter <ρ>, corresponding to the Latin letter <r>. Rhotics, a loanword from Latin *rhotica*, are thus r-like sounds. The use of the classificatory term *rhotic* from the beginning of Latin grammar to present-day linguistics implies the existence of a class of such sounds. However, the fact that this class is named by referring to a particular letter of the alphabet is telling, as will be discussed below. Languages or dialects displaying some variant of the phoneme /r/ in contexts where other closely related languages or dialects do not show this sound (such as syllable-final position in Irish English in comparison to Standard British English) are also sometimes called rhotic languages or dialects, as opposed to non-rhotic ones.

This chapter will discuss rhotic sounds, their justification as a natural class in phonology, and related phenomena, first by introducing the class of rhotic sounds, largely by discussing their distribution in the languages of the world, and by discussing the justification for the term "rhotic", outlining cases in which rhotics clearly seem to work as a class in phonological patterning (§2), in spite of the high degree of variability. The chapter will then proceed to discuss possible definitions of the class of rhotics in articulatory or acoustic terms (§3). Such definitions will turn out to be rather unsatisfactory, and possible alternatives will be introduced in §4. Finally, the use of the related terms "rhotacism" and "liquid" will be exemplified briefly in §5. One important general context of the discussion of rhotics is the nature of the relationship between phonetics and phonology. This issue will be taken up in the final discussion.

2 The class of rhotics

There is an extensional definition of the class of rhotic sounds which simply enumerates them by identifying those sounds which are denoted in the IPA system of sound identification by some letter shape related to <r> or its capital counterpart <R>. This is actually the most common method of defining rhotics extensionally; see Ladefoged and Maddieson (1996: 216). In the most recent version of the IPA chart of phonetic symbols,[1] the symbols listed in (1) fulfill this criterion, and could conveniently be called "rhotics." The IPA tables, in addition, recognize a secondary quality of rhoticity for vowels, and give [ɚ], and the small hook, as for [ɑ˞ e˞], as examples.

(1) *Rhotics according to the use of r-like symbols in the classification of the IPA*

	Alveolar	Retroflex	Uvular
Trill	r		ʀ
Tap or flap	ɾ	ɽ	
Fricative			ʁ
Approximant	ɹ	ɻ	
Lateral flap	ɺ		

Empty cells in (1) are those for which the notational system of the IPA provides no straightforward, i.e. simple, symbol. However, most if not all sounds missing in the classification (of place and manner) present in this table actually exist, even if they are rare. For example, the uvular approximant rhotic, which is not recognized by the IPA system of notation as given in (1), is part of the inventory of Standard German and Danish. A possible notation for this sound is [ʁ̞], combining the symbol for the voiced fricative with the diacritic for lowering; another notation in common use is the haček (ˇ). Another rhotic mentioned occasionally is the retroflex trill [r̢] (Hall 1997: 105–106; Whitley 2003: 82), as in the Dravidian language Toda. Whitley in fact presents examples of rhotics for all of the empty cells in (1), except for the lateral flap row. Furthermore, he argues that alveolar, retroflex, and uvular voiced fricatives are rhotics when they are non-sibilant. The feature sibilant then allows for a distinction between sibilant fricatives [z ʒ ʐ] and non-sibilant (rhotic) fricatives [ř ɽ̌ ʁ̌].[2] The alveolar trills and taps/flaps may also be fronted, i.e. dental, symbolized [r̪ ɾ̪].

There is surprisingly little disagreement on calling the sounds symbolized in (1), as well as the additional types mentioned here, rhotics. According to this list of rhotic sounds and its concomitant classification, rhotics may be characterized by an alveolar, retroflex, or uvular place of articulation, and by one of five manners of articulation (trill, tap/flap, fricative, approximant, lateral flap). However, this classification in terms of articulatory categories does not lead to any uniform definition of the class of rhotics.

As a result, there is no principled way of excluding other potential sounds. For example, while the alveolar and uvular trills ([r ʀ]) are always included in this class, it is unclear why the bilabial trill [ʙ] is not. The same problem holds for the taps/flaps: if alveolar [ɾ] and retroflex [ɽ] are rhotics, why is the labio-dental flap not a rhotic? These sounds just mentioned are rare (just like some of the rhotics referred to above), but this cannot count as evidence. There is generally no obvious reason why the category of rhotics should be restricted to the coronal (alveolar/postalveolar) and dorsal (velar/uvular) places of articulation, to the exclusion of others. We also note that all sounds given in (1) are meant to be voiced, but

[1] The standard source for this chart is the *Handbook of the International Phonetic Association* (IPA 2007). Whitley's table 1 (2003: 82) has the same set of rhotics as those in (1).

[2] As there is no standard (i.e. IPA) symbol for rhotic fricatives, I will use the ones introduced here (recommended by Whitley 2003).

voicelessness may occur for rhotics and can of course always be symbolized by adding the diacritic [̥]. The one feature which all the sounds in (1) have in common is that they are consonants, but, as noted, rhotic/rhotacized vowels exist as well.

(1) therefore immediately gives rise to a number of questions, such as these: Is there anything special about these sounds, and is there any possible feature or feature combination shared by these sounds? Is this list of candidate rhotics complete? Why should only the particular combinations of place and manner in (1) be called rhotics? Is the use of sound symbols represented by a version of the letter <r> arbitrary or not? Do rhotics, under whatever definition, behave as a class of their own with respect to phonological patterning?

The reader must be warned that there are at best provisional answers to all of the questions raised here. We will attempt to offer such answers by examining the distribution of rhotics in the languages of the world (§2.1), discussing the unified phonological behavior of rhotics as a class (§2.2), and looking at rhotics with respect to their dialectal variation (§2.3). Featural descriptions that attempt to express the unity of rhotics are discussed in §3.

2.1 *Rhotics in the languages of the world*

A first approach to rhotics would be to investigate their frequency and distribution in the languages of the world. The work by Maddieson (1984) provides the basic information on the distribution of r-sounds in the languages of the world.[3] This database, containing 316 languages, reveals the figures shown in Table 30.1 on the distribution of rhotics.

According to this count, the majority of languages have rhotic phonemes. Languages for which no rhotic phoneme has been reported often show some rhotic either as an allophone of some other phoneme or in a peripheral area. Matumbi, for example, has no rhotic phoneme in the "core" vocabulary but allows /r/ in loanwords, according to Odden (2006: 199). In the typical case, a language has exactly one r-phoneme, and it is very rare for a language to have more than two rhotic phonemes. Well-known languages with two r-phonemes are Spanish (Harris 1969; Lipski 1990) and Catalan (Padgett 2003; Wheeler 2005: 24–34). In both languages, one r-sound is usually a flap, and the other a trill. In these languages, the two rhotics are in contrast in the position between vowels only, as illustrated in (2), while in all other positions the two r-phonemes are in complementary distribution.[4]

Table 30.1 Distribution of rhotic phonemes (Maddieson 1984: 83)

	Number of rhotic phonemes									
	0		1		2		3		4	
Languages in sample	74	23.3%	183	57.7%	51	16.1%	8	2.5%	1	0.3%

[3] Unfortunately, Haspelmath *et al.* (2005) contains no information on rhotic sounds.

[4] In classical generative phonology, the two underlying r-phonemes are often analyzed as a singleton /r/ and a geminate /rr/. Harris (1969: 50ff.) argues for this proposal from the phonotactic behavior of the two phonemes.

(2) *Two rhotic phonemes in partial contrast*

a. *Spanish* (Harris 1969: 46)

pero	['peɾo]	'but'	*perro*	['pero]	'dog'
torero	[to'ɾeɾo]	'torero'	*torrero*	[to'reɾo]	'lighthouse keeper'

b. *Catalan* (Padgett 2003: 2)

['miɾə]	'(s/he) looks at'	['mir:ə]	'myrrh'
['paɾə]	'father'	['par:ə]	'grapevine'

One of the languages reported to contain three rhotic phonemes is the Australian language Warlpiri. According to Jagst (1975) and Nash (1980: 66ff.), Warlpiri has two rhotic flaps, alveolar /ɾ/ and retroflex /ɽ/, and an approximant /ɹ/. Jagst (1975: 27) calls the two (alveolar and retroflex) flaps "vibrants," but in fact seems to interpret them as flaps/taps. He also analyzes further allophones, namely a trilled [r] for the flap /ɾ/, and a glide [ɻ̥] for the approximant /ɹ/. A similar situation, with three rhotic phonemes, can be diagnosed for other Australian languages, e.g. Maung, Arabana-Wanganura, and Kariera-Ngarluma, and a few other languages, such as Sedang (Austro-Asiatic).

Languages with more than three rhotic phonemes are exceedingly rare, if they exist at all. The one example in Maddieson's database is Irish Gaelic, for which a re-analysis in terms of two rhotic phonemes seems plausible (see Bammesberger 1982 and Ní Chiosáin 1994 for different proposals for such a re-analysis). In summary, there does not seem to exist clear evidence for a language with more than three rhotic phonemes (a point also made by Hall 1997: 109), and most of the languages with three such phonemes are located in Australia.

As a final point on the distribution of rhotic sounds, note that the alveolar trill [r] is the most common rhotic, taking up 47.5 percent of all rhotic phonemes, according to the database of Maddieson (1984: 79). In the more comprehensive *UCLA Phonological Segment Inventory Database*,[5] containing phonemic data from 451 languages, around 40 percent of all languages have a dental and/or alveolar voiced trill, and 20 percent of the languages display a voiced alveolar flap. Only four languages show a uvular trill, namely Batak, French, German, and Moghol.

The predominance of dental-alveolar trills remains something of a puzzle, given the effort and precise fine-tuning needed for the articulation of the alveolar trill. As noted by a number of phoneticians (e.g. Barry 1997; Schiller 1999; Catford 2001), trilled rhotic [r] requires a highly developed skill of articulatory control and execution. These authors tend to see other rhotics as underachieved or somewhat defective alveolar trills. The statistical dominance, and in fact prototypical status, of the alveolar trill among the rhotics is in striking contrast to the considerable difficulties it seems to raise for articulation.

Further facts on the distribution of rhotics within the overall set are conveniently summarized by Maddieson (1984: 82), on the basis of his database of 316 languages:

[5] Available (May 2010) at http://web.phonetik.uni-frankfurt.de/upsid.html.

(3) *Quantitative generalizations for rhotics* (Maddieson 1984: 82)

a.	An r-sound is likely to be voiced.	308/316	97.5%
b.	An r-sound is likely to be dental or alveolar.	273/316	86.4%
c.	An r-sound is likely to be interrupted.	244/282[6]	86.5%
d.	A retroflex r-sound is likely to be a continuant.	20/38	52.6%
e.	An approximant r-sound is likely to be retroflex.	15/28	53.6%
f.	A fricative r-sound is likely to be retroflex.	5/10	50.0%

Overall, these results give a good impression of the prototypical rhotic. However, the frequency of all other rhotics is sufficiently large to warrant their inclusion in the class of rhotics. None of the rhotics mentioned so far is unlikely to exist. The question of whether the class of rhotics can be restricted to the lingual sounds (from dental to uvular) tabulated in (1) across the different manners must remain open. Sociolinguistic studies of British English (Foulkes and Docherty 2000) have pointed to a widespread use of a labio-dental type of rhotic among younger speakers of English (especially in the South and East of England) characterized by the use of forms such as *red* [ʋɛːd]. Given that there is no definition for rhotics beyond the enumeration of its members, it cannot be ruled out that this constitutes a sound change leading from a rhotic to a non-rhotic segment.

2.2 *The unity of rhotics*

We saw above in (2) that in Spanish and Catalan there are two rhotic phonemes. The fact that the contrast between these two r-phonemes is suspended in all contexts other than intervocalic provides further evidence for the class: the neutralized phoneme is one or the other of these rhotics (see also CHAPTER 11: THE PHONEME; CHAPTER 2: CONTRAST). This is just one example of a general tendency: the common r-phoneme in a language is subject to allophonic variation, and the allophones are often from the class called rhotics.

Most of the evidence for rhotics constituting a distinct class in phonology comes from the areas of language variation and change. The status of rhotics as a phonological class is discussed by Lindau (1985), Hall (1997: chapter 4), and Walsh Dickey (1997: 90–101), who present the following types of evidence for the role of rhotics as a systematic class in phonological patterning:

(4) a. the influence of rhotics on neighboring vowels
b. the phonotactic unity of rhotics
c. their place in syllabic structure
d. the allophonic variation of rhotics
e. the phonemic alternations of rhotics.

(a) Rhotics, of all the types discussed above, tend to interact more closely with neighboring vowels than do other vowel-adjacent consonants (see CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS). English provides one relevant

[6] The overall number of r-sounds is smaller here, because the sources do not always report the manner of articulation (Maddieson 1984: 78). More generally, underspecified information on r-sounds may have led to a predominance of prototypical r-sounds in the database, i.e. of alveolar trills.

example: before (actual or historical) /r/, only a limited set of vowel phonemes may appear. In Standard British English, for example, only lax vowels /ɪ ɛ ɑ u ɔ/ appear in pre-r position (e.g. *here, care, car, sure, more*), while all other vowels are neutralized to /ɜ/ (which may reflect a merger with historical /r/), as in *bird, word, heard.* The general tendency of rhotics not to undergo palatalization, as discussed by Walsh Dickey (1997) and Hall (2000), provides another example of (negative) rhotic–vowel interaction.

(b) As discussed by Walsh Dickey (1997: 91–92), rhotics in Australian languages are often prohibited from occurring in word-initial position (see CHAPTER 86: MORPHEME STRUCTURE CONSTRAINTS). One such language is Mullukmulluk, in which the two rhotic phonemes /ɾ/ and /r/ can appear in any position except word-initially (Birk 1975: 61). Given that Australian languages typically allow for two or three rhotic phonemes, this constraint focuses on rhotics as a class, and not on a single phoneme which happens to be a rhotic. There are other more detailed constraints on the placement of rhotics: for Warlpiri, Nash (1980: 76) notes that in a CVC sequence (with heterosyllabic Cs), the two consonants cannot be identical rhotic phonemes (i.e. two tokens from one of the phonemes /ɾ ɽ ɹ/). This constraint seems to have exceptions, but it still captures a pattern significant in its exclusive reference to the class of rhotics.

(c) For languages that allow clusters of more than one consonant to appear in onset and/or coda position, rhotics are typically assigned to the position immediately adjacent to the vowel of the respective syllable. That is, a template of the type CrVrC describes the phonotactic placement of rhotics rather well, with C standing for one or more consonants other than /r/. (8) below exemplifies such a patterning from German, but many languages with complex syllable constituents behave analogously (CHAPTER 49: SONORITY).

(d) In many languages, there is a great deal of allophonic variation for the usual single r-phoneme, with the allophones standing either in free variation or in complementary distribution. But the large number of these allophones are all drawn from the inventory of rhotic sounds. The Persian language (Farsi) provides an instructive example: the phoneme /r/ has trilled [r] as its main allophone according to Majidi (1986: 63–64, 2000: 41–43), but has three to four additional rhotic allophones in complementary distribution, as shown in (5). The phoneme /r/ in Persian does not have non-rhotic allophones.

(5) *Distribution of rhotics in Persian* (Majidi 1986: 63–64)

a. *flap* [ɾ] *intervocalically*
   /tare/ ta'[ɾ]e 'chive'
b. *voiced fricative* [ɹ̆] *in word-initial position*
   /ruz/ '[ɹ̆]uz 'day'
c. *partially or completely devoiced trills* [r̥] *adjacent to voiceless consonants and word-finally*[7]
   /babr/ 'bab[r̥] 'tiger'
   /xɒrkan/ xɒ[r̥]'kan 'collector of blackthorn'
d. *voiced trill* [r] *elsewhere*
   /arzɒn/ a[r]'zɒn 'cheap'

[7] Majidi (1986: 64) sees a tendency to distinguish the devoiced trills in terms of either partial or complete lack of voicing.

These counts demonstrate that, from one generation of dialect speakers to the next, up to 60 percent of rhotics are changed in their phonetic realization. What remains is a single rhotic phoneme in the respective word.

(6) *Intergenerational changes*

a. *Berg* 'mountain'

change in type of /r/	194/327 ≃ 60%
no change	133/327 ≃ 40%

b. *Rose* 'rose'

change in type of /r/	93/296 ≈ 31%
no change	203/296 ≈ 69%

c. *fahren* 'go, drive'

change in type of /r/	106/313 ≈ 34%
no change	207/313 ≃ 66%

Word-initial or syllable-initial rhotics change in about a third of the tokens, while rhotics in the coda, as in the lemma *Berg*, showed a large number of r-realizations for both generations of speakers. Furthermore, the most r-realizations changed across the generations. There is no reason to think that this phenomenon is restricted to this particular dialect. An even more striking case of such a change in rhotic realization is reported by Enderlin (1911: 168). In his study of the Alemannic (Swiss German) dialect of Kesswil, he notes that at the local school all 1st grade students had uvular [ʀ], while all 9th grade students had alveolar [r]. He also reports a 50 percent realization of both forms for 4th and 5th grade students.

In general, postvocalic, rhymal rhotics seem to be subject to more variation than rhotics in onset positions. In Table 30.2, these changes are classified according to the type of change found in the same dialect map, that for *Berg* 'mountain' in Bellmann *et al.* (1999: 463). The r-sounds listed next to "older generation" are those found for the older speakers; while the first row lists rhotics found for the younger generation. The check marks denote intergenerational changes for particular pairs of rhotics (i.e. in some particular location as selected in the dialect atlas) through the comparison of the two maps.

Table 30.2 r-conversions found for one lemma (*Berg* 'mountain')

		Younger generation						
		ɾ	ʀ	ʀ̥	ɹ	ə	ɐ	Ø
Older generation	ɾ	✓	✓	✓	–	✓	–	✓
	ʀ	✓	✓	✓	–	✓	–	✓
	ʀ̥	–	–	–	–	–	–	–
	ɹ	–	–	–	✓	✓	–	–
	ə	–	✓	✓	–	✓	–	✓
	ɐ	–	–	–	–	✓	–	–
	Ø	✓	–	–	–	✓	–	✓

all the subclasses. Similarly, Kohler concludes that a positive characterization of the phoneme /r/ in German, encompassing all its allophonic variations, is not possible, even for a single speaker: "only a negative characterization is possible" (Kohler 1995: 156).

Ladefoged and Maddieson (1996: 245) summarize on an even more pessimistic note:

> Although there are several well-defined subsets of sounds (trills, flaps, etc.) that are included in the rhotic class, the overall unity of the group seems to rest mostly on the historical connections between these subgroups, and on the choice of the letter "r" to represent them all.

This conclusion would deny the existence of rhotics as a phonetically defined class, and is rather pessimistic on the possibility of providing any coherent descriptions in phonetic terms. It falls back on the conventions of alphabetic writing, while these are themselves obviously in need of explanation. Conceivably, the spelling of some (class of) sounds by means of the letter <r> exerts some influence on the paths of historical change of these sounds. But to assume that this spelling has a pervasive cross-linguistic influence and thereby constitutes the sole basis of the development of a class of rhotic sounds worldwide does not seem to be well founded. Other well-documented sound changes such as spirantizations, consonant losses, or vowel shifts do not seem to be restricted by the spelling systems. For example, the Second Germanic Consonant Shift, changing /p/, /t/, and /k/ to fricatives or affricates (Iverson and Salmons 2006), was not prevented by the fact that the spelling of affected words was changed.

Both Lindau (1985) and Ladefoged and Maddieson (1996), while noting the lack of a convincing segmental definition, emphasize the role of rhotics as a phonologically relevant class, along the lines discussed in §2 above. In order to express this unity, Hall (1997) proposes to use a feature [±rhotic] as a classificatory feature for the rhotic/non-rhotic distinction. However, a substantive definition of this abstract feature does not seem to be available. In yet another attempt, Walsh Dickey (1997) proposes to define rhotics by means of the feature Laminal, the use of the tongue blade as opposed to the tongue tip, as an articulator subordinated in the feature hierarchy to the articulator Coronal, as presented in (7).

(7) *Feature structure of rhotics* (Walsh Dickey 1997: 106)

```
[liquid]
   |
Coronal
   |
 Apical
   |
Laminal
```

However, it is questionable whether all rhotics make use of this feature structure. It is certainly not the case for uvular rhotics. Furthermore, it is unclear why such a marked segment class (expressed here by a deep hierarchical stacking of several place features, as opposed to an underspecified representation indicating an unmarked class; see CHAPTER 27: THE ORGANIZATION OF FEATURES; CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION) should be present in the majority of languages. The question of manner features for rhotics is even more pressing: if the degree of opening for rhotics may range from vocalic to fricative, as the survey given above seems to suggest, it is unclear how the class of rhotics may be characterized as a whole. Furthermore, this raises the puzzle how segments at the extreme ends of this dimension are to be classified: for example, when is a voiced uvular fricative [ʁ] a rhotic, and when is it not? The following section will propose an approach from a different angle. The failure to find a common denominator for rhotics in terms of acoustic or articulatory features does not preclude the possibility that it can be found eventually, but chances seem to be slight.

4 Alternative proposals

Featural descriptions proposed for rhotics, as discussed in the preceding section, have in common that they attempt to characterize rhotics in purely segmental terms. But we have already seen above that rhotics are tightly connected to their positions within larger phonotactic patterns, at least with respect to the conditions for their allophonic variants or for patterns of complementary distributions. This observation raises the question whether rhotics should in fact be described in terms of purely segmental categories, or in segmental terms at all.

An alternative view would be to capitalize on the observation that rhotics appear in a particular well-defined syllabic position, namely the one immediately adjacent to the vowel. This view, proposed by Wiese (2001), and previously by Selkirk (1984), relies primarily on the apparently uniform behavior of all types of rhotics in terms of their syllabic constraints, and suggests that this is in fact the defining and constant property of the class of rhotics. On this view, the search for constant segmental properties is futile, because it starts from an incorrect presupposition, i.e. that classificatory features are by necessity segmental features. In contrast, it seems that rhotics are very stable with respect to their phonotactic behavior. In particular, their slot in the structure of a syllable does not seem to change with a change in their segmental make-up. For example, Hall (1993) discusses a (lower Rhine) variety of German in which the rhotic phoneme, in coda position, varies between a vowel and a voiceless fricative (*Tor* [toɐ̯] 'gate' *vs.* *Sport* [ʃpɔχt] 'sport'). In other words, a fricative rhotic obeys the same constraints on syllabic placement as a trill or approximant rhotic, or even a rhotacized vowel. The proposal then is that rhotics are defined as a particular relative point on the sonority scale, the point between vowels and laterals.

Another pattern discussed in Wiese (2003) casts doubts on the segmental approach to rhotics as a class. (8) presents those onset clusters of present-day Standard German which consist of any stop followed by a sonorant (the velar nasal /ŋ/ is excluded from such clusters on principled grounds). Examples given in the cells of the table include rather marginal clusters (such as /pn/ or /tm/ occurring in a few rare words only), but in all of these rare cases, there is no tendency to replace the clusters in question by a more natural cluster or to break it up by a process of epenthesis.

(8) *Initial clusters of stop + sonorant in Standard German*

	/r/	/l/	/n/	/m/
/p/	*Preis*	*Platz*	*Pneu*	—
/b/	*braun*	*blau*	—	—
/t/	*Traum*	—	—	*Tmesis*
/d/	*drei*	—	—	—
/k/	*Kreis*	*klug*	*Knie*	*Khmer*
/g/	*grau*	*Glaube*	*Gnade*	*Gmünd*

Not all possible stop–sonorant clusters are attested. As shown in (8), /r/ can be combined with any of the existing stops, while all other sonorant consonants are restricted in some way or other with respect to the preceding stops. Basically, all homorganic clusters, such as /tl/ or /pm/, are ill-formed in (8), but there might be additional restrictions ruling out /bn/ and /dm/. For discussion of these, see Hall (1992: 65–80) and Wiese (2000: 261–269). But crucially, in contrast to the non-rhotic sonorants, no restrictions at all hold for stop–sonorant clusters involving /r/. As the examples in (8) demonstrate, this phoneme productively combines with any preceding stop, and none of the clusters involving /r/ is marginal. While it would be possible to simply make /r/ exempt from the ban on homorganic clusters, such a move would simply beg the question why this is the case. It seems more plausible to say that place features of /r/ simply do not count, or, in an underspecification approach to features, are not present (see CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION; CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION). The homorganicity ban in place for other sonorants cannot apply, then, leading to the complete set of stop–/r/ clusters shown in (8).

Similar patterns can also be demonstrated for other clusters in German (such as fricative–stop clusters), and for similar clusters in other languages (e.g. English, Italian, Basque, Lithuanian). This line of argumentation, arguing that rhotics are defined as those sounds which bear a sonority value between that of vowels (including glides) and the next lower sonority class, is supported by the fact that the freedom of rhotics to combine with a preceding stop is independent of the particular type of the r-phoneme present in the respective variety of German – place features as well as manner features of the rhotic phoneme are always irrelevant.

Finally, the proposal gives an answer to a puzzle noted above, namely that one and the same segment may sometimes be classified as a rhotic and sometimes not. This is particularly obvious in the case of voiced uvular fricatives: for Standard French, [ʁ] is generally seen as a rhotic; for Classical Arabic, what is apparently the same segment is not (Watson 2002: 13). But if rhotics are defined in terms of their phonotactic behavior, this is less mysterious: in French, the segment in question appears between obstruents and vowels, as in the initial cluster in *frais* [fʁɛ] 'fresh' or in the final cluster in *carte* [kaʁt] 'card', while in Classical Arabic the respective segment patterns with other fricatives in terms of phonotactics, even after the rhotic [r] as in *farġ* [farʁ] 'width'. In addition, Arabic has a tap or trill /r/ which is analyzed as a rhotic, because it behaves accordingly in terms of phonotactic patterning: e.g. San'ani Arabic *sirt* 'I/you (MASC SG) went'; *lribt* 'I fed' (Watson 2002: 67, 73). Dutch is another language which shows both a phoneme /r/ (with a great deal of variation) and a voiced velar or uvular fricative /ɣ~ʁ/ (see Booij 1995: 7–8).

be a considerable difference between these two levels. In some recent theories of phonology, emphasis has been placed on the seamless integration of phonetic and phonological representations. In the model of Articulatory Phonology, for example, all representations consist of articulatory gestures, i.e. movements in the vocal tract, which constitute the domain for phonological as well as phonetic representations (for a survey see Browman and Goldstein 1992; CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATION). In a treatment of English /r/ and its alternations with "zero," McMahon *et al.* (1994: 303) argue that /r/ consists of two articulatory gestures, a palatal constriction and a pharyngeal constriction. It remains to be seen to what extent this approach is adequate for other rhotics, in particular the prototypical trill and the uvular varieties. The discussion of rhotics in this chapter raises the question of whether a gestural – or any other – representation in terms of uniform and rather concrete units is possible for the characterization of rhotics.

REFERENCES

Bammesberger, Alfred. 1982. *Essentials of Modern Irish.* Heidelberg: Carl Winter Verlag.

Barry, William J. 1997. Another R-tickle. *Journal of the International Phonetic Association* 27. 34–45.

Bellmann, Günter, Joachim Herrgen & Jürgen Erich Schmidt. 1999. *Mittelrheinischer Sprachatlas*, vol. 4: *Konsonantismus.* Tübingen: Niemeyer.

Birk, David B. W. 1975. The phonology of MalakMalak. In Sharpe *et al.* (1975), 59–78.

Blevins, Juliette & Andrew Garrett. 1998. The origins of consonant–vowel metathesis. *Language* 74. 508–556.

Booij, Geert. 1995. *The phonology of Dutch.* Oxford: Clarendon Press.

Braune, Wilhelm & Ingo Reiffenstein. 2004. *Althochdeutsche Grammatik*, vol. 1: *Laut- und Formenlehre.* Tübingen: Niemeyer.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180.

Catford, J. C. 2001. On Rs, rhotacism and paleophony. *Journal of the International Phonetic Association* 31. 171–185.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Enderlin, Fritz. 1911. *Die Mundart von Kesswil im Oberthurgau: Mit einem Beitrage zur Frage des Sprachlebens.* Frauenfeld: Huber & Co.

Foulkes, Paul & Gerard J. Docherty. 2000. Another chapter in the story of /r/: "Labiodental" variants in British English. *Journal of Sociolinguistics* 4. 213–222.

Göschel, Joachim. 1971. Artikulation und Distribution der sogenannten Liquida r in den europäischen Sprachen. *Indogermanische Forschungen* 76. 84–126.

Hall, T. A. 1992. *Syllable structure and syllable-related processes in German.* Tübingen: Niemeyer.

Hall, T. A. 1993. The phonology of German /r/. *Phonology* 10. 83–105.

Hall, T. A. 1997. *The phonology of coronals.* Amsterdam & Philadelphia: John Benjamins.

Hall, T. A. 2000. Typological generalizations concerning secondary palatalization. *Lingua* 110. 1–25.

Harris, James W. 1969. *Spanish phonology.* Cambridge, MA: MIT Press.

Haspelmath, Martin, Matthew S. Dryer, David Gil, Bernard Comrie & Hans-Jörg Bibiko (eds.) 2005. *The world atlas of language structures.* Oxford: Oxford University Press.

Hulst, Harry van der & Nancy Ritter (eds.) 1999. *The syllable: Views and facts.* Berlin & New York: Mouton de Gruyter.

IPA. 2007. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet.* Cambridge: Cambridge University Press.

Iverson, Gregory K. & Joseph C. Salmons. 2006. Fundamental regularities in the Second Consonant Shift. *Journal of Germanic Linguistics* 18. 45–70.
Jagst, Lothar. 1975. Ngardilpa (Warlpiri) phonology (language of the Warnayaka tribe, a subtribe of the Warlpiri tribe). In Sharpe *et al.* (1975), 21–57.
Kohler, Klaus J. 1995. *Einführung in die Phonetik des Deutschen*. 2nd edn. Berlin: Schmidt.
Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford & Malden, MA: Blackwell.
Lindau, Mona. 1978. Vowel features. *Language* 54. 541–563.
Lindau, Mona. 1985. The story of /ɪ/. In Victoria A. Fromkin (ed.) *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 157–168. Orlando: Academic Press.
Lipski, John M. 1990. Spanish taps and trills: Phonological structure of an isolated opposition. *Folia Linguistica* 24. 153–174.
Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.
Majidi, Mohammad-Reza. 1986. *Strukturelle Grammatik des Neupersischen (Farsi)*, vol. 1: *Phonologie: Eine paradigmatisch-syntagmatische Darstellung*. Hamburg: Buske.
Majidi, Mohammad-Reza. 2000. *Laut- und Schriftsystem des Neupersischen*. Hamburg: Buske.
Marotta, Giovanna. 1999. The Latin syllable. In van der Hulst & Ritter (1999), 285–310.
McMahon, April, Paul Foulkes & Laura Tollfree. 1994. Gestural representation and Lexical Phonology. *Phonology* 11. 277–316.
Nash, David. 1980. Topics in Warlpiri grammar. Ph.D. dissertation, MIT.
Ní Chiosáin, Máire. 1994. Irish palatalisation and the representation of place features. *Phonology* 11. 89–106.
Ní Chiosáin, Máire. 1999. Syllables and phonotactics in Irish. In van der Hulst & Ritter (1999), 551–575.
Odden, David. 2006. Kimatuumbi: Phonology. In Keith Brown (ed.) *Encyclopedia of language and linguistics*. 2nd edn, vol. 6, 198–206. Oxford: Elsevier.
Ortmann, Albert. 1998. Consonant epenthesis: Its distribution and phonological specification. In Wolfgang Kehrein & Richard Wiese (eds.) *Phonology and morphology of the Germanic languages*, 51–76. Tübingen: Niemeyer.
Padgett, Jaye. 2003. Systemic contrast and Catalan rhotics. Unpublished ms., University of California, Santa Cruz.
Rice, Keren. 2005. Liquid relationships. *Toronto Working Papers in Linguistics* 24. 31–44.
Schiller, Niels O. 1999. The phonetic variation of German /r/. In Matthias Butt & Nanna Fuhrhop (eds.) *Variation und Stabilität in der Wortstruktur: Untersuchungen zu Entwicklung, Erwerb und Varietäten des Deutschen und anderer Sprachen*, 261–287. Hildesheim: Olms Verlag.
Scobbie, James M. 2006. (R) as a variable. In Keith Brown (ed.) *Encyclopedia of language and linguistics*. 2nd edn, vol. 10, 337–344. Oxford: Elsevier.
Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*, 107–136. Cambridge, MA: MIT Press.
Sharpe, Margaret C., Lothar Jagst & David B. W. Birk (eds.) 1975. *Papers in Australian linguistics* 8. Canberra: Australian National University.
Uffmann, Christian. 2007. Intrusive [r] and optimal epenthetic consonants. *Language Sciences* 29. 451–476.
Ulbrich, Christiane & Horst Ulbrich. 2007. The realisation of /r/ in Swiss German and Austrian German. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 1761–1764. Saarbrücken: Saarland University.
Van de Velde, Hans & Roeland van Hout (eds.) 2001. *'r-atics: Sociolinguistic, phonetic and phonological characteristics of /r/*. Brussels: Université Libre de Bruxelles.

Walsh Dickey, Laura. 1997. The phonology of liquids. Ph.D. dissertation, University of Massachusetts, Amherst.

Watson, Janet C. E. 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Wheeler, Max W. 2005. *The phonology of Catalan*. Oxford: Oxford University Press.

Whitley, M. Stanley. 2003. Rhotic representation: Problems and proposals. *Journal of the International Phonetic Association* 33. 81–86.

Wiese, Richard. 2000. *The phonology of German*. Oxford: Oxford University Press.

Wiese, Richard. 2001. The phonology of /r/. In T. A. Hall (ed.) *Distinctive feature theory*, 335–368. Berlin & New York: Mouton de Gruyter.

Wiese, Richard. 2003. The unity and variation of (German) r. *Zeitschrift für Dialektologie und Linguistik* 70. 25–43.

31 Lateral Consonants

Moira Yip

Laterals are extremely common, and yet they are something of a phonological puzzle. For most consonants, there is fairly general agreement on which natural classes they belong to, and clear expectations of how they pattern in phonological processes. So for example [m] is Labial and [+nasal], and as such it patterns consistently with the other labials [p b], and the other nasals [n ŋ]. But for laterals this is not the case, as we shall see. Their behavior is highly variable across languages. This chapter gives an overview of their somewhat perplexing behavior, summarizes extant proposals for their phonological representation, and ends by advocating a proposal using violable constraints that allows for this variability.

I begin in §1 with some background on their articulation and acoustics, and how they are acquired by the child. §2 summarizes the types of laterals found in natural language. §3 surveys their roles in syllable structure. §4 discusses their sonorancy, voicing, and continuancy characteristics. §5 looks at alternations between laterals and other sounds. §6 asks whether there is a feature [lateral]. §7 looks at the positioning of [lateral] in a theory of feature geometry. I show that there are serious problems associated with any single choice of superordinate node, and instead in §8 I propose an approach based on violable feature co-occurrence constraints. §9 concludes.

1 Phonetics, perception, and acquisition

1.1 *Types of laterals and their frequency in the world's languages*

Over 80 percent of languages have one or more lateral consonants (Maddieson 1984). Laterals are defined by Ladefoged and Maddieson (1996: 183) as "sounds in which the tongue is constricted in such a way so as to narrow its profile from side to side so that a greater volume of air flows around one or both sides than over the center of the tongue." The palatograms in Figure 31.1 show clearly that the tip of the tongue makes contact with the roof of the mouth, but the sides do not.

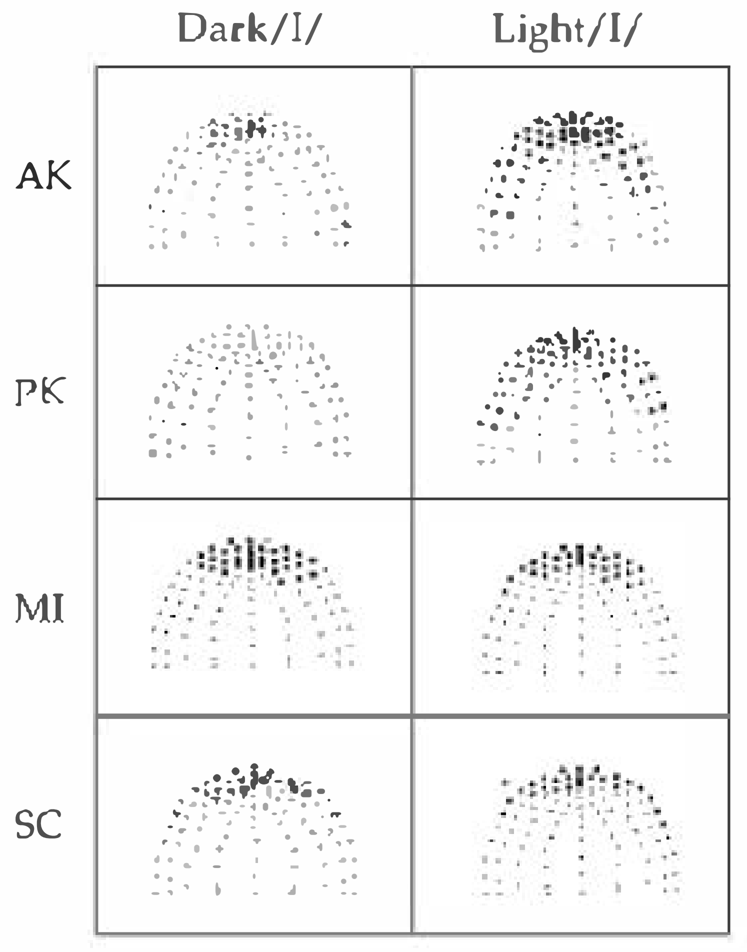


Figure 31.1 Linguo-palatal contact profiles of dark and light laterals for four speakers of American English, showing mid-sagittal contact in the alveolar and pre-palatal regions but no or limited contact at the sides. From Narayanan *et al.* (1997)

The approximant versions have some turbulence at the (incomplete) stricture; resonant versions like Standard German coda /l/ do not. Since the vocal tract is not fully obstructed in the typical approximant /l/, they are among the most sonorous of consonants (see CHAPTER 49: SONORITY). They frequently contrast with other sonorants, especially rhotics, and the class of laterals and rhotics is referred to as the liquids (see CHAPTER 8: SONORANTS). In other languages (such as Japanese) a language has only a single liquid, and it may vary between a lateral [l] and a more rhotic tap or flap [ɾ]. The variation may depend on context (such as syllable position) or they may be in free variation. Speakers of such languages famously have trouble perceiving and producing the difference between [l] and [r] when learning languages like English (see Iverson *et al.* 2003). There are also lateral obstruents, which will be discussed below.

1.2 *Articulation, acoustics, and acquisition of laterals*

Most laterals are dental or alveolar in articulation, but the tongue body is also frequently implicated, as shown for English by Sproat and Fujimura (1993). Gick *et al.* (2006) studied the articulation of laterals in six languages: Western Canadian English, Quebec French, Serbo-Croatian, Korean, Beijing Mandarin, and Squamish Salish. They found that all their laterals had an anterior tongue gesture in all syllable positions. More interestingly, in coda position all also have a dorsal gesture, which starts slightly earlier than the anterior gesture. Some, but not all, have a dorsal gesture in onset position too, in which case it is roughly simultaneous with the anterior gesture. The dorsal gesture in coda position may result from biomechanical causes, such as "active lateral compression of the

4.1 Sonorancy and voicing

There is general agreement that the vast majority of laterals are sonorants. But *how* sonorant are they? Based on widely accepted phonological evidence, Parker (2008) posits the following sonority hierarchy for sonorant consonants (see also CHAPTER 49: SONORITY):

(7) *Relative sonority of sonorant consonants*

glides > flaps > laterals > trills > nasals

Using data from Quechua (Peruvian) Spanish and US English, he then sets out to show that this ranking has a measurable phonetic correlate, namely acoustic intensity. Although he finds a very good statistical match for his overall notion, [l] in onset position frequently comes out as louder (and thus more sonorous) than the glides, contra expectations. Parker suggests that this is because the glides may become more "obstruent-like" in onset position, especially in Spanish. According to Harris and Kaisse (1999), this is rather pervasive. In Argentinian Spanish it is carried to an extreme, as shown by alternations such as *le*[j] 'law' *vs. le.*[ʒ]*es* 'laws'. But even in Castilian, the onset glide [j] frequently hardens to the non-strident palatal fricative [ʝ] or even to its non-continuant counterpart.

Once we look outside the prototypical approximant laterals, we quickly encounter laterals that are not sonorant at all. Examples of obstruent laterals include not only the obvious fricatives, affricates, and clicks, but also languages in which [l] patterns with the voiced obstruents, such as Southern Min, which has [l] instead of [d]. Modern Southern Min completely lacks [d]. Not only is [l] the modern reflex of historical *d, but underlying /p t k/ voice to [b l g] foot-internally (Hsu 1996), and /b l g/ nasalize to [m n ŋ] before nasal vowels. In some Bantu languages, like Ikalanga, historical *d has become /l/, and synchronically under velarization /l/ becomes the stop [gw], suggesting that it may still be an obstruent. (A reviewer correctly notes that in both these cases there could be intermediate steps involved, and /l/ may not be underlyingly obstruent.) There are even a few languages in which *all* the laterals appear to be obstruents: Tlingit for example has two lateral fricatives and three lateral affricates, but no lateral approximants.

Finally, turning to voicing, like other sonorants laterals are typically voiced, but voiceless versions are found, for example in Toda. The theoretical implications of the fact that most (but not all) laterals are voiced sonorants will be discussed in §7.

4.2 Continuants or not?

There has been little agreement as to whether ordinary approximant laterals behave like stops or like continuants, and thus whether they are [±continuant]. Alternations with stops are common, but alternations with fricatives do not seem to be found. Nonetheless, Holt (2002) argues that they are both [+continuant] and [−continuant] and Mielke's (2005) survey, as his title "Ambivalence and ambiguity in laterals and nasals" makes clear, finds an almost even 50/50 patterning as [±continuant]. He attributes this to their ambiguous phonetic cues, and suggests that it is precisely phonetically ambiguous segments that are likely to behave ambivalently in the phonology. I shall return to this issue in §9.

Mielke gives these examples of each pattern. In Basque, laterals (but not rhotics) pattern with nasals in assimilating in place to a following consonant. The class of segments that undergo the rule is [+sonorant, −continuant]. Mielke points out that there is no way to define this class other than by using [−continuant] to include /l/ and the nasals, but not /r/.

(8) /l/ *patterning with non-continuants in Basque* (Hualde 1991: 96)

egu[m] *berri*	'new day'
egu[n̪] *denak*	'every day'
egu[ɲ] *ttiki*	'small day'
gu[ŋ] *gorri*	'red day'
ata[l̪] *denak*	'every section'
ata[ʎ] *ttiki*	'small section'

(I will suggest an alternative explanation for the Basque case in §8.3.)

Conversely, in Finnish stems may end in the five coronals /t s n r l/. Before /n/-initial suffixes, [−continuant] /t/ *undergoes* total assimilation, but [+continuant] /s r l/ *trigger* total assimilation.

(9) /l/ *patterning with continuants in Finnish*
(Sulkala and Karjalainen 1992: 387–388)

		active potential	*2nd part active*	
a.	/avat/	[avannut]	[avannee]	'open'
b.	/pur/	[purrut]	[purree]	'bite'
	/nous/	[noussut]	[noussee]	'rise'
	/tul/	[tullut]	[tullee]	'come'

Mielke quotes Kaisse (2000) on the issue. She points out that the status of laterals hinges on whether [−continuant] is defined in terms of complete occlusion in the oral tract ("vowel tract" in *SPE*; Chomsky and Halle 1968: 318) or complete occlusion in the mid-sagittal region of the oral tract (see also CHAPTER 13: THE STRICTURE FEATURES). Laterals have complete occlusion only in the mid-sagittal region, not elsewhere, so they qualify as [−continuant] only under the latter definition. Mielke suggests that this makes laterals phonetically ambiguous, and that the boundary for the natural classes of [+continuant] and [−continuant] may vary cross-linguistically, placing laterals in different classes in different languages. The implication is that feature values may not be universally fixed for such segments, but may "emerge" on the basis of observable phonetic properties. One might also note that other rationales for cross-linguistic differences in feature specifications have been advanced, particularly the system of contrasts in the language in question. See Morén (2006) for an account on these lines for Serbian laterals.

5 Alternations involving laterals

When laterals alternate with other segments, it is usually with ones that are minimally different either articulatorily, acoustically, or both. Since most laterals are coronals, a small change in the type of closure so that it is complete will produce a coronal stop. If the lateral finds itself subject to nasal spreading, so that the velum is lowered, a nasal sonorant is the likely result. If the tongue shape is inverted, so that the closure is made with the sides but not the midline of the tongue, then a rhotic is formed. And if the closure is eliminated, but the tongue body remains high, a high vocoid results. Some of these changes also produce a sound that is acoustically still quite similar to the lateral, meaning that the change may have originally been driven in whole or in part by a misperception of the signal. All of these are common changes, and are discussed below.

5.1 *Stops*

In some languages, /l/ alternates with stops; more specifically, it appears to replace the voiced coronal stop /d/ in some contexts. In Palenquero Spanish (Piñeros 2003), /d/ is in free variation with [l] in some words:

(10) /d̪ed̪o/ → ['le.lo] ~ ['d̪e.ðo] 'finger'

In some Bantu languages, like Ikalanga, historical *d has become /l/, but under velarization /l/ becomes the stop [gʷ], suggesting that it may still be an obstruent. Conversely, in Southern Min (which has /l/ instead of /d/ in its phoneme inventory), if syllable-final /t/ ends up intervocalically (especially before an unstressed vowel), it voices and becomes not [d] but [l] (although descriptions vary, and as it is a brief flap or tap, it is not entirely clear how lateral its articulation is; see Hsu 1996).

5.2 *Nasals*

Historically, Cantonese had two distinct phonemes /l/ and /n/. Both could occur syllable-initially, as in the contrast between [lei] 'reason' and [nei] 'you'. Only /n/ could occur syllable-finally. However, in the last 50 years a gradual sound change has been taking place, and is now nearly complete for younger speakers (Bauer and Benedict 1997). Initial /n/ is being replaced by /l/, so that 'you' and 'reason' are now both pronounced as [lei]. As a result, [l] and [n] can be treated as allophones of a single phoneme, with [l] as the syllable-initial variant and [n] as the syllable-final one. However, actual alternations do not exist, because the language has essentially no resyllabification.

In the Min dialects of Chinese (including Southern Min, mentioned earlier), voiced stops and nasals are in complementary distribution, with [b g] occurring only before oral vowels and [m ŋ] only before nasal vowels. These data are from the Chaoyang dialect:

(11) bi:ʔ^{11} 'hide' mẽ:^{53} 'fast'

The language has no alveolar voiced [d], but the reflex of historical *d is [l] before oral vowels and [n] before nasal vowels. As a result [l] and [n] are in complementary distribution, and this is productive. In Chaoyang onomatopoeia, words are reduplicated and one onset is replaced by [l]. However, if the vowel is nasal it is replaced instead by [n]. See Yip (2001) for details.

leaving only the more vocalic Dorsal gesture (see CHAPTER 75: CONSONANT–VOWEL PLACE FEATURE INTERACTIONS).

A rather different but equally well-known case comes from the Cibaeño dialect of Spanish, in which both /l/and /r/ become the palatal glide [j] in coda position (Guitart 1985; Harris 1985; Alba 1988).

(15) *celda* [sejda] 'cell'
cerda [sejda] 'bristle'

Since the gesture that is preserved here is the Coronal one, Johnson and Britain's approach cannot deal with these facts unmodified. However, there is evidence that, unlike English coda /l/, Spanish coda [l] is clear, not dark, with a smaller Dorsal gesture. This might then explain why /l/ vocalizes as [j], not [w], although the picture is less clear if one studies a range of Spanish and Portuguese dialects. See Quilis *et al.* (1979) and Recasens and Espinosa (2005) for details.

Many dialects of Spanish, especially in Latin America, have replaced the palatal lateral [ʎ] by the glide [j]. This sound change is known as *yeismo*. (In Buenos Aires Porteño Spanish, this change has gone one step further in a change called *zheismo*, with [j] spirantizing to [ʒ]. See Harris and Kaisse (1999) for details.)

(16) *llorar* [ʎ]*orar* [j]*orar* 'to cry'
ella *e*[ʎ]*a* *e*[j]*a* 'she'

In Serbian, rather unusually, /l/ vocalizes to [ɔ]. See Morén (2006) for an interesting account within a Parallel Structures Model. Historically, vocalization of laterals is also common. In Germanic, compare English *old* to Dutch *oud*. In Polish, dark [ɫ] changed to [w] everywhere, even in onsets. In Romance, Latin *l has developed variously to [w/u], e.g. in French (*caldus* to *chaud*), or to [j/i] (in both onset and coda), as found in Italian *flos* to *fiore* and Portuguese *multus* to *muito*.

The inverse of vocalization of laterals, the lateralization of glides, seems to be much rarer (CHAPTER 15: GLIDES). Li (1974) documents a case in which *j > [l] in some Formosan languages, and a voiced coronal fricative [z ð] in others. Interestingly, in no case does *w > [l], even though in *l*-vocalization /l/ becomes [w] more often than [j]. This suggests that there must be a coronal gesture already present (as there is for [j]) for the creation of a novel lateral.

6 The feature [lateral]: Is it necessary?

Phonologists routinely use a feature [lateral] to distinguish /l/ from /r/, but some linguists (Spencer 1984, Brown 1995, and most recently Walsh Dickey 1997) have argued that it can be dispensed with. If a language has [l] but no [r], one might define [l] by the features [+consonant, +sonorant, −nasal], and [lateral] would be redundant. However, if [l] contrasts with [r], as it does in many languages, this will not suffice. They could perhaps be distinguished by the feature [continuant], with /r/ as [+continuant] and /l/ as [−continuant], but this is not without problems (see §4.2, and also van der Weijer 1995). Walsh Dickey (1997: 55) distinguishes them by means of Place features, with /l/ having complex Corono-Dorsal place and /r/ non-primary Laminal. Dissimilation of /l/ to [r] is loss of a secondary

Dorsal articulation. The lateral articulation of [l] is, for her, "a necessary phonetic consequence of a phonological Corono-Dorsal complex place structure." It is worth having a slightly closer look at her arguments and her proposal.

Walsh Dickey points out, correctly, that the strongest arguments for a feature come from its role in defining natural classes. The many languages with only one lateral can of course never offer this type of evidence. She lists three types of potentially significant evidence from languages with more than one lateral: (i) co-occurrence restrictions on different types of laterals; (ii) positional restrictions which cover all types of laterals in a language; and (iii) phonological processes which need to refer to all laterals, both sonorant and obstruent. She concludes that no such evidence exists, and I have also not encountered any convincing cases.

However, she pays a high price for the absence of [lateral]. In particular, she has to greatly complicate the internal structure of the coronal node by including Laminal (which in turn may or may not be [dental]), Apical (which in turn may or may not be [back]), and secondary Dorsal (which may have yet another Dorsal specification below it, to account for velarized "dark" laterals). Finally, any other sound which might have been thought to involve secondary Dorsal articulations, such as velarized coronal consonants, would require some extra specification if they did *not* have a lateral release. I therefore tentatively conclude that the feature [lateral] is still useful, and probably necessary.

Positive evidence for the feature [lateral] comes from its active role in the phonology of many languages, despite Walsh Dickey's claims to the contrary. In Eastern Catalan (and Sanskrit), for example, [lateral] spreads onto nasals to create a lateral nasal: /nl/ → [l̃l] in /son les tres/ → [sol̃les tres] (Mascaró 1976). There are well-known phonological processes that involve only [l] and [r], and in which they either dissimilate, as in Latin, where the suffix /-alis/ surfaces as [-aris] after a lateral root: *nav-alis vs. milit-aris* (Steriade 1987), or assimilate, as in Sundanese, where the infix /-ar-/ surfaces as [-al] after a preceding /l/: [k-ar-usut] *vs.* [l-al-əga] (see Cohn 1992 for details). Several of these processes are long distance, and can cross over other Coronals, making a Place feature account tricky. I conclude that the feature [lateral] cannot be dispensed with.

I should note that for the remainder of this chapter I shall treat [lateral] as a privative feature, but the results would not be materially affected if it were to turn out to be binary, as Steriade (1987) argues.

7 Feature geometry and the feature [lateral]: Two competing models

It has been proposed that distinctive features are related to each other by a hierarchical feature geometry (Sagey 1986, Clements and Hume 1995, and many others; see also CHAPTER 27: THE ORGANIZATION OF FEATURES). For example, the features Labial, Coronal, and Dorsal are dominated by a Place node. In such a model, we must then ask where the feature [lateral] is located. Early proposals spent little time worrying about the placement of [lateral], and tended to put it directly under the root node. Clements and Hume (1995: 293) opt for this, but admit that its position is open to dispute. Subsequently, two competing detailed proposals for the placement of lateral have been put forward, and are shown below: it might

Similar facts hold in Tamil (Beckman 1998), and in English: *we*[l̪θ] *wealth*, but *whe*[lk] *whelk*, although the English case may be purely phonetic, since it is not structure-preserving. The interesting fact is that in Basque and all these other languages laterals do not lose their laterality as the targets of Place assimilation, contra the predictions of the Coronal model.

Finally, there are languages in which Place contrasts are lost – a phenomenon described by Trigo (1988) and others as the delinking of the Place node – but laterality survives. In Caribbean Spanish (Trigo 1988: 71) place features are neutralized in codas: /d/ deletes, /s/ becomes [h], and all nasals become velar; /r/ and /l/ are unchanged.

(24)	a.	/βerdad/	→	[βerða]	'truth'	
	b.	/ines/	→	[ineh]	'Ines'	
	c.	/album/	→	[albuŋ]	'album'	(*optional*)
		/tren/	→	[treŋ]	'train'	
		/desdeɲ/	→	[desdeŋ]	'disdain'	
	d.	/tonel/	→	[tonel]	'barrel'	
		/par/	→	[par]	'pair'	

The data in this section are problematic for the Coronal model.

7.2 *The sonorant voicing (SV) model*

What about the sonorant voicing (SV) theory? Rice and Avery (1991) base their claims on three main arguments. Firstly, the SV node is the target node when lateral spreads to other sonorants. Secondly, a process of de-sonorantization can be seen as delinking of the SV node. Thirdly, rules which spread both nasality and laterality can be viewed as spreading the SV node (see Table 31.2).

Table 31.2 Predictions of the SV model

Prediction	*Under sonorant voicing?*	
	FOR	*AGAINST*
All laterals should be voiced sonorants	Laterals are usually voiced sonorants: *many languages*	Voiceless laterals: *Tahltan* Obstruent laterals: *Min, Bantu* Affricate laterals *Tahltan, Zulu*
SV spreading should spread [lateral]	Yes: *Sanskrit*	No: *Polish*
When [lateral] spreads, it should seek SV targets	Yes: *Toba Batak*	Laterals skipped by harmony that targets sonorants: no cases known
SV spreading onto laterals should remove [lateral]	Yes: *Itsekiri*	No: *English*
Loss of SV removes [lateral]	Yes: *Yagaria*	No: *Koyukon, Angas*

7.2.1 *Supporting evidence for the SV model*

Apart from the observation that most laterals are voiced sonorants, there are four other types of evidence in support of the proposal.

Firstly, SV spreading sometimes spreads [lateral], as in Sanskrit. Before laterals, all obstruents voice, and if they are coronal they also lateralize (Whitney 1889: 54; no examples of non-coronals are given):

(25) tat labhate → tal labhate
trin lokan → tril̃ lokan

Secondly, [lateral] spreading sometimes seeks SV targets to attach to, as in Toba Batak (Hayes 1986). Coronal sonorants assimilate to a following liquid:

(26)	*Spreading*	*No change*
	nr → rr	ln rn lr
	nl → ll	mr ml ŋl ŋr
	rl → ll	rr ll nn

If what spreads is the SV node, and if [lateral] is its dependent, then laterality will be carried along too.

Thirdly, when SV spreads onto a lateral target, replacing the original SV specification, the original lateral specification may be lost. The following facts from Itsekiri (Nigeria; Piggott 1991, cited in Brown 1995: 64) are often cited, and very similar facts hold in Southern Min Chinese and in Yoruba.

(27) lā → nā 'ask the price of'

This type of nasal harmony can be seen as spreading the superordinate SV node, carrying nasality with it, and removing the original SV node, laterality and all.

Fourthly, if an SV node delinks, lateral may be removed too, as in Yagaria. Here a coalescence process removes voicing and converts a sonorant [ʟ] to an obstruent [t], and in the process laterality also goes.

(28) gipaʔ-ʟoʔ → gipatoʔ 'at the door'

For further data on Yagaria, see §2, where these data are discussed in a different context.

7.2.2 *Counterevidence to the SV model*

There are various types of counterevidence to the SV model, as listed in the right-hand column of Table 31.2. The first problem is the existence of laterals that are not voiced sonorants, being either voiceless, obstruent or both. But there are other problems.

Firstly, SV spreading does not always spread [lateral]. The following data show a post-lexical process of voicing assimilation in Krakow and Posnan Polish (Dorota Glowacka, personal communication; Madelska and Witaszek-Samborska 1998):

(29) *brat* → *bra*[d] *Doroty/Natalji/Iwony/Luizy* 'brother of X'
syn → *sy*[n] *Luizy* 'son of Luiza'

Similar facts hold in all dialects between verbal prefixes and roots:

(30) s-kɔɲʧɨtɕ 'to end' z-bitɕ 'to break' z-lit̪ɨtɕ 'to count'

In the SV model, voicing in sonorants is represented by the presence of an SV node, and so this must be the active node that spreads. If [lateral] is under this node, it too should spread, but it does not, contra the predictions of the SV model.

Secondly, SV spreading into lateral targets sometimes leaves the laterality intact. This is the case in English, where liquids after voiceless aspirated stops become voiceless:

(31) [bl]*eak* [pl̥]*ease* [gl]*eam* [kl̥]*ean*
[br]*eam* [pr̥]*een* [gr]*een* [kr̥]*eam*

If sonorant voicing is denoted by the presence of an SV node, the devoicing would presumably mean that the SV node had been delinked, and one would then expect loss of [lateral] as well, but no such thing happens.

Thirdly, if the SV node of a lateral delinks for other reasons, rendering it voiceless, laterality should also be lost, but this is not the case in the Athapaskan language Koyukon (Rice 1994), which devoices syllable-final sonorants and continuants, including /l/. For the lateral, the result is a voiceless lateral fricative [ɬ]. Final stops are plain voiceless unaspirated.

(32) [nəɣælə] 'your (sc) trap' [xæɬ] 'trap'
[səʔɔɣə'] 'my snowshoes' [ʔɔx] 'snowshoes'
[nizuni] 'that which is good' [nizun̥] 'it is good'

Under the SV hypothesis, where [lateral] is under SV and devoicing of sonorants means removal of the SV node, laterality should also disappear, but it does not, contra the predictions of the SV model.

In the case of [lateral], then, the need created by theories of universal feature geometry to commit to a single location for the feature in all languages creates problems. Luckily, there are alternatives to these two proposals. Hegarty (1989) and Bao (1992) argue that [lateral] is simply a dependent of the Root node. Yip (2004: 5) goes further, and agrees with Padgett (1995, 2002) that (at least with respect to the behavior of [lateral]) features can be treated as an unstructured set of which [lateral] is a member, and that feature geometry as such is redundant. The next section lays out this proposal.

8 A feature co-occurrence constraint approach

8.1 Inventories

Suppose we capture the fact that the least marked lateral is a coronal sonorant by means of two universally fixed hierarchies of feature co-occurrence markedness constraints.

(33) a. *LateralObstruent >> *LateralSonorant
b. *LateralLabial >> *LateralDorsal >> *LateralCoronal

The idea is that these constraints are violated if a segment bears both features, at least if both are primary articulations. For example, the intention is that *LateralLabial bans bilabial or labio-dental segments with a lateral release. If the labiality is secondary, as in [l^w], it may be acceptable. Both the fixed rankings are phonetically grounded. Lateral release means the airflow is not easily obstructed, so lateral obstruents are more marked than lateral sonorants (33a), and lateral release is easiest if only the tip of the tongue is used to make closure, so coronals are the least marked laterals (33b).

Placing the faithfulness constraints at different points in these hierarchies gives us a typology of segmental inventories like those below:

(34) *Typology of lateral place of articulation*

a.	*LatLab >> *LatDors >> *LatCor >> Faith	Either no laterals (Maori), or placeless ones (Cambodian)
b.	*LatLab >> *LatDors >> Faith >> *LatCor	Common type, with Coronal laterals only (English)
c.	*LatLab >> Faith >> *LatDors >> *LatCor	New Guinea type, with velar and coronal laterals (Mid-Waghi), or perhaps palatal laterals
d.	Faith >> *LatLab >> *LatDors >> *LatCor	Laterals at all POAs (unattested)

(35) *Typology of lateral sonorants and obstruents*

a.	*LatObs >> *LatSon >> Faith	Languages with no laterals
b.	*LatObs >> Faith >> *LatSon	Common language type, with sonorant laterals
c.	Faith>> *LatObs >> *LatSon	Languages with both obstruent and sonorant laterals

In this way, we can correctly describe the inventories of laterals found in the world's languages, with the possible exception of Tlingit, which is reported to have lateral obstruents but no lateral sonorants. The complete absence of labial laterals also remains unexplained.

8.2 The spreading of [lateral]

How can these mini-grammars characterize the behavior of laterals shown in §7? First, let us consider the behavior of [lateral] in spreading processes, such as in Place assimilation or perhaps the spreading of SV under pressure from the Syllable Contact Law. Let us assume that assimilation involves a violation of the Ident family of faithfulness constraints, such as Ident-Place and Ident-Son, features are shared, but not laterality, and the segment remains a nasal. For full details, see Yip (2003, 2005).

We see then that variability in the spreading behavior of laterals in assimilation is handled without great difficulty by these minimally different OT grammars.

8.3 Laterals as the targets of spreading

I now turn to cases where lateral is (potentially) lost on the target of assimilation, or in neutralization. Whether or not it is lost depends on the relative ranking of Share-F and Ident-Lat, as shown below:

(38) a. Loss of [lateral] under (Place) assimilation: Educated Havana Spanish
Share-F >> Ident-Lat, Ident-Place

b. Retention of [lateral] under (Place) assimilation: Basque
IdentLat >> Share-F >> Ident-Place

A parallel analysis can be constructed for the cases in which SV spreading does or does not obliterate laterality.

The Basque case deserves a little more attention, because the Basque facts were used by Mielke to argue that laterals behave as [−continuant], since like nasals they undergo place assimilation, whereas [+continuant] [r] does not. Under the account offered in this section, however, there is another explanation available, as suggested by a reviewer. Nasals assimilate to all places, but laterals assimilate only to coronals, since *LatLab and *LatDors are high-ranked. Suppose there is also a set of constraints regulating the co-occurrence of a feature [rhotic], such that only a single rhotic exists. If these constraints and Ident[rhotic] all outrank Share-F, the rhotic will fail to assimilate, and continuancy need not be invoked.

Finally, laterality may or may not be lost when Place or SV is neutralized, for example in codas. Here the grammar is modeled on the ones in (38), but with *Coda-F instead of Share-F.

I have now outlined analyses that explain all the variation in the behavior of laterality, without reference to feature geometry, and using only rather simple markedness constraints on the co-occurrence of [lateral] with the various Place features, and with the feature [sonorant]. These interact with familiar faithfulness constraints and with markedness constraints that create pressure for change, such as Share-F, SyllableContact or *Coda-F. The clear preferences for coronal and sonorant laterals are captured by positing universal rankings of the relevant feature co-occurrence constraints. As always in OT, these constraints are violable, and this correctly predicts the observed cross-linguistic variation in lateral behavior that is such a problem for a fixed universal feature geometry.

9 Conclusion

The topic of this chapter has been laterals. As I hope is clear, they do not fit tidily into current phonological theories, especially when it comes to their distinctive feature make-up. Considering how common they are in language, this is something of an embarrassment. After all, they are neither unusual, prone to disappear over time, nor hard to acquire (that is, even if the child acquires them late, they get them sooner or later, which is more than can be said for [s] and [ɹ] for some speakers of British English). It is therefore incumbent on a good theory to accommodate them.

We have seen that their behavior is more variable across languages than that of most other sounds. So, for example, [t] always behaves like a voiceless coronal stop, and its distinctive features and their placement in the feature geometry are rarely if ever in dispute. So why should laterals be different, and are there other classes of sounds that exhibit comparable variability?

Variable behavior might be seen whenever the features are most readily produced and perceived on a certain type of segment, but can with some effort also be produced and perceived on sounds of other types too. For example, [strident] sounds, in which the turbulence produced at the point of constriction is sufficiently strong, and/or where the ensuing airstream then hits a sharp obstacle like the teeth, is easy to produce with the tip or blade of the tongue, but hard to produce elsewhere. We derive from this a constraint hierarchy *[Labial, strident] >> *[Coronal, strident]. Languages which contrast [f] and [ɸ], like Ewe, arguably violate the former as well as the latter. Turbulent airflow also requires a period of incomplete closure, or continuancy, so we also derive *[−cont, strident] >> *[+cont, strident]. Languages that violate the former have strident affricates, which have often been argued to be strident stops. In principle, then, the interactions of these constraints might also produce comparable variation to that we have seen with laterals.

For other features, no such variation is to be expected. [anterior] and [distributed] refine the type of contact the tip or blade of the tongue makes with the roof of the mouth. As such they can only be present in Coronals, and a sound that is [Dorsal, +anterior] is phonetically uninterpretable.

Mielke, by contrast, takes the variability in behavior of "ambivalent segments" like laterals to be an argument against universally defined distinctive features. Instead, he argues for "emergent distinctive features" based on phonetic similarity. Laterals, for example, may pattern with either continuants (16 languages) or non-continuants (61 languages) because, like continuants, they do not have totally blocked airflow, but like non-continuants they do have "a blockage of airflow past the primary structure." It is not clear how his proposals bear on those cases in §7.1 where the variability of laterals concerns their coronality rather than their continuancy, and thus where no appeal to variation in continuancy would seem to explain their behavior. It might be possible, however, to develop an extension of his approach from the starting observation that laterals are produced with both coronal and dorsal tongue gestures. Languages might therefore differ in which gesture they choose to interpret as a distinctive place feature for laterals. I leave the fleshing out of this idea for future research.

A substantial part of this chapter has focused on where the feature [lateral] sits in the feature geometry, but the proposal outlined in §7 and §8 makes no use of feature geometry at all. In this respect it is entirely compatible with the proposals of Padgett (1995, 2002), in which features form classes, but in which these classes are not embodied as superordinate nodes (with the exception of a root node). Instead of asking which node dominates [lateral], one would ask which class(es) of features [lateral] belongs to. There is then no reason why it might not belong to more than one class, for example Place and SV (or Manner), or to none. The feature co-occurrence proposal goes further, however, in that as far as [lateral] is concerned it does not even make use of feature classes. For die-hard proponents of feature geometry, I should note that it would be possible to combine the feature co-occurrence proposal with feature-geometric representations if other phenomena make this desirable, but only if [lateral] were directly under the root node.

REFERENCES

Alba, Orlando. 1988. Estudio sociolingüístico de la variación de las líquidas finales de palabra en el español cibaeño. In Robert M. Hammond & Melvyn C. Resnick (eds.) *Studies in Caribbean Spanish dialectology*, 1–12. Washington, DC: Georgetown University.
Anderson, John M. & Colin J. Ewen. 1987. *Principles of dependency phonology*. Cambridge: Cambridge University Press.
Bao, Zhiming. 1992. A note on [Lateral]. Unpublished ms., Ohio State University.
Bauer, Robert S. & Paul K. Benedict. 1997. *Modern Cantonese phonology*. Berlin & New York: Mouton de Gruyter.
Beckman, Jill N. 1998. Positional faithfulness. Ph.D. dissertation, University of Massachusetts, Amherst. Published 1999, New York: Garland.
Blevins, Juliette. 1994. A place for lateral in the feature geometry. *Journal of Linguistics* 30. 301–348.
Botma, Bert. 2004. Phonological aspects of nasality: An element-based dependency approach. Ph.D. dissertation, University of Amsterdam.
Botma, Bert, Colin J. Ewen & Erik Jan van der Torre. 2008. The syllabic affiliation of post-vocalic liquids: An onset-specifier approach. *Lingua* 118. 1250–1270.
Bradley, Travis G. 2006. Contrast and markedness in complex onset phonotactics. *Southwestern Journal of Linguistics* 25. 29–58.
Brown, Cindy. 1995. The feature geometry of lateral approximants and lateral fricatives. In Harry van der Hulst & Jeroen van der Weijer (eds.) *Leiden in last*, 41–88. The Hague: Holland Academic Graphics.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.
Cohn, Abigail C. 1992. The consequences of dissimilation in Sundanese. *Phonology* 9. 199–220.
Evans, Nicholas D. 1995. Current issues in the phonology of Australian languages. In Goldsmith (1995), 723–761.
Flemming, Edward. 2002. *Auditory representations in phonology*. London & New York: Routledge.
Gick, Bryan, Fiona Campbell, Sunyoung Oh & Linda Tamburri-Watt. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* 34. 49–72.
Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.
Guitart, Jorge M. 1976. *Markedness and a Cuban dialect of Spanish*. Washington, DC: Georgetown University Press.
Guitart, Jorge M. 1985. Variable rules in Caribbean Spanish and the organization of phonology. In Frank H. Nuessel (ed.) *Current issues in Hispanic phonology and morphology*, 28–33. Bloomington: Indiana University Linguistics Club.
Hale, Mark & Charles Reiss. 2008. *The phonological enterprise*. Oxford: Oxford University Press.
Harris, James W. 1985. Autosegmental phonology and liquid assimilation in Havana Spanish. In Larry D. King & Catherine A. Maley (eds.) *Selected papers from the 13th Linguistic Symposium on Romance Languages*, 127–148. Amsterdam: John Benjamins.

Harris, James W. & Ellen M. Kaisse. 1999. Palatal vowels, glides and obstruents in Argentinian Spanish. *Phonology* 16. 117–190.

Hayes, Bruce. 1986. Assimilation as spreading in Toba Batak. *Linguistic Inquiry* 17. 467–499.

Hegarty, Michael. 1989. An investigation of laterals and continuancy. Unpublished ms., MIT.

Heid, Sebastian & Sarah Hawkins. 2000. An acoustical study of long-domain /r/ and /l/ coarticulation. *Proceedings of the 5th Seminar on Speech Production: Models and Data*, 77–80. Munich: Institut für Phonetik und Sprachliche Kommunikation, Ludwig-Maximilians-Universität.

Holt, D. Eric. 2002. The articulator group and liquid geometry: Implications for Spanish phonology present and past. In Caroline R. Wiltshire & Joaquim Camps (eds.) *Romance phonology and variation*, 85–99. Amsterdam & Philadelphia: John Benjamins.

Hsu, Chai-Shune. 1996. Voicing underspecification in Taiwanese word-final consonants. *UCLA Working Papers in Linguistics* 2: *Papers in Phonology* 3. 1–24.

Hualde, José Ignacio. 1991. *Basque phonology*. London & New York: Routledge.

Inkelas, Sharon & Yvan Rose. 2008. Positional neutralization: A case study from child language. *Language* 83. 707–736.

Iverson, Gregory K. & Hyang-Sook Sohn. 1994. Liquid representation in Korean. In Young-Key Kim-Renaud (ed.) *Theoretical issues in Korean linguistics*, 1–19. Stanford: CSLI.

Iverson, Paul, Patricia K. Kuhl, Reiko Akahane-Yamada, Eugen Diesch, Yohich Tohkura, Andreas Kettermann & Claudia Siebert. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87. B47–B57.

Johnson, Wyn & David Britain. 2007. L-vocalisation as a natural phenomenon: Explorations in sociophonology. *Language Sciences* 29. 294–315.

Kaisse, Ellen M. 2000. Laterals are [–continuant]. Unpublished ms., University of Washington.

Ladefoged, Peter. 2006. *A course in phonetics*. 5th edn. Boston: Thomas Wadsworth.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford & Malden, MA: Blackwell.

Li, Paul Jen-Kuei. 1974. Alternations between semi-consonants and fricatives or liquids. *Oceanic Linguistics* 13. 163–186.

Macken, Marlys A. 1980. The child's lexical representation: The '*puzzle-puddle-pickle*' evidence. *Journal of Linguistics* 16. 1–17.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian. 1987. Linguolabials. *Journal of the Acoustical Society of America* 81. S65.

Maddieson, Ian. 2008. Lateral consonants. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.) *World atlas of language structures online*, ch. 8. Munich: Max Planck Digital Library 2008. http://wals.info/feature/description/8.

Madelska, L. & M. Witaszek-Samborska. 1998. *Zapis fonetyczny*. Poznań: Wydawnictwo Naukowe UAM.

Mascaró, Joan. 1976. Catalan phonology and the phonological cycle. Ph.D. dissertation, MIT.

Mielke, Jeff. 2005. Ambivalence and ambiguity in laterals and nasals. *Phonology* 22. 169–203.

Mithun, Marianne & Hasan Basri. 1986. The phonology of Selayarese. *Oceanic Linguistics* 25. 210–254.

Morén, Bruce. 2006. Consonant–vowel interactions in Serbian: Features, representations and constraint interactions. *Lingua* 116. 1198–1244.

Nacaskul, Karnchana. 1978. The syllabic and morphological structure of Cambodian words. In Philip N. Jenner (ed.) *Mon-Khmer studies VII*, 183–200. Honolulu: University Press of Hawaii.

Narayanan, Shrikanth S., Abeer A. Alwan & Katherine Haker. 1997. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data, part I: The laterals. *Journal of the Acoustical Society of America* 101. 1064–1077.

Ohala, John J. 1974. Phonetic explanation in phonology. *Papers from the Annual Regional Meeting, Chicago Linguistic Society: Parasession on natural phonology*. 251–274.

Padgett, Jaye. 1991. Stricture in feature geometry. Ph.D. dissertation, University of Massachusetts, Amherst.

Padgett, Jaye. 1995. Feature classes. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 385–420. Amherst, MA: GLSA.

Padgett, Jaye. 2002. Feature classes in phonology. *Language* 78. 81–110.

Paradis, Carole & Jean-François Prunet (eds.) 1991. *The special status of coronals: Internal and external evidence.* San Diego: Academic Press.

Parker, Steve. 2008. Sound level protrusions as physical correlates of sonority. *Journal of Phonetics* 36. 55–90.

Piggott, Glyne L. 1991. The geometry of sonorant features. Unpublished ms., McGill University.

Piñeros, Carlos. 2003. Accounting for the instability of Palenquero voiced stops. *Lingua* 113. 1185–1222.

Quilis, A., M. Esgueva, M. L. Gutiérrez Araus & M. Canterero. 1979. Características acústicas de las consonantes laterales españolas. *Lingüística Española Actual* 1. 233–343.

Recasens, Daniel & Aina Espinosa. 2005. Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /ł/: Evidence from two Catalan dialects. *Journal of the International Phonetic Association* 35. 1–25.

Rice, Keren. 1994. Laryngeal features in Athapaskan languages. *Phonology* 11. 107–147.

Rice, Keren & Peter Avery. 1991. On the relationship between laterality and coronality. In Paradis & Prunet (1991), 101–124.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.

Scobbie, James M. & Alan Wrench. 2003. An articulatory investigation of word final /l/ and /l/ sandhi in three dialects of English. In M. J. Solé, D. Recasens & J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*, 1871–1874. Barcelona: Causal Productions.

Shaw, Patricia A. 1991. Consonant harmony systems: The special status of coronal harmony. In Paradis & Prunet (1991), 125–157.

Smith, Neil V. 2010. *Acquiring phonology: A cross-generational case-study.* Cambridge: Cambridge University Press.

Spencer, Andrew. 1984. Eliminating the feature [lateral]. *Journal of Linguistics* 20. 23–43.

Sproat, Richard & Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21. 291–311.

Steriade, Donca. 1987. Redundant values. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23(2). 339–362.

Sulkala, Helena & Merja Karjalainen. 1992. *Finnish.* London & New York: Routledge.

Trigo, Loren. 1988. On the phonological behavior and derivation of nasal glides. Ph.D. dissertation, MIT.

Walsh Dickey, Laura. 1994. Representing laterals. *Papers from the Annual Meeting of the North East Linguistic Society* 25. 535–550.

Walsh Dickey, Laura. 1997. The phonology of liquids. Ph.D. dissertation, University of Massachusetts, Amherst.

Weijer, Jeroen van de. 1995. Continuancy in liquids and in obstruents. *Lingua* 96. 45–61.

Whitney, William Dwight. 1889. *Sanskrit grammar, including both the classical language, and the older dialects of Veda and Brahmana.* 2nd edn. Cambridge, MA: Harvard University Press.

Yip, Moira. 1989. Feature geometry and cooccurrence restrictions. *Phonology* 6. 349–374.

Yip, Moira. 1992. Prosodic morphology in four Chinese dialects. *Journal of East Asian Linguistics* 1. 1–35.

Yip, Moira. 2001. Segmental unmarkedness versus input preservation in reduplication. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory*, 206–228. Cambridge: Cambridge University Press.

Yip, Moira. 2003. Some real and not-so-real consequences of comparative markedness. *Theoretical Linguistics* 29. 53–64

Yip, Moira. 2004. Lateral survival: An OT account. *International Journal of English Studies* 4(2). 25–51.

Yip, Moira. 2005. Variability in feature affiliations through violable constraints: The case of [lateral]. In Marc van Oostendorp & Jeroen van de Weijer (eds.) *The internal organization of phonological segments*, 63–91. Berlin & New York: Mouton de Gruyter.

Yu, Alan C. L. 2007. *A natural history of infixation*. Oxford: Oxford University Press.

Volume I: General Issues and Segmental Phonology

Volume II: Suprasegmental and Prosodic Phonology

Volume III: Phonological Processes

Volume IV: Phonological Interfaces

Volume V: Phonology across Languages

The Blackwell Companion to Phonology is a major reference work drawing together 124 new contributions from leading scholars in the field. Led by a team of renowned international scholars, the *Companion* represents a diverse range of approaches and methodologies to key phenomena in phonological research. In contrast to other handbooks and reference works currently available for phonology, these volumes focus on phenomena and case studies that highlight past and ongoing debates in the field. Interdisciplinary connections, such as those with the social and computational sciences, are covered, as well as statistical and experimental techniques. The *Companion* will be a touchstone for future phonological theorists, giving an overview of data and insights which any theory of phonology should be able to cover.

Simultaneously offering broad coverage and a high level of detail, *The Blackwell Companion to Phonology* is a landmark work available in both print and online editions that will be indispensable to students and researchers in the field for years to come.

www.companiontophonology.com

Cover image: Theo van Doesburg, Color scheme for Café Aubette, Strasbourg, France 1926. New York, Museum of Modern Art (MoMA). Ink and gouache on paper, 27.2 x 62.9 cm. Gift of Lily Auchincloss, Celeste Bartos, and Marshall Cogan 593.1982 © 2010. Digital image, The Museum of Modern Art, New York / Scala, Florence.

Cover design by Nicki Averill

WILEY-BLACKWELL
www.wiley.com/wiley-blackwell

VOLUME II
Suprasegmental and Prosodic Phonology

Edited by
van Oostendorp,
Ewen, Hume,
and Rice

THE BLACKWELL COMPANION TO
Phonology

WILEY-BLACKWELL

VOLUME II
Suprasegmental and Prosodic Phonology

Edited by
Marc van Oostendorp
Colin J. Ewen
Elizabeth Hume
Keren Rice

THE BLACKWELL COMPANION TO Phonology

WILEY-BLACKWELL

MARC VAN OOSTENDORP is Senior Researcher at the Department of Variationist Linguistics at the Meertens Institute of the Royal Netherlands Academy of Arts and Sciences, and Professor of Phonological Microvariation at the University of Leiden. He holds an M.A. in Computational Linguistics and a Ph.D. from Tilburg University.

COLIN J. EWEN is Professor and Chair of English Language and Linguistics at the University of Leiden. He holds an M.A. in English Language and Linguistics and a Ph.D. from the University of Edinburgh. He is editor (with Ellen Kaisse) of the journal *Phonology*.

ELIZABETH HUME is Professor and Chair of the Department of Linguistics at The Ohio State University. She holds a Ph.D. and M.A. in Linguistics from Cornell University, an M.A. in French and Social Psychology of Language from McMaster University, and a B.A. in French from Université Laval.

KEREN RICE is University Professor and Canada Research Chair in Linguistics and Aboriginal Studies at the University of Toronto. She holds an M.A. and Ph.D. in Linguistics from the University of Toronto and a B.A. from Cornell University.

This edition first published 2011

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

The Blackwell companion to phonology / edited by Marc van Oostendorp . . . [et al.].
p. cm. — (Blackwell companions to linguistics series)
Includes bibliographical references and index.
ISBN 978-1-4051-8423-6 (hardcover : alk. paper)
1. Phonetics. 2. Grammar, Comparative and general—Phonology.
I. Oostendorp, Marc van, 1967–
P217.B53 2011
414—dc22 2010042206

A catalogue record for this book is available from the British Library.

Set in 10/12pt Palatino by Graphicraft Limited, Hong Kong

1 2011

Brief Contents

Volume I

Volume II

Volume III

Volume IV

Volume V

Full Table of Contents

Volume I

Volume II

Suprasegmental and Prosodic Phonology

Volume III

Phonological Processes

Volume IV

Phonological Interfaces

Volume V

Phonology across Languages

32 The Representation of Intonation

Amalia Arvaniti

1 Introduction

It is a well-known truism that no utterance is ever produced in a strict monotone; all utterances, in all languages, show some pitch modulation. Such changes in pitch – impressionistically described as rises and falls – are due to changes in fundamental frequency or F0, the physical property of the speech signal that is determined by the basic rate of vibration of the vocal folds and gives rise to the percept of pitch.

Although pitch modulations exist in all languages, their origin and function differ, in that pitch patterns may be specified either at both the lexical and phrasal levels or only at the phrasal level, resulting in more or less dense tonal specifications, respectively (Gooden *et al.* 2009). The term *intonation* is used to refer to phrasal tonal patterns, while the terms *pitch accent* and *tone* are traditionally used to refer to lexical tonal specifications (chapter 45: the representation of tone). Simplifying somewhat, in languages like English, Italian, Greek, and many other European languages the entire F0 contour is specified at the phrasal level by means of a complex interplay between metrical structure, prosodic phrasing, syntax, and pragmatics; these factors determine where pitch movements will occur and of what type they will be. In languages referred to as *tone languages* – such as Mandarin, Thai, and Igbo – most syllables are lexically specified for tone and tonal changes affect lexical meaning; in languages often referred to as *pitch accent languages* – such as Japanese, Swedish, and Serbian – tone operates in a similar fashion, except that at most one syllable in each word is lexically specified for tone. In both tone and pitch accent languages additional tonal patterns are specified at the phrasal level. Here the focus is on languages without lexical tonal specifications, since it is the intonation of these languages that has been mostly examined.

Determining the structure of pitch modulation and the primitives that make up pitch contours in languages without lexical tone is challenging, since F0 changes are not as discrete and easily identifiable as in "tonal" languages, their connections to segmental material are less easy to determine, and associated meanings are harder to pinpoint since they deal with information structure and pragmatic interpretation rather than lexical semantics. The following examples illustrate these points. In Figure 32.1, two utterances are shown, *Me?!* and *A*

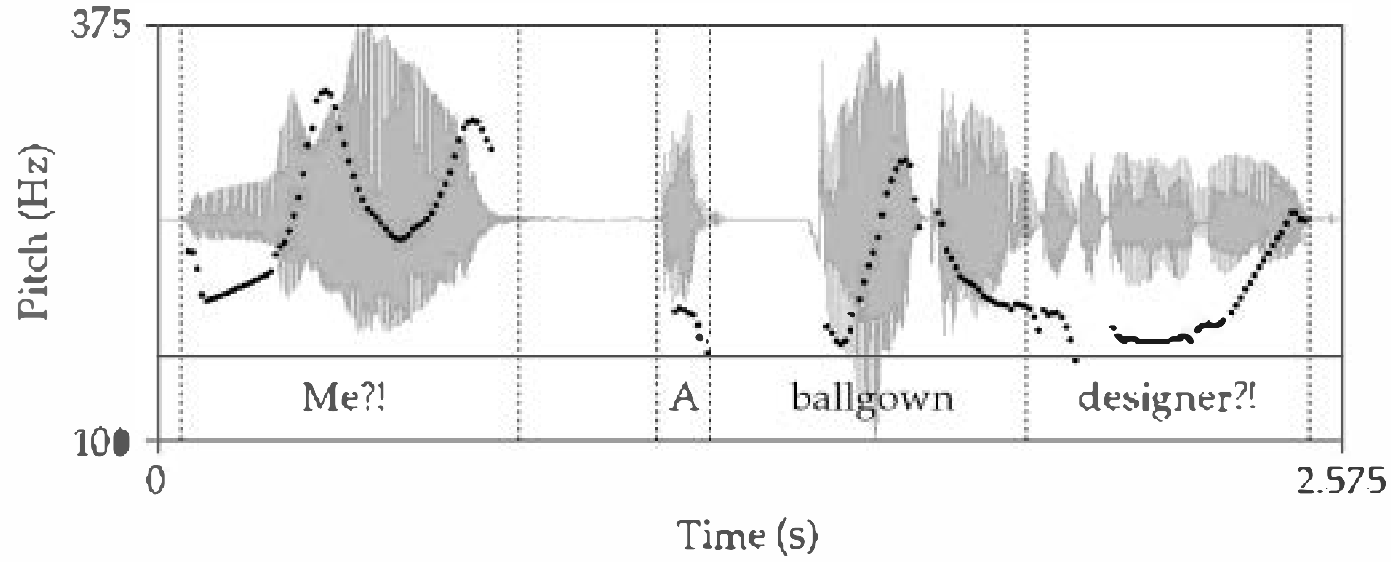


Figure 32.1 Waveforms and F0 contours of two English utterances illustrating the incredulity contour (Hirschberg and Ward 1992); on the left, *Me?!*; on the right, *A ballgown designer?!*. Vertical lines indicate word boundaries

ballgown designer?!, both using the rise-fall-rise melody that implies incredulity (Ward and Hirschberg 1985; Hirschberg and Ward 1992). They are plausible responses to a career advisor's pronouncement that, according to test results, designing ballgowns is the recommended career choice for the speaker, who has all along dreamed of becoming an aerospace engineer (for similar examples, see Ladd 2008: 45–46). Although the short contour can be informally described as rise–fall–rise, the longer contour cannot be described in a similar fashion, as it shows a long low-level stretch between a rise–fall and a final rise. In Figure 32.2, Greek contours very similar to the English ones in Figure 32.1 are shown, though in the case of Greek these contours are used for wh-questions (Arvaniti and Baltazani 2005; Arvaniti and Ladd 2009). As can be seen, the same issue with overall shape arises here as well. Further, as Arvaniti and Baltazani (2005) note, the Greek melody in Figure 32.2 can also be used for polite requests employing an imperative; e.g. [ˈðose sti maˈria ˈliɣo neˈraci] 'give Maria some water' (*lit.* give to Maria a-little water-DIM). Finally, Figure 32.3 illustrates two instances of another English melody: unlike the contours in Figures 32.1 and 32.2, which look different from each other but convey the same meaning in each case, the contours of Figure 32.3 are realizations of the same melody but convey different meaning, depending on the utterance: the melody used is

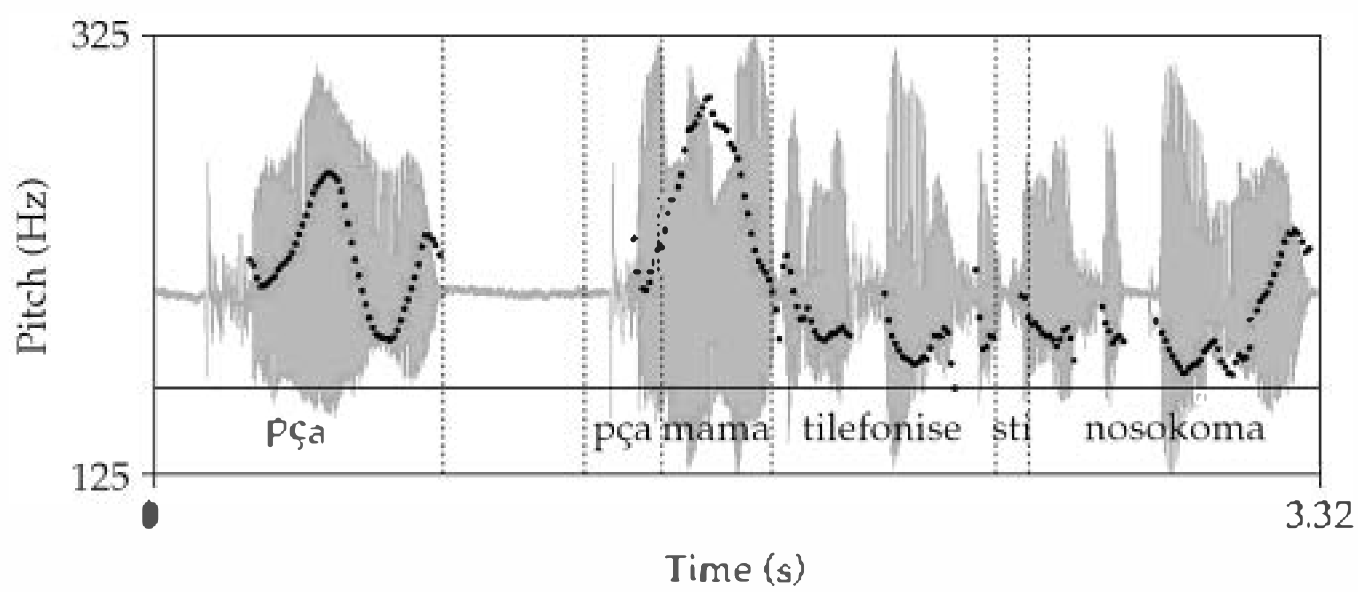


Figure 32.2 Waveforms, transcriptions, and F0 contours of two Greek wh-questions, on the left, [ˈpça] 'which (FEM)', on the right, [ˈpça maˈma tileˈfonise sti nosoˈkoma] 'which mom called the nurse?' Vertical lines indicate word boundaries

2.4 *Superpositional models*

In a number of configurational models, contours are said to be composed of two elements, a general trend and local perturbations which "ride" on this overall movement. This conception of intonation was also espoused by Bolinger, who distinguished *accentuation* from *intonation*, using accentuation to refer to pitch movements (*accents*) on stressed syllables, and intonation to refer to the general course of F0, "the rise and fall of pitch as it occurs along the speech chain" (Bolinger 1986: 194).

The IPO system is one such superpositional model, in that its primitives and contours are seen as localized movements superposed on a larger declination component, which is taken to be largely automatic and due to the drop in subglottal pressure during the course of an utterance (declination reset and local movements, on the other hand, are seen as actively controlled by the speaker; 't Hart *et al.* 1990: ch. 5). The exact role of declination and its physiology are still a matter of debate (see e.g. Pierrehumbert and Beckman 1988: ch. 3; Gussenhoven 2004: ch. 6), though evidence such as that provided for Japanese downtrends by Pierrehumbert and Beckman (1988) does not support the idea of declination playing as important a role as the IPO scholars envisioned.

Superpositional models have also been presented by Fujisaki (1983, 2004), Gårding (1983, 1987), and Thorsen (1980, 1985, 1986). Simplifying somewhat, in Fujisaki's system a *phrase command* results in the rising–falling course of F0 throughout an utterance (or a part thereof), with *accent commands* being responsible for more localized perturbations. Gårding (1983, 1987) posits grids (quasi parallel lines) within which most local F0 minima and maxima can be fitted; the overall range and direction of the grid (rising or falling or a combination thereof) reflect functional differences between utterances, such as the distinction between statements and questions. Similarly, Thorsen (1980: 1022) suggests that the rate of F0 drop in Danish is directly related to utterance function: "falling intonation contours are associated with declarative, intermediate contours with nonfinal, and flat contours with interrogative sentences." Due to the connection between communicative functions and overall F0 trends, the models of Thorsen and Gårding face similar issues to gestalt models with respect to meaning. On the other hand, Fujisaki's model, which does not rely on meaning distinctions, must resort to counterintuitive solutions – such as negative accent commands and phase commands that span linguistically arbitrary stretches – in order to adequately describe the course of F0 in languages other than Japanese (e.g. Fujisaki *et al.* 1997 on Greek; Fujisaki *et al.* 2005 on Mandarin; Gu *et al.* 2007 on Cantonese; for a discussion of these problems, see Ladd 2008: 23ff.; Arvaniti and Ladd 2009).

3 Pitch levels as primitives

3.1 *Early level-based models*

Descriptions of intonation by means of level tones date from the American structuralists (Pike 1945; Trager and Smith 1951; Hockett 1955; Trager 1961). In these systems, intonation is analyzed by means of four levels, extra-high, high, mid, and low. The level tones of these analyses are meant to be phonological abstractions equivalent to phonemes (CHAPTER 11: THE PHONEME); as such, they are said to be defined relative to each other, rather than each representing a specific pitch range.

These early analyses were heavily criticized by configurationalists, most notably Bolinger (1951), who questioned the claim that the four levels are relative, pointing out that if this assertion is taken at face value, it is not possible to distinguish combinations such as 123 from 234, although, theoretically, such combinations should be distinct. Thus, Bolinger concluded that level tones cannot be relative but must "rove each in its own bailiwick" (1951: 200), and set out to test this hypothesis by recording utterances in various *ways* (as Bolinger termed them, i.e. melodies differing in various aspects) and having listeners judge them for similarity or appropriateness for a given purpose, such as appeasing a child. His results showed that contours analyzed as contrastive in the system of Trager and Smith (whom he particularly targeted) are perceived as similar by listeners, while others, analyzed as allophones of the same basic melody, are considered by listeners to be contrastive. Bolinger used his results to argue that a system with four levels can be at the same time too powerful and not powerful enough to capture contrastive and allophonic variations in the intonational system of English. His results led him to reject level tone analyses as untenable and to propose instead that melodies are gestalts.

Bolinger's critique reflects the assumptions of concreteness and bi-uniqueness prevalent at the time. It is clear, for example, that Bolinger expected the different levels to faithfully represent the entire course of an utterance's F0, and to do so in such a way that the pitch range of each level did not overlap with that of others at any point in the utterance. Further, his comment that F0 forms "a continuous line that can be traced on a piece of paper" (Bolinger 1951: 206), coupled with his distinction between monotones, which he accepts, and level tones, which he does not (e.g. Bolinger 1986: 29), suggest that he expected a level-based representation to be phonetically realized as a series of sustained pitch levels. It is obvious that if these assumptions are adopted, a level-based analysis is unworkable on both phonological and phonetic grounds (on the latter, see Xu and Sun 2002; Dilley and Brown 2007).

Although Bolinger's critique was well accepted, it is fair to note that many of the assumptions he attributes to the structuralists are not found in their works. Pike (1945), Hockett (1955), and Trager (1961) all note that levels represent "only those points in the contour crucial to the establishment of its characteristic rises and falls" (Pike 1945: 26); with the exception of terminal junctures, these points associate with stressed syllables. Similarly, the structuralists noted that absolute pitch levels are not signifiicant as such, and recognized the existence of both level tones and contours (e.g. Trager and Smith's terminal juncture phonemes). Further, Pike (1945) discusses at length the fact that level tones need not be realized as a series of sustained pitches but can be realized as glides, especially when they are found close to each other, as happens in short utterances, for example. Nevertheless, as a result of configurationalist critiques, level-based analyses were largely abandoned in the following decades. Research within the generative framework focused primarily on the description of tone languages and no theoretical position was strongly taken in favor either of levels or of configurations with respect to intonation.

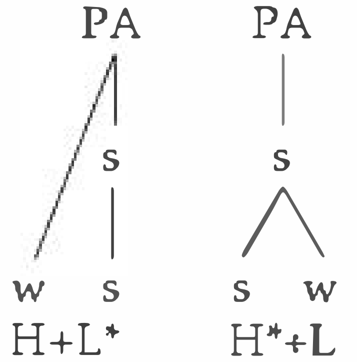
The role of underspecification cannot be underestimated: it leads to a clear understanding that the number of tones need not match the number of tone-bearing units (TBUs), so that both strings of tonally unspecified TBUs and instances of several tones associating with the same TBU are possible. Thus, in the AM framework, the two English contours in Figure 32.1 are both analyzed as L*+H L- H%. By using the same representation for these two contours, AM captures the fact that they are instantiations of the same melody, thereby generalizing beyond surface form. AM can also account for the systematic differences between contours like those in Figure 32.1, which, as mentioned earlier, were particularly problematic for the British school. In *Me?!*, all three tonal events must co-occur with the only syllable of this utterance; hence the obvious lengthening of *me* (720 msec) and the swift movement from one tonal target to the next. In *A ballgown designer?!*, L*+H is associated with the metrically strongest syllable in the utterance, i.e. *ball*, and it co-occurs with it (showing the peak delay expected for this accent; e.g. Pierrehumbert and Steele 1989; Arvaniti and Garding 2007). The H% is realized on the last syllable, which is the one showing a rise. The L-, which is associated with the ip boundary, spreads between the L*+H and H%, accounting for the fall and low-level stretch of F0 (for details on the realization of the L- in such contours, see Grice *et al.* 2000; Barnes *et al.* 2006). A similar analysis applies to the Greek wh-questions shown in Figure 32.2, analyzed as L*+H L- !H% (where !H refers to a downstepped H tone; Grice *et al.* 2000; Arvaniti and Baltazani 2005; Arvaniti and Ladd 2009).

Overall, the AM model avoids several pitfalls of previous analyses. First, by formally separating stress from intonation and providing a mechanism for their interaction, the AM model incorporates the insights of Bolinger about pitch accents without requiring distinct accentual and phrasal components to account for pitch contours. In addition, the use of only H and L tones avoids the problems noted by Bolinger (1951) with respect to level tones. At the same time, by treating the issue of pitch range as a matter of phonetic realization, AM avoids the problems that plagued the British analyses due to the confounding of linguistic and paralinguistic aspects of pitch range (cf. the disagreements regarding whether high falls and low falls are distinct entities). Further, by making explicit the separation between the phonetics and phonology of intonation, the AM model provides a principled account of the context-dependent variation of tones, a point that was not explicitly addressed in previous models, which mostly confounded contours and representations. Finally, the use of underspecification provides a parsimonious and elegant way of capturing both the similarities of melodies and the differences in phonetic realization that arise from the properties of the metrical structure with which a melody associates. In this way, the model can account for both local phonetic detail and abstract phonological form, something that configurational and full specification models cannot do (for extensive discussions of this point, see Pierrehumbert and Beckman 1988; Arvaniti *et al.* 2006a; Arvaniti and Ladd 2009). Finally, since the degree of underspecification can vary, AM can account for languages with dense tonal specifications, such as Mandarin, as well as for languages with more sparse specifications, such as English. In short, then, AM not only provides an answer regarding the nature of intonational primitives, but crucially addresses the even more fundamental question of what should be represented phonologically when it comes to intonation, an issue that most other theories have not tackled by focusing exclusively on faithful representations of entire F0 curves.

determined if none of the properties that may define them is stable. Finally, it is not clear how the notion of an indivisible unit can be defended for the Greek accents at all, since the beginning and ending points of the rise do not behave as one. Their relative autonomy is demonstrated by the fact that they align independently of each other and are not similarly affected by tonal crowding, which typically results in the undershooting of the L, while the realization of the H remains largely unaltered (Arvaniti *et al.* 2000). This pattern is difficult to account for if the rise is a unit, in which case one would more plausibly expect a curtailment of the entire pitch movement.

It is thus clear that dynamic tones cannot account for some of the attested patterns. On the other hand, level tones can be used not only for the representation of loosely defined rises and falls, as in Greek, but also for rises and falls that are more closely knit. Such units have been reported by Frota (2002) for European Portuguese. Specifically, Frota found that in the falling accent indicating broad focus, the H and L are timed with respect to distinct segments (similarly to the Greek case), but the fall of the accent indicating narrow focus shows a constant timing relationship between the H and L tones (similar to that discussed by Pierrehumbert 1980 for L*+H and L+H* in English). This difference between the two accents of European Portuguese can be represented by means of a hierarchical representation of tones shown in (1), as first proposed by Grice (1995a, 1995b) and adopted by Frota, or it can be treated as an issue of phonetic realization, as argued in Arvaniti *et al.* (2006b). Either way, it is clear that while level tones can adequately describe all attested tonal patterns, dynamic tones cannot. In short, then, both the empirical evidence and phonological considerations of parsimony and descriptive adequacy make a theory based on level tones preferable.

(1)



5 Conclusion

The original debate about levels *vs.* configurations (Bolinger 1951) focused on the issue of whether melodies are gestalts or should be seen as composites of primitives. Independently of this distinction, Bolinger's views, espoused by many before him and since, are based on the idea that intonational contours should be represented in their entirety, either as a series of primitives, or as a "line [. . .] on a piece of paper." Current understanding suggests that couching the problem in these terms is misleading, as neither type of representation is likely to be correct: as shown, gestalt approaches cannot account in a satisfactory manner for either intonational meaning or intonational form; yet representations that fully specify the course of F0, either in terms of dynamic tones or in terms of levels, do not fare much better. Due to the particularities of intonation, especially the fact that its realization depends on the metrical structure of the utterance with

Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2010. The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation. *Proceedings of Speech Prosody 2010*. Available (August 2010) at http://speechprosody2010.illinois.edu/papers/100445.pdf.

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.

Bishop, Judith & Janet Fletcher. 2005. Intonation in six dialects of Bininj Gun-wok. In Jun (2005b), 331–361.

Bolinger, Dwight L. 1951. Intonation: Levels versus configurations. *Word* 7. 199–210.

Bolinger, Dwight L. 1964. Around the edge of language: Intonation. *Harvard Educational Review* 34. 282–293. Reprinted in Bolinger (1972), 19–29.

Bolinger, Dwight L. (ed.) 1972. *Intonation: Selected readings*. Harmondsworth: Penguin.

Bolinger, Dwight L. 1986. *Intonation and its parts: Melody in spoken English*. London: Edward Arnold.

Botinis, Antonis. 1989. *Stress and prosodic structure in Greek: A phonological, acoustic, physiological and perceptual study*. Lund: Lund University Press.

Broe, Michael B. & Janet B. Pierrehumbert (eds.) 2000. *Papers in laboratory phonology V: Acquisition and the lexicon*. Cambridge: Cambridge University Press.

Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.

Bryant, Gregory A. & Jean E. Fox Tree. 2005. Is there an ironic tone of voice? *Language and Speech* 48. 257–277.

Büring, Daniel. 2007. Intonation, semantics and information structure. In Gillian Ramchand & Charles Reiss (eds.) *The Oxford handbook of linguistic interfaces*, 445–473. Oxford: Oxford University Press.

Calhoun, Sasha. 2010. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language* 86. 1–42.

Chen, Yiya. 2006. Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics* 34. 176–201.

Chen, Yiya. 2010. Post-focus F_0 compression: Now you see it, now you don't. *Journal of Phonetics* 38. 517–525.

Cohen, Antonie & Johan 't Hart. 1968. On the anatomy of intonation. *Lingua* 19. 177–192.

Cohn, Abigail C., Cécile Fougeron & Marie Huffman (eds.) Forthcoming. *The Oxford handbook of laboratory phonology*. Oxford: Oxford University Press.

Cooper, William E. & John Sorensen. 1981. *Fundamental frequency in sentence production*. Heidelberg: Springer.

Crystal, David. 1972. The intonation system of English. In Bolinger (1972), 110–136.

Dainora, Audra. 2001. An empirically based probabilistic model of intonation in American English. Ph.D. dissertation, University of Chicago.

Dainora, Audra. 2006. Modeling intonation in English: A probabilistic approach to phonological competence. In Goldstein *et al.* (2006), 107–132.

de Jong, Kenneth J. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97. 491–504.

Dilley, Laura C. & Meredith Brown. 2007. Effects of pitch range variation on f_0 extrema in an imitation task. *Journal of Phonetics* 35. 523–551.

D'Imperio, Mariapaola. 2000. The role of perception in defining tonal targets and their alignment. Ph.D. dissertation, Ohio State University.

D'Imperio, Mariapaola. Forthcoming. Prosodic representations. (Part II: Tonal alignment). In Cohn *et al.* (forthcoming).

D'Imperio, Mariapaola & David House. 1997. Perception of questions and statements in Neapolitan Italian. *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech 1997)*, vol. 1, 251–254. Rhodes, Greece.

D'Imperio, Mariapaola, Jacques Terken & Michel Pitermann. 2000. Perceived tone "targets" and pitch accent identification in Italian. In Michael Barlow (ed.) *Proceedings of the 8th International Conference on Speech Science and Technology*, 206–211. Canberra: Australian Speech Science and Technology Association.

D'Imperio, Mariapaola, Barbara Gili Fivela & Oliver Niebuhr. 2010. Alignment perception of high intonational plateaux in Italian and German. *Proceedings of Speech Prosody 2010*. Available (August 2010) at http://speechprosody2010.illinois.edu/papers/100186.pdf.

Elordieta, Gorka & José Ignacio Hualde. 2003. Tonal and durational correlates of accent in contexts of downstep in Lekeitio Basque. *Journal of the International Phonetic Association* 33. 195–209.

Fletcher, Janet, Esther Grabe & Paul Warren. 2005. Intonational variation in four dialects of English: The high rising tune. In Jun (2005b), 390–409.

Fougeron, Cécile & Sun-Ah Jun. 1998. Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics* 26. 45–69.

Frota, Sónia. 2002. Tonal association and target alignment in European Portuguese nuclear falls. In Gussenhoven & Warner (2002), 387–418.

Fujisaki, Hiroya. 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. In Peter F. MacNeilage (ed.) *The production of speech*, 39–55. New York: Springer.

Fujisaki, Hiroya. 2004. Information, prosody and modeling – with emphasis on tonal features of speech. In Bernard Bel & Isabelle Marlien (eds.) *Speech Prosody* 2004. Available (August 2010) at www.isca-speech.org/archive/sp2004/sp04_001.pdf.

Fujisaki, Hiroya, Sumio Ohno & Takashi Yagi. 1997. Analysis and modelling of fundamental frequency contours of Greek utterances. *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech 1997)*, vol. 2, 465–468. Rhodes, Greece.

Fujisaki, Hiroya, Changfu Wang, Sumio Ohno & Wentao Gu. 2005. Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command–response model. *Speech Communication* 47. 59–70.

Gårding, Eva. 1983. A generative model of intonation. In Anne Cutler & D. Robert Ladd (eds.) *Prosody: Models and measurements*, 11–25. Berlin: Springer.

Gårding, Eva. 1987. Speech act and tonal pattern in standard Chinese: Constancy and variation. *Phonetica* 44. 13–29.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT. Published 1979, New York: Garland.

Goldstein, Louis, Douglas Whalen & Catherine T. Best (eds.) 2006. *Laboratory phonology 8*. Berlin & New York: Mouton de Gruyter.

Gooden, Shelome, Kath-Ann Drayton & Mary Beckman. 2009. Tone inventories and tune-text alignments: Prosodic variation in "hybrid" prosodic systems. *Studies in Language* 33. 354–394.

Gordon, Matthew. 2005. Intonational phonology of Chickasaw. In Jun (2005b), 301–330.

Gordon, Matthew. 2008. Pitch accent timing and scaling in Chickasaw. *Journal of Phonetics* 36. 521–535.

Grabe, Esther, Brechtje Post, Francis Nolan & Kimberley Farrar. 2000. Pitch accent realization in four varieties of British English. *Journal of Phonetics* 28. 161–185.

Grabe, Esther, Greg Kochanski & John Coleman. 2003. Quantitative modelling of intonational variation. *Proceedings of Speech Analysis and Recognition in Technology, Linguistics and Medicine 2003*. Available (August 2010) at www.phon.ox.ac.uk/oxigen/publications.php.

Grice, Martine. 1995a. *The intonation of interrogation in Palermo Italian: Implications for intonation theory*. Tübingen: Niemeyer.

Grice, Martine. 1995b. Leading tones and downstep in English. *Phonology* 12. 183–233.

Grice, Martine, D. Robert Ladd & Amalia Arvaniti. 2000. On the place of phrase accents in intonational phonology. *Phonology* 17. 143–185.

Gu, Wentao, Keikichi Hirose & Hiroya Fujisaki. 2007. Analysis of tones in Cantonese speech based on the command–response model. *Phonetica* 64. 29–62.

Gussenhoven, Carlos. 1999. Discreteness and gradience in intonational contrasts. *Language and Speech* 42. 283–305.

Gussenhoven, Carlos. 2000. The boundary tones are coming: On the non-peripheral realization of boundary tones. In Broe & Pierrehumbert (2000), 132–151.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos & Natasha Warner (eds.) 2002. *Laboratory phonology 7*. Berlin & New York: Mouton de Gruyter.

Halliday, M. A. K. 1967. *Intonation and grammar in British English*. The Hague: Mouton.

Halliday, M. A. K. 1970. *A course in spoken English: Intonation*. Oxford: Oxford University Press.

Harrington, Jonathan, Janet Fletcher & Mary E. Beckman. 2000. Manner and place conflicts in the articulation of accent in Australian English. In Broe & Pierrehumbert (2000), 40–51.

Hart, Johan 't & Antonie Cohen. 1973. Intonation by rule: A perceptual quest. *Journal of Phonetics* 1. 309–327.

Hart, Johan 't & René Collier. 1975. Integrating different levels of intonation analysis. *Journal of Phonetics* 3. 235–255.

Hart, Johan 't, René Collier & Antonie Cohen. 1990. *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Hayes, Bruce & Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9. 47–96.

Hirschberg, Julia & Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise–fall–rise intonation contour in English. *Journal of Phonetics* 20. 241–251.

Hirst, Daniel & Albert Di Cristo. 1998. A survey of intonation systems. In Daniel Hirst & Albert Di Cristo (eds.) *Intonation systems: A survey of twenty languages*, 1–44. Cambridge: Cambridge University Press.

Hirst, Daniel, Albert Di Cristo & Robert Espesser. 2000. Levels of representation and levels of analysis for the description of intonation systems. In Merle Horne (ed.) *Prosody: Theory and experiment*, 51–87. Dordrecht: Kluwer.

Hockett, Charles F. 1955. *A manual of phonology*. Baltimore: Waverly Press.

Hualde, José Ignacio, Gorka Elordieta, Inaki Gaminde & Rajka Smiljanić. 2002. From pitch accent to stress accent in Basque. In Gussenhoven & Warner (2002), 547–584.

Jones, Daniel. 1972. *An outline of English phonetics*. 9th edn. Cambridge: Heffer & Sons.

Jun, Sun-Ah. 2005a. Korean intonational phonology and prosodic transcription. In Jun (2005b), 201–229.

Jun, Sun-Ah (ed.) 2005b. *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Kingdon, Roger. 1958. *The groundwork of English Intonation*. London: Longmans, Green & Co.

Knight, Rachael-Anne & Francis Nolan. 2006. The effect of pitch span on intonational plateau. *Journal of the International Phonetic Association* 36. 21–38.

Ladd, D. Robert. 1980. The structure of intonational meaning: Evidence from English. Bloomington: Indiana University Press.

Ladd, D. Robert. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Ladd, D. Robert. 2008. *Intonational phonology*. 2nd edn. Cambridge: Cambridge University Press.

Ladd, D. Robert, Ineke Mennen & Astrid Schepman. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107. 2685–2696.

Ladd, D. Robert, Astrid Schepman, Laurence White, Louise May Quarmby & Rebekah Stackhouse. 2009. Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics* 37. 145–161.

Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT.

Leben, William R. 1976. The tones of English intonation. *Linguistic Analysis* 2. 69–107.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.

Liberman, Mark. 1975. The intonational system of English. Ph.D. dissertation, MIT.

Liberman, Mark & Janet B. Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*, 157–233. Cambridge, MA: MIT Press.

Morén, Bruce & Elizabeth C. Zsiga. 2006. The lexical and post-lexical phonology of Thai tones. *Natural Language and Linguistic Theory* 24. 113–178.

Myers, Scott. 2003. F0 timing in Kinyarwanda. *Phonetica* 60. 71–97.

O'Connor, J. D. & G. F. Arnold. 1973. *Intonation of colloquial English*. London: Longman.

Ohala, John J. 1983. Cross-language view of pitch: An ethological view. *Phonetica* 40. 1–18.

Ohala, John J. & William G. Ewan. 1973. Speed of pitch change. *Journal of the Acoustical Society of America* 53. 345.

Pan, Ho-Hsien. 2008. Focus and Taiwanese unchecked tones. In Chungmin Lee, Matthew Gordon & Daniel Büring (eds.) *Topic and focus: Cross-linguistic perspectives on meaning and intonation*, 195–213. Dordrecht: Springer.

Pierrehumbert, Janet B. 1980. The phonology and phonetics of English intonation. Ph.D. dissertation, MIT.

Pierrehumbert, Janet B. & Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.

Pierrehumbert, Janet B. & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.) *Intentions in communication*, 271–311. Cambridge MA: MIT Press.

Pierrehumbert, Janet B. & Shirley Steele. 1989. Categories of tonal alignment in English. *Phonetica* 46. 181–196.

Pike, Kenneth L. 1945. *The intonation of American English*. Ann Arbor: University of Michigan Press.

Prieto, Pilar. 2005. Stability effects in tonal clash contexts in Catalan. *Journal of Phonetics* 33. 215–242.

Prieto, Pilar & Francisco Torreira. 2007. The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35. 473–500.

Prieto, Pilar, Chilin Shih & Holy Nibert. 1996. Pitch downtrend in Spanish. *Journal of Phonetics* 24. 445–473.

Redi, Laura. 2003. Categorical effects in the production of pitch contours in English. In M. J. Solé, D. Recasens & J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*, 2921–2924. Barcelona: Causal Productions.

Rialland, Annie. 2007. Question prosody: An African perspective. In Carlos Gussenhoven & Tomas Riad (eds.) *Tones and tunes*, vol. 1: *Typological studies in word and sentence prosody*, 35–62. Berlin & New York: Mouton de Gruyter.

Rietveld, Toni & Carlos Gussenhoven. 1995. Aligning pitch targets in speech synthesis: Effects of syllable structure. *Journal of Phonetics* 23. 375–385.

Silverman, Kim E. A. 1987. The structure and processing of fundamental frequency contours. Ph.D. dissertation, Cambridge University.

Silverman, Kim E. A. & Janet B. Pierrehumbert. 1990. The timing of prenuclear high accents in English. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 72–106. Cambridge: Cambridge University Press.

Smiljanić, Rajka. 2006. Early vs. late focus: Pitch-peak alignment in two dialects of Serbian and Croatian. In Goldstein *et al.* (2006), 494–518.

Steedman, Mark. 2000. Information structure and the syntax–phonology interface. *Linguistic Inquiry* 31. 649–689.

Sundberg, Johan. 1979. Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7. 71–79.

Swerts, Marc, Emiel Krahmer & Cinzia Avesani. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30. 629–654.

Thorsen, Nina. 1980. A study of the perception of sentence intonation: Evidence from Danish. *Journal of the Acoustical Society of America* 67. 1014–1030.

Thorsen, Nina. 1985. Intonation and text in Standard Danish. *Journal of the Acoustical Society of America* 77. 1205–1216.

Thorsen, Nina. 1986. Sentence intonation in textual context: Supplementary data. *Journal of the Acoustical Society of America* 80. 1041–1047.

Trager, George L. 1961. The intonation system of American English. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott & J. L. M. Trimm (eds.) *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday*, 266–270. London: Longmans, Green & Co. Reprinted in Bolinger (1972), 83–86.

Trager, George L. & Henry L. Smith. 1951. *An outline of English structure*. Norman, OK: Battenburg Press.

Venditti, Jennifer, Kikuo Maekawa & Mary E. Beckman. 2008. Prominence marking in the Japanese intonation system. In Shigeru Miyagawa & Mamoru Saito (eds.) *The Oxford handbook of Japanese linguistics*, 456–512. Oxford: Oxford University Press.

Ward, Gregory & Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall–rise intonation. *Language* 61. 747–776.

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 21. 55–105.

Xu, Yi. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46. 220–251.

Xu, Yi & Xuejing Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111. 1399–1413.

33 Syllable-internal Structure

Anna R. K. Bosch

> There is no simple discovery procedure for determining phonological syllable structure (which, like phonological representations in general, may not be in a one-to-one relationship with systematic phonetic syllabification, and which may not necessarily conform to native speaker intuitions about syllable division). The nature of the mechanism which assigns syllabification (defines possible syllables) for a given language is an empirical hypothesis, whose confirmation depends on the extent to which linguistically significant generalizations can be expressed under it (Feinstein 1979: 255).

1 Introduction

Although the syllable has been used by generations of linguists both in language description and in phonological theory, it is still surprising to see the variety of opinions and arguments on the topic. As Einar Haugen expostulated as early as 1956, "everyone talks about syllables, but no one seems to do anything about defining them" (Haugen 1956: 196). Since that time we have had fifty years or more of attempts to outline and define the syllable and its constituents, perhaps without coming much closer to solid agreement. Although the terms syllable, onset, rhyme, nucleus, and coda remain in common usage among phonologists, we cannot yet point to invariant acoustic or articulatory evidence for these constituents. As Haugen continues, "the only real basis for assuming their existence is that speakers of the language can utter them separately, dividing utterances into sequences that seem natural when pronounced alone." We point to this evidence again and again as certain proof that there is "something" called the syllable; perhaps the everyday linguistic knowledge of the speaker is the single constant throughout phonological research on syllable structure. And yet, as Feinstein emphasizes in the quote appended above, we also draw a certain distinction between the "speaker's syllable" and the "phonological syllable." The phonological syllable is defined empirically by "linguistically significant generalizations," while the speaker's syllable is defined simply and automatically (when the decision is indeed simple and automatic) in careful speech, or by a number of – also empirical – experimental methods exploring external evidence, such as language games and other speaker behavior (CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY).

As Goldsmith (1990) points out, there are at least two competing, or perhaps parallel, views of the syllable that have influenced phonological theory over the past century or more: just as light can behave as both wave and particle, the syllable too has been shown to demonstrate both a wave-like property based on sonority (CHAPTER 49: SONORITY), and a piece-like, or chunk-like, division into smaller constituents, such as onset and rhyme. While there are still good reasons to hold on to a wave-shaped understanding of the syllable, defining the syllable and its properties with reference to the peaks and valleys of sonority shaping each one, the present chapter will nonetheless focus exclusively on theories of the syllable which specifically address the question of constituent structure. The question of syllable structure can be understood as a question about the nature of linguistic representation (cf. Anderson 1985); as Anderson claims, "most of the history of twentieth-century phonology is the history of theories of representations, devoted to questions such as 'What is the nature of the phoneme, morphophoneme, morpheme, etc?' " (Anderson 1985: 9), to which we might add the question that concerns us here: "What is the nature of the syllable?" This chapter begins with a brief historical sketch tracing the early arguments in favor of different representations of syllable structure, followed by an overview of different models of the internal structure of the syllable. A final section reviews the conclusions of experimental studies as they adduce evidence for or against internal constituents of the syllable. (See also CHAPTER 109: POLISH SYLLABLE STRUCTURE; CHAPTER 115: CHINESE SYLLABLE STRUCTURE; CHAPTER 56: SIGN SYLLABLES.)

2 Early twentieth-century discussions of syllable-internal structure

Early twentieth-century linguists asked themselves the same question, "what is the nature of the syllable?" Saussure proposes an impressionistic account of the syllable as composed of a succession of explosive and implosive articulatory movements; while all speech consists of an alternating series of these movements, the syllable boundary itself is marked by "the passage from an implosion to an explosion in a chain of sounds" (Saussure 1922). Arguing for the syllable as a unit of phonology, he reasons that "the regular coincidence of a mechanical principle and a definite acoustical effect assures the implosive-explosive combination of a right to existence in phonology" (1922: 57). Saussure goes so far as to claim that these opening and closing articulatory motions, which he distinguishes from acoustic sonority, are themselves the irreducible units of the syllable; further, a close examination of Saussure's diagrams demonstrates that he considers the vocalic peak to form a part of the implosion, hinting at something like a rhyme: "Whenever a particular phoneme is more open than the following one, the impression of continuity persists; . . . an implosive link, like an explosive one, obviously can include more than two elements if each has wider aperture than the following one" (Saussure 1922: 56). Saussure's footnote demonstrating the syllabification of the word *particularly* is noteworthy, presenting an early sketch of an onset-rhyme division within the syllable: a vowel and following tautosyllabic consonant are both "implosive," according to Saussure's terminology, while the prevocalic consonants are "explosive." (Superscript arrow-heads indicating explosion [<] and implosion [>] are placed immediately above each alphabetic graph in Saussure's text.)

(1) < > > < >> < > < > > < >
[p a r t i k i u l a r l i]

Kuryłowicz develops a notion of syllable structure which was clearly influenced by Saussure, referring to "the initial (explosive) consonant group and the final (implosive) consonant group as they relate to the vocalic center"[1] (Kurylowicz 1949). Here Kuryłowicz quite explicitly associates the structure of both "semantic" and "phonic" systems, presenting a "table of correspondence," or equivalence, which presages hierarchical structure within the syllable, as in (2), in parallel with the semantic functions of subject, predicate, etc.

(2) *Table of correspondence* (Kuryłowicz 1949, reprinted in Hamp *et al.* 1966: 230, translated A. R. K. Bosch)

semantic system	*phonic system*
proposition	syllable
predicate	vowel
subject	initial consonant group
additional information	final consonant group, etc.

That is, just as a proposition consists of subject, predicate, and additional information, a syllable can be seen to consist of initial consonant group, vowel, and final consonant group. Peak and coda are grouped into a constituent in Kuryłowicz (1948), on the basis of co-occurrence restrictions which are found between peak and coda, but not between onset and peak.

Although the term "vocalic peak" is already in use by the time of Saussure's writing, Selkirk (1982) credits Hockett (1955) with the terms "onset" and "coda"; the use of "rhyme" to refer to the conjunction of peak and coda is attributed to Fudge (1969), although of course the informal usage of this term to describe poetic form dates from the seventeenth century.[2]

3 Evidence for constituents within the syllable

Reviewing arguments for syllable-internal structure from Saussure onward, we find that evidence for structure within the syllable is typically modeled on evidence for the syllable itself; in a comprehensive overview of syllable theories, Blevins (1995) outlines four traditional arguments in favor of the syllable itself as a constituent. She notes that (a) the syllable has been employed as the domain within which phonological processes or constraints may apply; (b) the syllable edge is identified as the locus for the application of processes or constraints; (c) the syllable itself may be picked out as a "target structure," e.g. for the application of language games or for the assignment of stress or tone; and finally (d) field linguists recount that native speakers can express intuitions regarding

[1] "le groupe consonantique initial (explosif) et le groupe consonantique final (implosif) par rapport au centre vocalique" [translation A. R. K. Bosch].

[2] The *Oxford English Dictionary* cites Samuel Butler (1663), "For Rhime the Rudder is of Verses, With which like Ships they stear their courses."

the number of syllables per word or utterance. So, for example, nasalization may spread within a syllable (a); a syllable-final consonant may be devoiced (b); syllables may be independently manipulated in language games such as the French Verlan (c); finally, field linguists commonly report formal and informal studies of speakers who are easily able to count syllables, or who pause between syllables when exaggerating slow and careful speech (d).

All these are common examples of the utility of the constituent "syllable"; however, not all of these arguments provide evidence for sub-syllabic constituents. Upon closer examination, only (a) and (c) usefully apply in evaluating syllable-internal constituents. First, as argued in (a), the constituents onset or rhyme have been argued to serve as phonological domains: Davis (1992) argues from Italian that the choice of the article (*il* or *lo*) depends on the constituent structure of the following onset (CHAPTER 55: ONSETS; CHAPTER 38: THE REPRESENTATION OF SC CLUSTERS). And the constituent structure of the rhyme – short vowel, long vowel, or vowel + consonant – may play a crucial role in stress assignment in quantity-sensitive languages. In addition (as argued in (c)), the separate constituents onset and rhyme may be singled out as "target structures" for the application of language games. In the American children's game "ubby dubby," popularized by the 1970s television show *Zoom!*, the sequence [əb] is inserted between each onset and rhyme: *hello* becomes [həbɛləbou]. Numerous language games play on the identification of onset and rhyme as target structures, and studies suggest that speech errors may operate on onset and rhyme sequences as single units (see discussion of experimental evidence, below). Thus evidence for sub-syllabic constituents derives primarily from data suggesting that onset and rhyme function as phonological domains or as target structures for other linguistic behavior. However, when we return to examine common evidence for the syllable as a constituent, parallel arguments for syllable-internal structure are not as convincing. Evidence such as (b) that refers to syllable-edges as targets (e.g. devoicing a syllable-final obstruent; CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION) in fact says nothing about syllable-internal structure per se: a syllable boundary, without reference to constituent structure, could identify this position (see Steriade 1999, for example). And finally, evidence in (d) from slow or careful speech by native speakers may provide clues as to syllable count or syllable boundaries, but generally provides little insight into sub-syllabic constituents, without additional manipulations such as we find in studies of language games, etc.

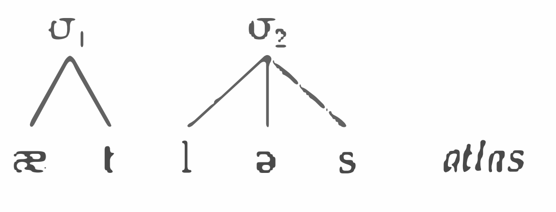
Nonetheless, we frequently uncover parallels between discussions of syllable structure and discussions of syllable-internal structure. This is made explicit within the framework of prosodic phonology (Selkirk 1982; Nespor and Vogel 1986): here the internal structure of the syllable is seen as a natural extension of the higher-level prosodic structure, to which the syllable naturally belongs (see CHAPTER 40: THE FOOT; CHAPTER 51: THE PHONOLOGICAL WORD; CHAPTER 84: CLITICS; CHAPTER 50: TONAL ALIGNMENT). After laying out arguments for the syllable as a constituent, Selkirk goes on to conclude that:

> The same three reasons leading to the postulation of the syllable can be shown to motivate the existence of privileged groupings of segments within the syllable which must be thought of as constituent-like linguistic units themselves . . . an internally-structured tree quite analogous to a tree representing syntactic structure (Selkirk 1982: 237).

3.1 "Flat" models of syllable-internal structure

Various models of syllable-internal structure have been proposed over the past century of linguistic study, from an entirely flat structure consisting primarily of syllable boundary markers to a more highly articulated hierarchical structure. Kahn (1980), for example, proposes the simplest two-tier flat structure consisting of syllable nodes (σ_1, σ_2, etc.) on one tier, associated directly with the segments of phonetic (or phonological) representation, as in (3).

(3) *No internal constituent structure* (e.g. Kahn 1980)



For Kahn, the discrete segments are "associated" with the syllable node, and among his syllable-building principles is one akin to the no-crossing constraint of Goldsmith's (1976) autosegmental phonology; "given the way the term 'syllable' is understood, it would seem nonsensical to speak of discontinuous syllables" (Kahn 1980: 36). Kahn explicitly cites Goldsmith manuscripts from 1974 and 1975, and in a footnote outlines his claim that he himself is working in an autosegmental framework, "because all theories of the syllable, including my own, are 'autosegmental' in that they involve parallel analyses of phonological material into (traditional) segments and syllables" (Kahn 1980: 61; see also CHAPTER 14: AUTOSEGMENTS).

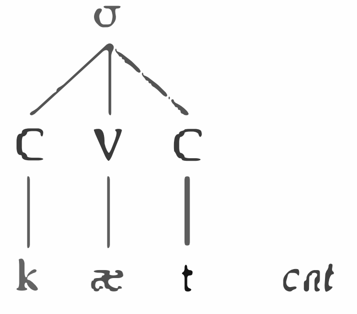
A contemporary version of Kahn's flat structure is echoed in a recent textbook, Hayes (2009). Hayes takes a non-committal position on the internal structure of the syllable: while he prefers constituent (tree) structure to the simple use of boundary symbols to identify syllables, he makes no claim about constituency within the syllable. Introducing the terms onset, coda, and nucleus, he explains that

> In some theories, the onset, nucleus, and coda are described as constituents (they are daughters of the syllable node σ, and dominate segments). This book will use "onset," "nucleus," and "coda" merely as useful descriptive terminology (Hayes 2009: 251).

In diagrams throughout this textbook, as in Kahn (1980), segments are dominated directly by the syllable node itself, without intervening structure.

A related flat structure with an intervening CV tier is proposed by Clements and Keyser (1983) in (4) (see also CHAPTER 54: THE SKELETON).

(4) *Syllable with intervening* CV *tier* (e.g. Clements and Keyser 1983)



Among the options that do incorporate some representation of internal structure, however, ternary branching structure represents perhaps the simplest

This is essentially what Venneman (1972) terms the "Law of Initials": that "medial syllable-initial clusters should be possible word-initial clusters" (Venneman 1972: 11); or what Bell (1976) terms "the Kuryłowicz condition," that "initial and final clusters of medial syllables conform to the same constraints as those in initial and final syllables" (Bell 1976: 255).[6]

On the other hand, there are languages which place more restrictions on word-internal consonant clusters, and allow extra consonants at word edges, such as Polish (Kenstowicz 1994: 262ff.). A number of chapters in Féry and van de Vijver (2003) tease out just these inconsistencies between word-internal clusters and clusters found in either word-initial or word-final position (Cho and King 2003; Green 2003; Kiparsky 2003; Wiltshire 2003). See also Dixon (1970) and many others, as well as CHAPTER 36: FINAL CONSONANTS.

3.4 *Moraic phonology and syllable-internal structure*

While the syllable itself has been in wide use throughout the past century or more, clear arguments providing evidence of the internal structure of the syllable are less common than one might expect. Many arguments which state a convincing case for the syllable as a phonological unit in fact fail to motivate syllable-internal structure. The constituent structure of onset, rhyme, nucleus, and coda intersects in complex ways with a moraic theory of syllable organization, as sketched in (14) below. While a moraic analysis often appears to supersede one employing syllable structure, in fact the notion of moraic weight is interwoven with an understanding of syllable structure, in particular the structure of the rhyme.

(14) *The mora in the prosodic hierarchy*

Prosodic Word
|
Foot
|
Syllable
|
Mora

The mora provides a useful means of representing syllable weight, in quantity-sensitive languages where this is required. In languages such as English or Latin, a syllable with a short vowel is monomoraic, while syllables with a long vowel (VV), or vowel + coda consonant (VC), are bimoraic. We note, however, that only consonants in the rhyme may be moraic; onset consonants never contribute to

[6] Recall that in this article Bell argues against "the distributional syllable." Nevertheless, he does not entirely conclude that phonology can do without the syllable: "Let us, however, guard against too narrow a view, against confusing a tool with the problem. 'Defining the syllable' and 'proving the existence of the syllable' are pseudo-problems. Segment organization is the problem. If assumption of a syllabic unit leads to explanation of regularities of segment organization, so much the better. If not, we will be awaiting a more general theory of organization, and the syllable may enter the museum's Hall of Scientific Constructs, taking its place beside ether, the noble savage, and the like" (Bell 1976: 261).

syllable weight (CHAPTER 47: INITIAL GEMINATES; CHAPTER 55 ONSETS). Thus we return to some notion of constituent structure within the syllable, if only to identify the domain in which moras are projected. Furthermore, whether the mora truly serves as a "constituent" is unclear; while the syllabic nucleus is typically affiliated with a mora, the affiliations of onset and non-moraic coda consonants are less clear. Non-moraic elements are sometimes associated with the syllable node directly, or sometimes argued to share the mora with the nuclear vowel. Essays in Ziolkowski *et al.* (1990) demonstrate the range of arguments regarding moraic structure within the syllable. Hyman (1985) originally suggested that a syllable-initial consonant links to the mora of the following vowel, creating what looks like a "body–coda" structure. More commonly, the syllable-initial consonant is assumed to associate directly to the syllable node, as in Hayes (1989) and many others. For a more recent position favoring the mora, Yip argues explicitly that the evidence in favor of the constituents onset and rhyme "is scanty and inconsistent" (Yip 2003: 779), and relies on a moraic model of the syllable to account for the behavior of pre-nuclear glides in English and Mandarin Chinese. I leave it to other contributors to this *Companion* to tease out the intricacies of moraic phonology in more detail (see CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE; CHAPTER 40: THE FOOT; CHAPTER 41: THE REPRESENTATION OF WORD STRESS).

4 Experimental studies

Experimental studies, both of acoustic properties of speech and of human behavioral responses to syllabification tasks, have also been constructed to explore this question of the nature of the syllable. The great majority of experimental work over the past fifty years involves studies of simple syllabification, with a view to accounting for syllable boundaries. Most of these studies place a particular focus on the syllabification of an intervocalic consonant or consonants; see for example studies on Dutch, Finnish, German, French, Japanese, and English (e.g. Fallows 1981; Gillis and DeSchutter 1996; Schiller *et al.* 1997; Berg and Niemi 2000; Content *et al.* 2001; Goslin and Frauenfelder 2001; Ishikawa 2002; Redford and Randall 2005). Still, a number of experimental studies have been adduced to test the validity of the internal constituents of the syllable; the majority of these studies focus on the primary constituents of onset and rhyme. Evidence regarding hierarchical structure within the syllable is mixed, with arguments drawn from language games (both "traditional" and invented/experimental), slips of the tongue, perceptual studies, investigations with children, and other experimental paradigms (CHAPTER 96: EXPERIMENTAL APPROACHES IN THEORETICAL PHONOLOGY) to investigate whether onset, rhyme, nucleus, and coda are or are not syllable constituents.

A series of experiments by Treiman and co-authors argue in favor of an onset–rhyme structure, based on subjects' performance on various word games (Treiman 1986; Fowler *et al.* 1993; Treiman *et al.* 1994; Treiman *et al.* 1995). Most recently, Kapatsinski (2009) claims to show from an experimental study that English speakers are able to learn rhyme–affix associations more easily than body–affix associations, basing this argument on the claim that associations should be easier to learn within rather than across constituents, given a hierarchical structure to the syllable. However, Pierrehumbert and Nair (1995) replicate Treiman's word game paradigm merely substitutions of segmental units – errors in "selection" – but instead may be examples of gestural intrusion or mis-timing (Pouplier and Goldstein 2005).

Various studies conducted on languages other than English appear to show that if there is an onset–rhyme distinction it may be a language-specific one; experiments with native speakers of Korean indicate that Korean syllables "contain a cohesive CV or body unit, in contrast to the VC or rhyme unit of English" (Yoon and Derwing 2001); one of the five experiments described here studied preliterate children, indicating that literacy could not be a confounding factor. A study involving Chinese–English bilingual children found "a preference for matching body over rime in Chinese, and for matching rime over body in English," concluding that there must be cross-language differences in processing spoken syllables (Wang and Cheng 2008).

Acoustic studies have examined timing relationships within the syllable to identify syllable constituents. Conducting an acoustic study on English disyllables and casual speech vowel reduction, in triplets of words such as *support/sport/s'pport* (reduced *support*) Fokes and Bond (1993) conclude that there were no invariant acoustic cues determining syllabicity. While the authors concede that *sport* and *s'pport* may in fact be phonetically distinct, the study found no invariant cues to distinguish them. Certainly, attempts to isolate acoustic or articulatory invariants of the syllable date from as early as Stetson's (1928) "chest pulse" theory; however, there is no current consensus on either acoustic or articulatory definitions of the syllable, let alone of structure internal to the syllable. While Selkirk (1982: 340) set the stage for a good deal of ensuing research with her note that "other phonological, or shall we say phonetic, phenomena such as duration and closeness of transition between segments might also be taken as revealing of the immediate constituent structure of the syllable," we still find very little clear evidence of any invariant property pointing to syllable-internal hierarchical structure.

5 Conclusion

Despite the lack of phonetic evidence for invariant acoustic or articulatory measures of syllable structure, research in this area too continues apace. As Ladefoged noted:

> There is no single muscular gesture marking each syllable . . . (but) there is evidence . . . that speakers organize the sequences of complex muscular events that make up utterances in terms of a hierarchy of units, one of which is the size of a syllable; and it is certainly true that speakers usually know how many syllables there are in an utterance. We will therefore assume that a neurophysiological definition is possible, even if one cannot at the moment state it in any way (Ladefoged 1971: 81).

Even those who argue against the use of syllable structure to account for phonotactics acknowledge the usefulness of the terms referring to syllable-internal structure: onset, peak, coda, and even rhyme. "Syllable structure, whether directly perceived or inferred, is an undeniable aspect of phonological representations," claims Steriade (1999), although she goes on to argue against employing syllable constituents in a phonological analysis, concluding that syllable position "does not condition segment realization." Steriade argues instead that knowledge of syllable structure, and syllable edges in particular, derives from or is founded on the speaker's perception of word-based phonotactic regularities. Her claim is essentially that we have put the cart before the horse in arguing that phonotactic constraints are built upon syllable structure; instead, these phonotactic regularities may be precisely what allow us to identify syllable position.

In any case, the labels we use to identify internal constituents of the syllable – onset, coda, and rhyme – remain convenient terminology, and seem likely to remain in common usage. Nevertheless, it also seems clear that a conservative view of linguistic structure – a view shaped by Occam's razor, perhaps – would concede that these terms, while useful, may not be supported by empirical evidence. Acoustic and experimental studies offer only mixed results, while language-specific phonological studies continue to differ widely in their use of (and claims for) some particular organization of syllable-internal structure. How we use syllable structure to represent the patterns and organization of human language will differ depending on the questions we ask and the problems we confront in the specific languages we investigate. Syllable structure may turn out to be an organizational tool, rather than an object available for independent manipulation.

REFERENCES

Álvarez, Carlos, Manuel Carreiras & Manuel Perea. 2004. Are syllables phonological units in visual word recognition? *Language and Cognitive Processes* 19. 427–452.

Anderson, Stephen R. 1985. *Phonology in the twentieth century.* Chicago: University of Chicago Press.

Barlow, Jessica. 2001. Individual differences in the production of initial consonant sequences in Pig Latin. *Lingua* 111. 667–696.

Bell, Alan. 1976. The distributional syllable. In Alphonse Juilland (ed.) *Linguistic studies offered to Joseph Greenberg*, vol. 2, 249–262. Saratoga, CA: Anma Libri.

Berg, Thomas & Jussi Niemi. 2000. Syllabification in Finnish and German: Onset filling vs. onset maximization. *Journal of Phonetics* 28. 187–216.

Blevins, Juliette. 1995. The syllable in phonological theory. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 206–244. Cambridge, MA & Oxford: Blackwell.

Botma, Bert, Colin J. Ewen & Erik Jan van der Torre. 2008. The syllabic affiliation of postvocalic liquids: An onset-specifier approach. *Lingua* 118. 1250–1270.

Cho, Young-mee Yu & Tracy Holloway King. 2003. Semisyllables and universal syllabification. In Féry & van de Vijver (2003), 183–212.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable.* Cambridge, MA: MIT Press.

Content, Alain, Ruth K. Kearns & Uli H. Frauenfelder. 2001. Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language* 45. 177–199.

Davis, Stuart. 1989. On a non-argument for the rhyme. *Journal of Linguistics* 25. 211–217.

Davis, Stuart. 1992. The onset as a constituent of the syllable: Evidence from Italian. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26. 71–79.

Derwing, Bruce L. 1992. A pause-break task for eliciting syllable boundary judgments from literate and illiterate speakers: Preliminary results for 5 diverse languages. *Language and Speech* 35. 219–235.

Dixon, R. M. W. 1970. Olgolo syllable structure and what they are doing about it. *Linguistic Inquiry* 1. 273–276.

Fallows, Deborah. 1981. Experimental evidence for English syllabification and syllable structure. *Journal of Linguistics* 17. 309–317.

Feinstein, Mark. 1979. Prenasalization and syllable structure. *Linguistic Inquiry* 10. 245–278.
Féry, Caroline & Ruben van de Vijver (eds.) 2003. *The syllable in Optimality Theory*. Cambridge: Cambridge University Press.
Firth, J. R. & B. B. Rogers 1937. The structure of the Chinese monosyllable in a Hunanese dialect. *Bulletin of the School of Oriental Studies* 8. 1055–1074.
Fischer-Jørgensen, Eli. 1952. On the definition of phoneme categories on a distributional basis. *Acta Linguistica* 7. 8–39.
Fokes, Joann & Z. S. Bond. 1993. The elusive/illusive syllable. *Phonetica* 50. 102–123.
Fowler, Carol A., Rebecca Treiman & Jennifer Gross. 1993. The structure of English syllables and polysyllables. *Journal of Memory and Language* 32. 115–140.
Fromkin, Victoria A. (ed.) 1973. *Speech errors as linguistic evidence*. The Hague: Mouton.
Fudge, Erik C. 1969. Syllables. *Journal of Linguistics* 5. 253–286.
Fudge, Erik C. 1987. Branching structure within the syllable. *Journal of Linguistics* 23. 359–377.
Geudens, Astrid & Dominiek Sandra. 2003. Beyond implicit phonological knowledge: No support for an onset–rime structure in children's explicit phonological awareness. *Journal of Memory and Language* 49. 157–182.
Geudens, Astrid, Dominiek Sandra & Heike Martensen. 2005. Rhyming words and onset–rime constituents: An inquiry into structural breaking points and emergent boundaries in the syllable. *Journal of Experimental Child Psychology* 92. 366–387.
Gillis, Steven & Georges DeSchutter. 1996. Intuitive syllabification: Universals and language specific constraints. *Journal of Child Language* 23. 487–514.
Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.
Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Oxford & Cambridge, MA: Blackwell.
Goslin, Jeremy & Uli H. Frauenfelder. 2001. A comparison of theoretical and human syllabification. *Language and Speech* 44. 409–436.
Green, Antony Dubach. 2003. Extrasyllabic consonants and onset well-formedness. In Féry & van de Vijver (2003), 238–253.
Hamp, Eric P., Fred W. Householder & Robert Austerlitz. 1966. *Readings in Linguistics II*. Chicago: University of Chicago Press.
Harris, John. 1994. *English sound structure*. Oxford: Blackwell.
Haugen, Einar. 1956. Syllabification in Kutenai. *International Journal of American Linguistics* 22. 196–201.
Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.
Hayes, Bruce. 2009. *Introductory phonology*. Malden, MA & Oxford: Wiley-Blackwell.
Hayes, Rachel A., Alan M. Slater & Christopher A. Longmore. 2009. Rhyming abilities in 9-month-olds: The role of the vowel and coda explored. *Cognitive Development* 24. 106–112.
Hindson, Barbara Anne & Brian Byrne. 1997. The status of final consonant clusters in English syllables: Evidence from children. *Journal of Experimental Child Psychology* 64. 119–136.
Hoard, James E. 1971. Aspiration, tenseness, and syllabication in English. *Language* 47. 133–140.
Hockett, Charles F. 1947. Peiping phonology. *Journal of the American Oriental Society* 67. 253–267.
Hockett, Charles F. 1955. *A manual of phonology*. Baltimore: Waverly Press.
Hooper, Joan B. 1972. The syllable in phonological theory. *Language* 48. 525–540.
Hulst, Harry van der & Nancy Ritter (eds.) 1999. *The syllable: Views and facts*. Berlin & New York: Mouton de Gruyter.
Hyman, Larry M. 1985. *A theory of phonological weight*. Dordrecht: Foris.
Ishikawa, Keiichi. 2002. Syllabification of intervocalic consonants by English and Japanese speakers. *Language and Speech* 45. 355–385.
Kahn, Daniel. 1980. *Syllable-based generalizations in English phonology*. New York: Garland.

Kandel, Sonia, Lucie Hérault, Géraldine Grosjacques, Eric Lambert & Michel Fayol. 2009. Orthographic vs. phonologic syllables in handwriting production. *Cognition* 110. 440–444.

Kapatsinski, Vsevolod. 2009. Testing theories of linguistic constituency with configural learning: The case of the English syllable. *Language* 85. 248–277.

Kaye, Jonathan. 1990. "Coda" licensing. *Phonology* 7. 301–330.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1985. The internal structure of phonological elements: A theory of charm and government. *Phonology Yearbook* 2. 305–328.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology* 7. 193–231.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kiparsky, Paul. 2003. Syllables and moras in Arabic. In Féry & van de Vijver (2003), 147–182.

Kuryłowicz, Jerzy. 1948. Contribution à la théorie de la syllabe. *Bulletin de la Société Polonaise de Linguistique* 8. 80–113.

Kurylowicz, Jerzy. 1949. Linguistique et théorie du signe. *Journal de Psychologie* 42. 170–180.

Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics*. Chicago: Chicago University Press.

Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.

McCarthy, John J. 1979. On stress and syllabification. *Linguistic Inquiry* 10. 443–465.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Pierrehumbert, Janet B. & Rami Nair. 1995. Word games and syllable structure. *Language and Speech* 38. 77–114.

Pike, Kenneth L. 1975a. Suprasegmentals in reference to phonemes of item, of process, and of relation. *Bibliotheca Phonetica* 11. 45–56.

Pike, Kenneth L. 1975b. Tests for prosodic features of pitch, quantity, stress. *Bibliotheca Phonetica* 11. 4–5.

Pike, Kenneth L. & Eunice V. Pike. 1947. Immediate constituents of Mazateco syllables. *International Journal of American Linguistics* 13. 78–91.

Pouplier, Marianne. 2007. Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech* 50. 311–341.

Pouplier, Marianne. 2008. The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics* 36. 114–140.

Pouplier, Marianne & Louis Goldstein. 2005. Asymmetries in the perception of speech production errors. *Journal of Phonetics* 33. 47–75.

Rand, Earl. 1968. The structural phonology of Alabaman, a Muskogean language. *International Journal of American Linguistics* 34. 94–103.

Redford, Melissa A. & Patrick Randall. 2005. The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics* 33. 27–46.

Saussure, Ferdinand de. 1922. *A course in general linguistics*. New York: Philosophical Library.

Scheer, Tobias. 2004. *A lateral theory of phonology*, vol. 1: *What is CVCV, and why should it be?* Berlin & New York: Mouton de Gruyter.

Schiller, Niels O., Antje S. Meyer & Willem J. Levelt. 1997. The syllabic structure of spoken words: Evidence from the syllabification of intervocalic consonants. *Language and Speech* 40. 103–140.

Selkirk, Elisabeth. 1982. The syllable. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part II, 337–383. Dordrecht: Foris.

Steriade, Donca. 1999. Alternatives to syllable-based accounts of consonantal phonotactics. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.) *Item order in language and speech*, 205–245. Prague: Karolinum Press.

Stetson, Raymond H. 1928. *Motor phonetics*. Haarlem: Société Hollandaise de Science.

Treiman, Rebecca. 1983. The structure of spoken syllables: Evidence from novel word games. *Cognition* 15. 49–74.

34 Precedence Relations in Phonology

Charles Cairns
Eric Raimy

1 Introduction

"Precedence" in phonology refers to the fact that elements occur in ordered sequences. An understanding of precedence relations is key to explicating notions of locality, adjacency, and left–right asymmetries, which have played significant roles in the phonological literature, especially since the advent of autosegmental phonology (Goldsmith 1976; CHAPTER 14: AUTOSEGMENTS). McCarthy (1989: 71) writes "nonlinear phonology imposes strict requirements of locality on phonological rules," and "locality … ensures that the elements referred to in phonological transformations and constraints are adjacent at some level of representation."

We must agree on a common set of questions in order to compare and contrast linguists' notions of precedence relations, a requirement that presupposes a formal framework. Since the earliest days of phonology, the sequential nature of speech and the progression of letters across the printed page have been assumed to be sufficient to understand precedence. It is always salutary to explicate tacit assumptions, so this chapter explores the implications of a formally rigorous understanding of precedence. The formal rigor is supplied by graph theory, a branch of mathematics that we will use to unpack the question of what it means for phonemes to appear in an ordered sequence (Wilson 1996 is one of several good introductions). Phonology is concerned with the characteristics of precedence in human language, so graph theory itself can be no more than a useful tool. But because graph theory is an explicit mathematical model that provides specific and well-understood possible answers to questions of precedence, we explore its implications in this chapter.

Once we have introduced the relevant aspects of graph theory, we go on to examine a sample of the claims and assumptions that have been made in the literature on phonological precedence, with varying degrees of explicitness. In particular, we examine the characterizations of various approaches to autosegmental phonology within graph theory. We will further explicate the nature of precedence in phonology by considering the basic operation of deletion. All phonologists must take as bedrock assumptions that phonemes appear in a sequence and that there exist phonological processes with the capacity to delete phonemes from a sequence. One interesting result of this exercise is that commonly assumed antagonistic theories of phonology converge on a common model of deletion.

Before proceeding, we first ask if precedence relations are primitives of phonological representations or if they are derived. Van der Hulst (2008), for example, proposes that precedence relations can be derived from underlying syllables. Consider a highly articulated theory of the syllable with labeled nodes, e.g. Fudge (1969, 1987); Cairns (1988); see also CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE. The idea is that if at the lexical level featural information such as [back] or [coronal] were stored in syllabic nodes like onset and rhyme, then perhaps the number and order of phonemes on the surface could be predicted from the inherent order of syllabic constituents. There are at least four natural limits to deriving precedence relations from syllable structure. First, syllabification is not always exhaustive; many languages are known to have sequences of unsyllabified consonants (Bagemihl 1991; Czaykowska-Higgins and Willett 1997; Vaux and Wolfe 2009). Second, it would in any case be necessary to specify the order among syllables. Consider the English loan from Tamil *catamaran*; hypothetical **matacaran* and **tamacaran* would serve as equally plausible loans into English, yet they differ from the existing word only by syllable order. The possibility of using foot structure to order syllables only moves this question higher in the prosodic hierarchy and requires more prosodic information to be stored in the lexicon. The necessity of stipulating explicit sequencing information in the lexicon cannot be escaped.

Third, reference to precedence relations at the segmental level is necessary to properly account for phonotactic constraints. Blevins (2003), building on Steriade (1999), shows numerous compelling examples where phonological sequencing generalizations and cross-linguistic universal patterns refer to properties of the phonological string and not to syllable structure. Of course, as Blevins points out, there are many cases "where phonotactic constraints and syllable structure appear to converge." She suggests that "this is because syllabifications are derivative of phonotactics, not vice versa" (2003: 393).

Finally, we are encouraged in our focus on segmental precedence by the fact that resyllabification is rampant throughout phonology and phonetics; the effervescence of the syllable makes it a poor candidate for the bearer of lexical precedence relations.

§2 sketches the basic elements of graph theory and their application to phonology. §3 presents the complications that arise in connection with considering precedence relations in autosegmental phonology and provides a plausible explanation of why the Obligatory Contour Principle (OCP) is so variably valid. §4 demonstrates how investigating the process of deletion in graph theory illuminates how different theoretical models of phonology converge on the same understanding of deletion in phonology. §5 demonstrates how theories of phonology that view segments as entities that occur in real time can benefit from the consideration of precedence in graph-theoretic terms. §6 shows how graph theory can illuminate issues in phonology such as the "no crossing constraint" and local *vs.* long-distance adjacency. §7 illustrates some extensions of graph theory approaches to different phonological and morphological phenomena. §8 concludes this chapter.

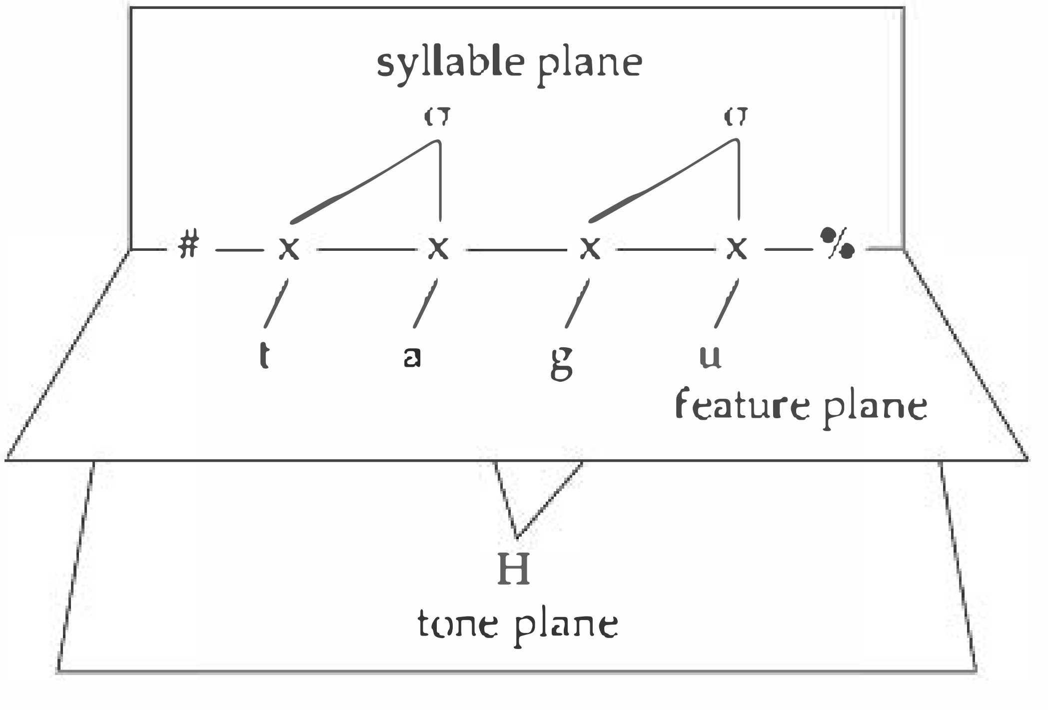
2 Precedence relations and graph theory

Phonologists differ about whether to view a phonological sequence as a string of discrete, point-like objects or as a series of possibly overlapping segments which exist in real time; Trubetzkoy (1939) described phonemes as essentially timeless (*zeitlos*), abstract entities like the dits and dots of Morse code, where each unit is unaffected by the producing process (although possibly influenced by proximal symbols). The alternative view, that phonological patterning is defined by the dynamic articulation of actual speech, goes back to Sweet (1877), Sievers (1881), and Saussure (1916). This serves as the fundamental assumption of such schools of thought as Articulatory Phonology (Browman and Goldstein 1986, 1989, 1990a, 1990b; CHAPTER 5: THE ATOMS OF PHONOLOGICAL REPRESENTATIONS), which views the elements that appear in phonological sequences to be events in real time.

The consensus that phonetic sequences involve continuous, overlapping elements does not extend to more abstract levels of phonology. Theories of morphological operations such as reduplication and infixation and of many phonological processes such as morphophonemics, deletion, etc., generally operate on representations made up of discrete, point-like objects, and these theories have achieved considerable descriptive success; we will proceed on this basis and turn to the event-based outlook later.

Consider the representation of the Margi word /tágú/ 'horse' (from Kenstowicz 1994: 312) in a generic version of autosegmental phonology in (1). This is a picture of what Coleman and Local (1991: 309) call "paddlewheel graphs." It is drawn so as to induce the reader to visualize three half planes, each emanating from a common line. The three-dimensional metaphor is useful because it helps us break down general questions of precedence into smaller and hence more manageable ones. This imagery can be misleading, however, because, as Coleman and Local (1991) demonstrate, autosegmental representations are not necessarily three-dimensional (i.e. non-planar, as defined in §6) in a mathematical sense.

(1) *Paddlewheel graph of Margi* /tágú/



Lines are referred to as "tiers" in autosegmental phonology, and the line defined by the intersection of the three planes depicted in (1) is the "anchor tier." The anchor tier contains the string of symbols # x x x x %, where # and % indicate the beginning and end, respectively, of the phoneme sequence (we ignore these now, but return to them below).

Elements are also arrayed in tiers on the "feature plane," the "syllable plane," and the "tone plane" of (1). The feature plane is shown with the alphabetic symbols *t, a, g, u* in lieu of the familiar feature trees (we employ this notation throughout the remainder of the chapter); in §3 we will see that this is more complicated than

(2) a. t — a — g — u

b. vertices: {t, a, g, u}

edges (unordered): {t a, a g, g u}

(2a) and (2b) contain the same information. (2a) depicts the graph in a visual manner, while (2b) defines the graph as a list of vertices and a list of edges. Each edge is defined by a pair of vertices, and these pairs are set off typographically from each other by commas in (2b). A list of edges suffices for many graphs, because the vertices can be determined from the list of edges. Note that the lists of vertices and edges are literally lists and not sets.

Because the edges in (2b) are undirected, it is called an undirected graph, where the two vertices that define each edge are unordered with respect to each other; for example, the edge {t a} is equivalent to {a t}. Because this type of graph contains information only about adjacency of vertices and does not specify any order, it appears to be a poor candidate for a model to represent phoneme strings. For one thing, lexical representations must contain ordering information. Consider the existence of pairs of words like *cat* and *tack* in English. The adjacency pairs (undirected edges) for these two words are identical; both words have the edge set {t æ, æ k}.

Beyond the obvious fact that ordering is distinctive is the observation made by de Lacy (2007) that if phonological graphs were undirected, we would predict that mirror-image rules or constraints like those that appeared in early versions of *SPE*-type phonology would be commonplace. An example would be rules of the form $x \rightarrow y$ / {z __, __ z}, as suggested by Bach (1968), Langacker (1969), and Anderson (1974). This means that x is rewritten y if it either precedes or follows z. If we were to adopt undirected graphs as the representation for phonological forms we would predict, contrary to fact, that mirror operations should be the most common operations found in phonology.

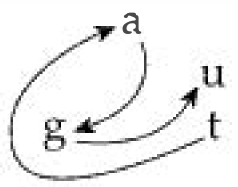
Because a theory containing only undirected edges cannot distinguish between *cat* and *tack* and makes false predictions about the directionality of phonological operations, we consider graphs where the vertices that specify each edge are ordered with respect to each other. Information about sequential order can be added to the graph in (2) by making it a "directed graph" (or "digraph"), as in (3). An edge in a digraph is an ordered pair of vertices; the vertex mentioned first in an edge specification precedes the one mentioned second.

(3) a. t → a → g → u

b. vertices: {t, a, g, u}

edges (ordered): {t a, a g, g u}

c.



(3a) is the graphic representation of the information in (3b). Because the edges in (3b) are ordered, they specify that *t* precedes *a*, *a* precedes *g*, and *g* precedes *u*. (3a) is not the only diagram consistent with (3b), however; for example, the diagram in (3c) is equally consistent with (3b), but not as convenient to read as (3a). They are formally equivalent; in fact, (3a), (3b), and (3c) all represent the same *g* or *u*, so in this sense it is true that if *t* precedes *a* and *a* precedes *g*, then *t* precedes *g*.

Before concluding this section, note that it is convenient to add explicit beginning and end symbols, which we represent as # and %, respectively. Positing these symbols allows us to say that all vertices on the anchor tier that are of degree two are available for serving as phonological segments, and only the abstract terminal symbols are of degree one. These symbols are also convenient to define the environments of initial and final segments, needed by many phonological processes. Also, chain graphs supplemented with special beginning and end symbols allow for the definition of free *vs.* bound morphemes: a free morpheme like *tagu* is a chain graph with # and % at beginning and end; a bound morpheme would lack one or both of these symbols. The graph in (4) is the type we will use to explicate precedence issues when we consider deletion in §4.

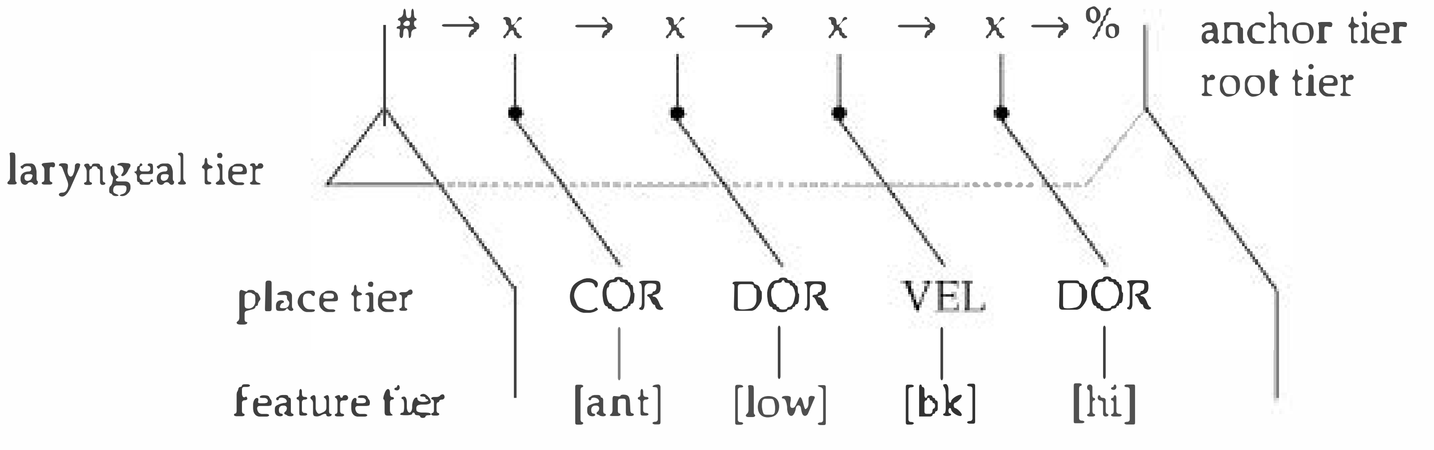
(4) a. # → t → a → g → u → %
b. vertices: {#, t, a, g, u, %}
edges (ordered): {# t, t a, a g, g u, u %}

This section has described the basic principles of graph theory as they apply to notions of phonological representations held by virtually all schools of thought in phonology. The overwhelming majority of phonologists operate on the assumptions, usually tacit, that sequences of phonemes have the properties of asymmetry, irreflexivity, and (in the qualified senses explicated above) transitivity, and the preceding paragraphs have shown that these are best understood in graph theory, where they derive from the properties of directed chain graphs. However, as stated at the beginning of this section, we have been considering only precedence relations on the anchor tier of (1). We now turn to a consideration of precedence relations in autosegmental phonology.

3 Precedence relations and autosegmental phonology

In the preceding section we adopted the expository convenience of depicting phonological content of segments on the anchor tier by means of phoneme symbols. This was a shorthand way of showing feature trees, sketched in (5), which is an elaboration of the diagram in (1). This section is devoted to analyzing the precedence relations among elements that are on different tiers, and we will see that this exercise provides insight into the OCP. A brief explanation of (5) is in order first.

(5) *Schematized autosegmental representation of* [tagu]



violate the requirement that phonological representations be equivalent to connected graphs; floating elements would be represented by vertices that are not connected to the rest of the graph. Consequently, "floating" features must be connected to the rest of a phonological representation; this seems to be consistent with current thinking on this topic (see CHAPTER 82: FEATURAL AFFIXES).

We will assume that precedence is encoded only on the anchor tier for the remainder of this chapter. This view improves our understanding of Odden's (1986, 1988) critique of the OCP as a language universal: in general, the OCP appears to behave in a cross-linguistically arbitrary way, because the locality of distinctive features is mediated by the timing tier. Consequently, whether two elements are adjacent or not must be determined either directly from the list of edges or by calculating whether a walk exists between the two elements. The implementation of a walk appears to coincide with proposals about "searches" in phonology by Mailhot and Reiss (2007), Samuels (2009), and Nevins (2010) (see §7). This perspective has the advantage that it supplies one universal account of adjacency, which has two different specifications: two elements are adjacent if they either share an edge or if there is a walk connecting them. Each language-particular process must, of course, specify which definition of adjacency is required for its application. The existence of two ways of specifying adjacency is plausibly a major reason why the OCP appears superficially to be so variably valid.

4 Deletion and precedence

Deletion is a fundamental phonological phenomenon that must be accounted for by any phonological theory. The naive view of deletion is that segments can be simply eliminated from a representation without any complicating entailments; this is based, of course, on the conception of precedence portrayed by the left to right array of symbols across the printed page. Deletion is far more complicated when we consider how precedence structures are altered when a segment is deleted. For purposes of explicating how deletion affects precedence relations, we return to using our Margi example and suppose that a phonological process of Margi were to delete the *g* from *tagu*. It does not matter what triggers this operation, nor in what phonological theory this operation is described; we are interested here in specifying precisely what it means to delete a segment when it is considered as a vertex in a directed chain graph.

In this section we describe the two characterizations of deletion (see CHAPTER 68: DELETION) that are possible within directed chain graphs. One consists of skipping over (or underparsing) the segment to be deleted, and the other involves merging two segments into one. Both appear to be empirically attested. A particularly significant part of this presentation is that two models of Optimality Theory (Containment and Correspondence) as well as the derivational model known as Precedence Based Phonology (Raimy 2000, 2009) converge on the representations and operations revealed by a graph-theoretic explication.

Without considerations about the nature of precedence, (11a) presents the naive mapping between representations undergoing the deletion of *g*, where the symbol ">" means merely "becomes," without reference to the nature of the operations involved. The question at hand is: once precedence relations are specified, how exactly does (11b) become (11c)?

All graphs representing phonological structure must be interpreted by some sort of phonetic implementation mechanism, so the question now before us is how to assure that this mechanism follows the added "*a* to *u*" precedence relation, and not the old "*a* to *g*" precedence relation. There are two general solutions to the favoring of the newly added precedence relation and they differ in whether computation in phonology is "parallel" or "derivational." "Parallel" computation is the type found in the containment model of OT, where there is a GEN function that creates a list of different candidates; there is also the EVAL function, which determines which candidate is the most harmonic given a language-specific ranking of CON. In this model, GEN would be free to create candidates that contain any number of novel precedence relations that do not exist in the input. Therefore, GEN will generate candidates with underparsed segments, and these candidates would indicate that phonetic implementation will follow the "deletion edge" and not the edges associated with the deleted *g*. Whether candidates with detours might emerge as the most harmonic will be the result of parochial constraint interactions. In short, preference for newly added precedence links that produce deletion can be implemented easily in the containment model of OT.

This approach to deletion can be straightforwardly carried over to the Correspondence model of OT (McCarthy and Prince 1995). The difference between the implementation of deletion by underparsing in the Containment and Correspondence models of OT is in the status of a representation with conflicting precedence specifications as in (13). Whereas the Containment model can directly produce this type of representation as its output with the effect that segments that are "detoured" around are the underparsed segments, the Correspondence model can produce the chain graph in (11c) directly from (11b) by deleting the vertex *g* and by adding the new edge {a u}, operations freely performed by GEN. This computation requires the comparison of both (11c) and (11b) in order to determine the MAX, DEP, and CONTIGUITY violations. Thus it appears that the Containment and Correspondence models of OT converge on deletion as the addition of a new precedence link.

Preference for newly added links and consequent deletion can be implemented in a derivational model of phonology by imposing an order on the list of edges. Recall that (13) is not a directed chain graph, a requirement for phonetic implementation. We return to this point in §7, but for now note that, as Idsardi and Raimy (forthcoming) point out, a derivational theory of phonology requires representations to have the characteristics of a directed chain graph only at the interface between modules. "Serialization" is a process that creates a walk from the beginning symbol # to the end symbol %; this walk creates a representation that is a directed chain graph. Idsardi and Shorey (2007) propose that serialization is a walk that uses the order of the list of edges to make decisions about which precedence link to follow when there is a choice. Thus, in (13b), when the walk reaches the *a* vertex, it must choose between following either the {a u} or the {a g} edge. This choice will be made by the order of the edge list, because one of these edges will be "first" and thus will be followed. In order for an added precedence link to produce deletion effects it must be added to the list of edges in a manner which places it before the precedence links associated with the segment to be deleted.

We emphasize that the detour approach to deletion is virtually the same in three theories of phonology, i.e. the Containment and Correspondence models of OT and the derivational model. The differences among the phonological theories arise only from specific details on how the deletion process is actually computed in the different models.

A second approach to deletion is also possible in graph-theoretic terms and can be understood as involving a type of coalescence, as suggested by de Lacy (2007) for OT. Another way to map a representation like (11b) to (11c) is to merge two of the vertices into one, without any addition of new precedence relations. We must first decide whether the "deleted" segment is merged with the preceding or following segment. For purposes of explication in our hypothetical example we merge the deleted *g* with the preceding *a*. The operation of vertex merger will be broken into steps in (14), so that important questions can be identified.

(14) *Deletion as vertex merger*

a. # → t → a → g → u → %
 vertices {#, t, a, g, u, %}
 edges (ordered) {# t, t a, a g, g u, u %}

b. # → t → [ag] → u → %
 vertices {#, t, [ag], u, %}
 edges (ordered) {# t, t [ag], [ag] [ag], [ag] u, u %}

c. # → t → [a] → u → %
 vertices {#, t, [a], u, %}
 edges (ordered) {# t, t [a], [a] u, u %}

(14a) presents the representation for *tagu* as in (4). Vertex merger (14b) coalesces every occurrence of *a* and *g* into a new vertex *ag*. This produces the representation in (14b), where the merger of the *a* and *g* is indicated by the [ag] composite segment. The inevitable result of this is the production of the edge {ag ag}, or the edge that loops back onto itself, thus violating the constraint requiring irreflexivity. This reflexive edge will produce a geminate or long version of the composite [ag] segment. Although this is not the desired result for plain deletion, this situation produces compensatory lengthening effects (Hayes 1989; Sloan 1991: 80–87) in a straightforward manner without recourse to moras (see §7).

The final steps that are required to produce (14c) are to eliminate the looping back arrow if it is not desired and to specify the phonetic interpretation of the vertex *ag*. There are a number of ways of accomplishing the former, one of which is to specify a parameter that allows languages to eliminate any "ref.exive" edges (edges that are defined by two mentions of the same vertex) whenever they are formed. Any theory of phonology that has the resources to separate the melodic content of a segment from its timing slot easily handles the phonetic interpretation aspect of this process. Some languages retain features of both segments in the composite vertex; for example classic coalescence in Sanskrit /a/ + /i/ > /e/. In our example, we want the segment to simply be interpreted as *a*, which will require some extra statement.

The preceding description of deletion in connection with compensatory lengthening and coalescence is theory-neutral, and will be implemented differently in different theories. The point is that considering explicit precedence relations shows how deletion, coalescence, and compensatory lengthening are deeply connected regardless of how different theories focus on the various surface effects of deletion.

One question that arises from this discussion is whether there is an embarrassment of wealth in the representational possibilities offered by graph theory to produce deletion. It appears that all the ways of accomplishing deletion are attested in different languages. For example, Tohono O'odham vowel syncope in reduplication is best characterized by the detour approach (Raimy 2000: 113–114). Chumash /l/-deletion (Raimy 1999: 82–83) exhibits deletion that is best characterized with the coalescence approach, as are examples of classic coalescence. Some languages appear to have both options available in competition with each other (Indonesian nasal assimilation; Raimy 2000: 99–112). Consequently, all of the analytical options based on different representational opportunities appear to be attested. Questions about deletion are thus typological in nature, where phonologists should ask which kind of deletion any particular instantiation in a particular language exhibits: does the type of deletion that occurs correlate with compensatory lengthening and/or coalescence processes, and what diagnostics are there to distinguish the different types of deletion?

An important theme of this section is that making precedence relations explicit via graph theory is useful to all theories of phonology. This should not be misunderstood as suggesting that there are no differences among different theories of phonology; our point is that specific differences with respect to precedence can be identified explicitly through differences in the necessary graph structures. Put another way, graph theory provides a reasonably neutral lingua franca to discuss and explore the nature of precedence in phonology. Of course, up until this point we have been discussing phonology whose elements are discrete, timeless entities; we now turn to considering the role of precedence in theories that view segments as containing real-time articulatory gestures.

5 Precedence and discrete point theory

One assumption in the preceding sections of this chapter that can and should be called into question is that phonology operates strictly on abstract discrete elements. This is the standard assumption in the class of theories of phonology that we will call formal phonology. Not all theories of phonology make this assumption. According to the theory of Articulatory Phonology (Browman and Goldstein 1986, 1989, 1990a, 1990b), vertices represent "gestures," i.e. abstract representations of dynamic articulatory events or actions. Thus the basic elements exist in real time, and are not abstract, timeless points. Such theories are also amenable to graph-theoretic representation, because they can be seen to have multiple precedence structures, one for each articulator in a segment. In fact, from a precedence structure point of view, it may be just the difference of whether there is a single abstract precedence structure or multiple more concrete articulator based precedence structures that produces an overall precedence structure that allows or disallows overlap.

Gafos (2002) argues that it is the nature of precedence that is at issue in the formal *vs.* gestural models of phonology. Gafos (2002: 270) observes that "the phonologically relevant notion of time is *overlap* of dynamic units"; this is in contrast to "linear order of static units [being] the only relevant notion of time in phonology." Both approaches still order phonological elements; they differ only in the nature of the precedence graphs. For formal phonology, the vertices are segments that are timeless points, and overlap is not definable. In gestural phonology, the vertices are gestures that have both spatial and temporal aspects, so there is a more complicated relation of overlap. Bird and Klein (1990) explicitly develop the model of precedence for phonological representations that can overlap. Another way of conceptualizing precedence in gestural phonology is to assign weight to the edges in a precedence graph to indicate the amount of overlap between each gesture; alternatively, the weights may indicate the actual time span between the gestures.

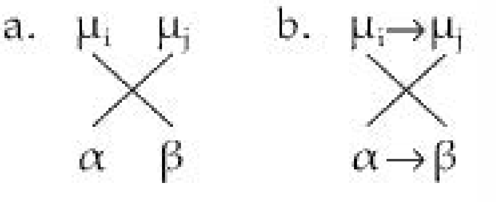
Any model of phonology can be cast by precedence graphs. Different models of phonology simply argue for or assume different characteristics for the relevant precedence graph. The different approaches to phonology – formal and gestural – are only incompatible if one assumes that either model will account for all phonological phenomena. Both models are necessary in their own domains and provide insights into different aspects of phonology. Formal phonology with atemporal segments represents phenomena closer to the morphology–phonology interface, while gestural phonology represents phenomena closer to the phonetics–phonology interface.

6 Graph theory as a tool for phonology

One major advantage of adopting graph theory as the formal underpinning of precedence in phonology is that it supplies general and specific knowledge that can be applied to uniquely phonological questions. Two topics in phonology that directly benefit from general knowledge of the nature of graphs are the no crossing constraint and locality in phonology.

The no crossing constraint originates in Goldsmith (1976: 27) as the statement "Association lines do not cross." Archangeli and Pulleyblank (1994: 39) present this constraint as a ban on the representation in (15a), which we augment with explicit precedence relations in (15b).

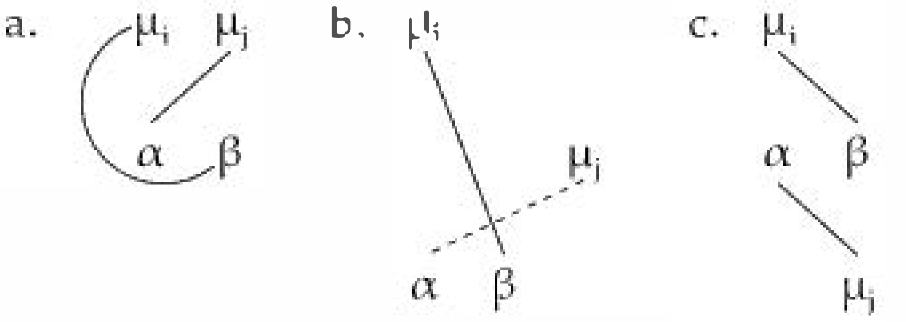
(15) *The no crossing constraint* (Archangeli and Pulleyblank 1994: 39)



The problem with (15) is that it appears to encode conflicting precedence relations. The association lines indicate that α and μ_j occur together, as do β and μ_i. The two tiers thus provide conflicting information about which feature precedes the other feature, because it encodes both $\alpha \rightarrow \beta$ and $\beta \rightarrow \alpha$. Two assumptions dictate that (15b) is ineluctably derived from (15a): that the left-to-right array of printed symbols encode directed edges and that association lines must be straight.

Coleman and Local (1991) argue that the no crossing constraint is an incoherent concept in autosegmental phonology, in part because there is no mathematical justification for insisting on straight lines; that restriction is merely a convention for drawing diagrams and has no formal content. (15a) is formally equivalent to all the diagrams in (16), which do not portray a no crossing constraint violation. (16a) is a graphic trick making an association line curved; (16b) makes the diagram appear three-dimensional, and (16c) simply moves one of the elements to the other side of a tier.

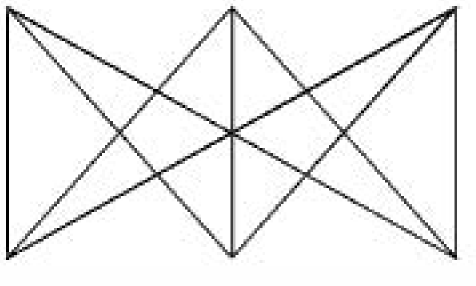
(16) *Graphic workarounds*



More importantly, Coleman and Local argue that the no crossing constraint implies that autosegmental representations must be "planar," an important question of graph theory that bears directly on when lines cross or not.

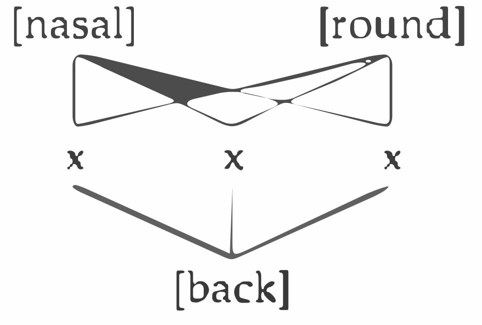
Kuratowski (1930) proves that two kinds of graph cannot be drawn within a single plane without lines crossing. One is the $K_{3,3}$ graph in (17). This is a bipartite graph, which means that the vertices can be partitioned into two sets such that each edge of the graph has one end in one set and the other in the other. The presence of a $K_{3,3}$ graph structure means that the graph which has it is "non-planar" if the lines are not to cross. According to Kuratowski's Theorem, a graph is non-planar if and only if it contains either the subgraph $K_{3,3}$ or K_5 (see Kuratowski 1930 for the K_5 graph, which does not appear to be relevant to phonology).

(17) *Kuratowski's $K_{3,3}$ graph*



Coleman and Local (1991) establish that most supposedly three-dimensional (non-planar) representations in the literature are in fact planar; on the other hand, they also point out that phonology must countenance non-planar representations. The autosegmental representation of the word *room* in Guyanese English in (18), adapted from Coleman and Local (1991: 330), is an example of a non-planar graph (i.e. it contains a $K_{3,3}$ graph); the three features [nasal], [round], and [back] are on independent tiers and each is associated with all three segments. The first x-slot represents the phoneme /r/, the second /u/, and the last /m/. Coleman and Local report that the three features depicted here spread to each of the three segments independently of one another, and therefore each should be on a separate tier. It is not possible to depict these associations in two-dimensional space without lines crossing.

(18)



The import of Coleman and Local's work is that autosegmental representations can be non-planar, which means that different distinctive features will not be able to cross lines by definition, because the association lines will be in different planes. The practical result is that phonologists must be explicit about the nature of precedence in phonological representations and distinguish between conventions on drawing convenient diagrams and formal properties of precedence.

Restrictions on precedence relations in phonology do not necessarily reside in the representations themselves but may result from how they are implemented. Graphs are "abstract data structures" (Aho *et al.* 1985) that can be implemented in many different ways. The manner of implementation affects what operations are easier (or possible) to perform than others. For example, we have so far portrayed phonological representations as a list of vertices (*x*-slots and associated features) and edges (precedence relations) as the basis for implementation. Some scholars (e.g. Heinz 2007) use adjacency tables as the basis for implementation, as exemplified for *tagu* in Table 34.1 (which is technically a local adjacency table). An adjacency table is made by listing the vertices of the graph as the headers for columns and rows and indicating in each cell whether the relevant precedence relationship holds or not.

The row headers in Table 34.1 indicate which segment precedes the column headers; Table 34.1 encodes that *t* is the word-initial segment, because no segment precedes it, indicated by the lack of any mark in the *t* column; *t* precedes *a*, as indicated by the mark in the *t* row's *a* column, *a* precedes *g*, and so on. This type of precedence encoding makes information about immediate precedence easily accessible, reflecting the frequency of phonological operations that are strictly local. A drawback to this type of precedence encoding is that long-distance relationships required to account for phenomena like long-distance assimilation and vowel harmony (see CHAPTER 77: LONG-DISTANCE ASSIMILATION OF CONSONANTS; CHAPTER 91: VOWEL HARMONY: OPAQUE AND TRANSPARENT VOWELS; CHAPTER 118: TURKISH VOWEL HARMONY; CHAPTER 123: HUNGARIAN VOWEL HARMONY) must be calculated by a walk (see §2) through the representation.

Adjacency tables can encode "transitive precedence." Table 34.2 does this by marking in each row all of the other segments that a particular segment precedes

Table 34.1 Immediate adjacency

		follows			
		t	a	g	u
precedes	t		x		
	a			x	
	g				x
	u				

Table 34.2 Transitive precedence

		follows			
		t	a	g	u
precedes	t		x	x	x
	a			x	x
	g				x
	u				

Just as there are two distinct ways to produce the deletion of a segment, there are two distinct ways to epenthesize a segment. The first way is to explicitly add precedence relations to and from a new segment from the relevant points in the representation. This would take (19a) and produce the representation in (20).

(20) $\# \rightarrow t \rightarrow a \rightarrow g \rightarrow u \rightarrow \%$, with $a \rightarrow r \rightarrow g$ (r below a and g)

The new structure added to (19a) to produce (20) is offset by having the segment *r* below the underlying segments. Notice the parallels between (19) and (13a), and recall that the actual graphic layout does not matter. The difference is that deletion is detouring around an old segment, while epenthesis is detouring through a new segment. Another parallel between the epenthesis and deletion representations is that both need to be "serialized" to resolve the precedence conflict present in them. As earlier in the chapter, all models of phonology have the relevant resources to produce the proper output to this form.

"Fission" is to epenthesis as coalescence is to deletion. To see this, consider the explicit representation of x-slots and melodies in (21), which shows the splitting of an x-slot to produce a new segment.

(21) a.

$$\begin{array}{cccccccccccc} & & t & & a & & g & & u & & \\ & & | & & | & & | & & | & & \\ \# & \rightarrow & x & \rightarrow & x_i & \rightarrow & x & \rightarrow & x & \rightarrow & \% \end{array}$$

b.

$$\begin{array}{cccccccccccccc} & & t & & a & & & & g & & u & & \\ & & | & & | & & & & | & & | & & \\ \# & \rightarrow & x & \rightarrow & x_i & \rightarrow & x_j & \rightarrow & x & \rightarrow & x & \rightarrow & \% \end{array}$$

This type of epenthesis raises the question how the new x-slot is formed without taking the associated melody along with it; reasons of space preclude a full exposition of this question here beyond pointing out that the interpretation of a bare x-slot will be determined on a language specific basis (see CHAPTER 67: VOWEL EPENTHESIS; CHAPTER 58: THE EMERGENCE OF THE UNMARKED). The differences between these two types of epenthesis suggest differences in the type of epenthetic segment. The approach in (20) supports a prespecified type of epenthetic segment, which cannot be derived from markedness conditions, while the approach in (21) suggests an "Emergence of the Unmarked" (McCarthy and Prince 1994) type of epenthetic segment, because an empty x-slot needs to be interpreted. A final note is that just as the coalescence approach to deletion creates an intermediate representation that can account for compensatory lengthening effects, so this fission approach can also account for simple segment lengthening effect by fissioning the melody along with the x-slot.

If segmental length is encoded on the x-tier as opposed to being marked by a mora (see Ringen and Vago 2010), there are two representations for true geminates (Hayes 1986; Schein and Steriade 1986) that follow naturally. The structure in (22a) is the traditional multiply linked melody representation, while (22b) is a novel looped segment representation for a geminate.

A minimal extension of the epenthesis as addition of a precedence link produces infixing phenomena. Toba Batak infixation, taken from Halle (2001), is presented in (25) (stress is suppressed for expository purposes).

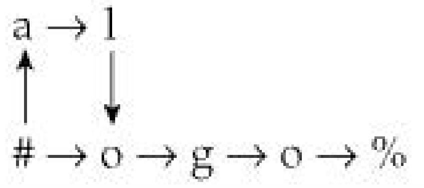
(25) *Nominalizer* -al- *in Toba Batak*

batuk	*b-al-atuk*	'ladder'
ogo	*al-ogo*	'wind'

This infixing pattern can be described as the affix *al* preceding the first vowel of the form. Whether *al* appears as an infix or a prefix is derived from whether the first vowel of the form is in word-initial position. This variation does not change the positional generalization about where this affix is concatenated to the base. (26) provides the precedence graphs for the examples in (25).

(26) a.

b.



As with previous examples, the addition of the affix *al* produces a representation that violates the characteristics of a chain graph, which forces the representation to be serialized at some point in time. See Yu (2007) for further discussion of infixation.

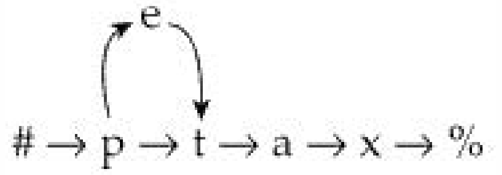
Our last example of (morpho)phonological processes that are illuminated by explicit precedence graphs is from Spokane, in (27). Repetitive morphology in Spokane is interesting because it combines elements of infixation and reduplication. The content of the repetitive morpheme is /e/; it appears to infix if the root begins with two consonants (27a), but causes reduplication if the base begins with a single consonant (27b).

(27) *Repetitive morphology in Spokane* (simplified, from Bates and Carlson 1992)

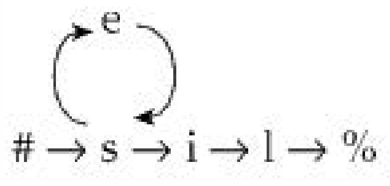
	repetitive	*base*			*repetitive*	*base*	
a.	petax	ptax	'spit'	b.	sesil	sil	'chop'
	qesip	qsip	'long ago'		kekul	kul	'make'

The precedence graphs that capture this behavior are shown in (28).

(28) a.



b.



REFERENCES

Aho, Alfred, John Hopcroft & Jeffrey Ullman. 1985. *The design and analysis of computer algorithms*. Reading, MA: Addison-Wesley Publishing Company.

Alber, Birgit. 2007. Deutsche und Italienische Kurzwörter im Vergleich. In Claudio Di Meola, Livio Gaeta, Antonie Hornung & Lorenza Rega (eds.) *Perspektiven Zwei*, 101–112. Rome: Istituto Italiano di Studi Germanici.

Alber, Birgit. 2009. The foot in truncation. Paper presented at the CUNY Conference on the Foot, January 2009. Available (May 2010) at http://www.cunyphonologyforum.net/footconf.php.

Anderson, Stephen R. 1974. *The organization of phonology*. New York: Academic Press.

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Bach, Emmon. 1968. Two proposals concerning the simplicity metric in phonology. *Glossa* 2. 128–149.

Bagemihl, Bruce. 1991. Syllable structure in Bella Coola. *Linguistic Inquiry* 22. 589–646.

Bates, Dawn & Barry F. Carlson. 1992. Simple syllables in Spokane Salish. *Linguistic Inquiry* 23. 653–659.

Bird, Steven & Ewan Klein. 1990. Phonological events. *Journal of Linguistics* 26. 33–56.

Blevins, Juliette. 2003. The independent nature of phonotactic constraints: An alternative to syllable-based approaches. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*, 375–403. Cambridge: Cambridge University Press.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6. 201–251.

Browman, Catherine P. & Louis Goldstein. 1990a. Gestural specification using dynamically defined articulatory structures. *Journal of Phonetics* 18. 299–320.

Browman, Catherine P. & Louis Goldstein. 1990b. Representation and reality: Physical systems and phonological structure. *Journal of Phonetics* 18. 411–424.

Cairns, Charles. 1988. Phonotactics, markedness and lexical representation. *Phonology* 5. 209–236.

Cairns, Charles & Eric Raimy. 2009. Architecture and representations in phonology. In Raimy & Cairns (2009), 1–16.

Coleman, John & John Local. 1991. The "No Crossing Constraint" in autosegmental phonology. *Linguistics and Philosophy* 14. 295–338.

Czaykowska-Higgins, Ewa & Marie Louise Willet. 1997. Simple syllables in Nxa'amxcin. *International Journal of American Linguistics* 63. 385–411.

de Lacy, Paul. 2007. The formal properties of phonological precedence. Paper presented at the CUNY Conference on Precedence Relations, January 2007. Available (May 2010) at http://www.cunyphonologyforum.net/forum.php.

Fudge, Erik C. 1969. Syllables. *Journal of Linguistics* 5. 253–286.

Fudge, Erik C. 1987. Branching structure within the syllable. *Journal of Linguistics* 23. 359–377.

Gafos, Adamantios I. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20. 269–337.

Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT. Published 1979, New York: Garland.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Oxford & Cambridge, MA: Blackwell.

Halle, Morris. 2001. Infixation versus onset metathesis in Tagalog, Chamorro, and Toba Batak. In Michael Kenstowicz (ed.) *Ken Hale: A life in language*, 153–168. Cambridge, MA: MIT Press.

Hayes, Bruce. 1986. Inalterability in CV phonology. *Language* 62. 321–351.
Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.
Heinz, Jeffrey. 2007. The inductive learning of phonotactic patterns. Ph.D. dissertation, University of California, Los Angeles.
Hulst, Harry van der. 2008. The syllable in RCVP: Structure and licensing. Paper presented at the CUNY Conference on the Syllable, January 2008. Available (May 2010) at http://www.cunyphonologyforum.net/syllconf.php.
Idsardi, William J. & Eric Raimy. Forthcoming. Three types of linearization and the temporal aspects of speech. In Theresa Biberauer & Ian Roberts (eds.) *Challenges to linearization*. Berlin & New York: Mouton de Gruyter.
Idsardi, William J. & Rachel Shorey. 2007. Unwinding morphology. Paper presented at the CUNY Phonology Forum Conference on Precedence Relations, January 2007. Available (May 2010) at http://www.cunyphonologyforum.net/forum.php.
Kahn, Daniel. 1976. Syllable-based generalizations in English phonology. Ph.D. dissertation, MIT.
Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.
Krämer, Martin. 1999. A correspondence approach to vowel harmony and disharmony. (ROA-293.)
Kuratowski, Kazimierz. 1930. Sur le problème des courbes gauches en topologie. *Fundamenta Mathematicæ* 15. 271–283.
Langacker, Ronald. 1969. Mirror image rules II: Lexicon and phonology. *Language* 45. 844–862.
Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT.
Mailhot, Frédéric & Charles Reiss. 2007. Computing long-distance dependencies in vowel harmony. *Biolinguistics* 1. 28–48.
McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12. 373–418.
McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17. 207–263.
McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.
McCarthy, John J. 1989. Linear order in phonological representation. *Linguistic Inquiry* 20. 71–99.
McCarthy, John J. & Alan Prince. 1994. The emergence of the unmarked: Optimality in prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Rutgers University (ROA-13).
McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.
Nevins, Andrew. 2010. *Locality in vowel harmony*. Cambridge, MA: MIT Press.
Odden, David. 1986. On the role of the Obligatory Contour Principle in phonological theory. *Language* 62. 353–383.
Odden, David. 1988. Anti antigemination and the OCP. *Linguistic Inquiry* 19. 451–475.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.
Raimy, Eric. 1999. Representing reduplication. Ph.D. dissertation, University of Delaware.
Raimy, Eric. 2000. *The phonology and morphology of reduplication*. Berlin & New York: Mouton de Gruyter.
Raimy, Eric. 2009. Deriving reduplicative templates in a modular fashion. In Raimy & Cairns (2009), 383–404.
Raimy, Eric & Charles Cairns (eds.) 2009. *Contemporary views on architecture and representations in phonology*. Cambridge, MA: MIT Press.

Ringen, Catherine & Robert M. Vago. 2010. Geminates: Heavy or long? In Charles Cairns & Eric Raimy (eds.) *Handbook of the syllable*. Leiden: Brill.
Rose, Sharon & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80. 475–531.
Samuels, Bridget. 2009. The structure of phonological theory. Ph.D. dissertation, Harvard University.
Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Lausanne & Paris: Payot.
Schein, Barry & Donca Steriade. 1986. On geminates. *Linguistic Inquiry* 17. 691–744.
Sievers, Eduard. 1881. *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. Leipzig: Breitkopf & Härtel.
Sloan, Kelly D. 1991. Syllables and templates: Evidence from Southern Sierra Miwok. Ph.D. dissertation, MIT.
Steriade, Donca. 1999. Alternatives to syllable-based accounts of consonantal phonotactics. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.) *Item order in language and speech*, 205–242. Prague: Karolinum Press.
Sweet, Henry. 1877. *A handbook of phonetics*. Oxford: Clarendon Press.
Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.
Vaux, Bert & Andrew Wolfe. 2009. The appendix. In Raimy & Cairns (2009), 101–143.
Wilson, Robin J. 1996. *Introduction to graph theory*. Longman: London.
Yu, Alan C. L. 2007. *A natural history of infixation*. Oxford: Oxford University Press.

35 Downstep

Bruce Connell

1 Introduction

Downstep is a pitch-lowering phenomenon that is widely recognized to occur in tone languages, particularly those of sub-Saharan Africa, in which it was first identified. It is also attested in several languages of the Americas, though only very rarely in Asia. The concept of downstep has also more recently been extended to account for phenomena associated with intonation in non-tonal languages,[1] and is indeed perhaps better known to non-specialists in this context. Downstep is most commonly described as the lowering influence of a low tone (L) on a following high tone (H), such that a new, lower, "ceiling" is set for all subsequent Hs within a specifiable domain or prosodic unit. One of the early and still primary areas of debate in the study of downstep pertains to the sameness or otherwise of downstep as effected by a surface L as opposed to an underlying or floating L. Another area of interest has to do with the implications of its recurrent and cumulative nature for the analysis of contrastive tone levels; a language with only two apparent contrastive lexical levels (e.g. H, L) will manifest several intervening levels in actual speech, including situations in which a H occurring late in an utterance may be realized at a lower pitch than a L early in the same utterance. Another debate centers on the analysis of downstep applying to tones other than H; while the paradigm case is the lowering of H in a language with two tones (H, L), downstep does occur in languages with more than two tones, and in some such languages, cases of downstepped mid (M) and L tones are also attested. Whether such cases parallel downstepping of H remains moot. In addition, there are apparent cases of "upstep," a tone raising phenomenon that ideally would be symmetrical to downstep in the details of its realization, though few if any of the attested cases are precisely so.

[1] An anonymous reviewer suggests the term "non-tonal language" is a misnomer, given that "there is no essential qualitative difference between tones based on their origin in the lexicon or in phrasal phonology." While I am sympathetic to the view that "non-tonal" is problematic, the claim that "there is no essential difference" is contentious. This discussion is outside the scope of the present chapter, but Hyman's (2001: 1368) definition of a tone language serves to distinguish the two types of language, a tone language being one, "in which an indication of pitch enters into the lexical realization of at least some morphemes."

Discussion as to how best to characterize downstep from a theoretical perspective has benefited from and contributed to general phonological theory, with proposals being shaped by phonetic and phonological approaches. This debate remains unresolved, with some scholars advocating the view that downstep is best accommodated in the phonetic implementation component (e.g. Poser 1984; Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988), and others arguing that it is phonological (e.g. Snider 1998, 1999).

The discussion of downstep was particularly fruitful in the 1960s and then again through the 1980s and 1990s, with the advent of autosegmental phonology (CHAPTER 14: AUTOSEGMENTS) and feature geometry (CHAPTER 27: THE ORGANIZATION OF FEATURES). Since then the debate has subsided somewhat, though two good general presentations of downstep have appeared in recent years, in Yip (2002) and Gussenhoven (2004). The most recent detailed argument for a particular theoretical approach to downstep is Snider (1999).

This chapter characterizes downstep and related tonal phenomena, summarizing with representative data the key issues and current views. In doing so, I draw on those sources just mentioned, as well as several other published (and some unpublished) studies that have contributed to our understanding of downstep.

The remainder of the chapter is divided as follows. §2 is an overview of work leading to the recognition and study of downstep in tone languages. This section defines the key terms used in the discussion, presents relevant illustrative data, and introduces some of the important theoretical issues involved. The third section then discuss downstep-related issues, including its distribution, downstepping of tones other than H and in languages with more than just H and L tones, and downstep in non-tonal languages. What triggers downstep is discussed in §4, phonetic aspects of downstep in §5, and issues pertaining to upstep and H-raising in §6. A particular personal concern has been the inconsistent and conflicting use of terminology found in the discussion of downstep and related pitch phenomena, and its importance for developing an adequate understanding of these phenomena; these issues are examined in §7. §8 presents instrumental evidence from the Bantoid language Mambila that bears on a resolution to the issues discussed in §7.

Inevitably, some aspects of the topic receive fuller treatment than others; in particular, the substantial literature in which the notion of downstep is used in the analysis of non-tonal languages does not get the attention it deserves. This is in large part due to space constraints, but it is to some extent also a reflection of my own expertise and familiarity with the subject, as well as an attempt to address a perceived imbalance in the general theoretical phonology literature; while downstep is primarily a phenomenon of tone languages, it has perhaps received greater attention for its operation in non-tonal languages, and this aspect of the discussion is more accessible to the general reader. For related discussion, see also CHAPTER 45: THE REPRESENTATION OF TONE; CHAPTER 114: BANTU TONE.

2 An overview of downstep in language study: Phonological issues

2.1 *Early studies*

The phenomenon now known as downstep was first noted in print well over a century ago, by Christaller (1875: 15). In his discussion of Fante (Kwa, Ghana)

(1) a. idiɣ(e) ɔbɔŋ ke ŋkut [¯ ¯ ¯ ¯ ¯ ¯ ¯]
'It isn't a *mosquito* that I see.'
b. idiɣ(e) ɔbɔŋ ke ŋkut [¯ ¯ ¯ _ – – –]
'It isn't a *piece of cane* that I see.'
c. idiɣ(e) ɔbɔŋ ke ŋkut [¯ ¯ ¯ – – – –]
'It isn't the *chief* that I see.'
d. ekpeɲɔŋ edi ufɔk [_ ¯ – – – – _]
'Ekpenyong came to the house.'
e. ekpeɲɔŋ emen inuen ɔɲɔŋ edi ufɔk [– ¯ ¯ ¯ – – – – – – – – – _]
'Ekpenyong picked up the bird and came home.'

The first three sentences, which differ only in the italicized words (i.e. the tone pattern of /ɔbɔŋ/) show, in (1a) all Hs, in (1b) Hs where an intervening L has conditioned a lowering of following Hs, and in (1c) a similar lowering of the last four Hs, though without a preceding L to effect the lowering. Sentences (1d) and (1e) show successive lowering of what are analyzable as Hs, with only the initial and final tones being L. The "terracing" seen with the Hs in these two sentences shows the classic effect of downstep.

These data reveal the unsatisfactory nature of an analysis that sees the lowered tones as M. First, such an analysis would require /ké ŋꜜkút/ 'that I see' in sentences (1a)–(1c) to be identified first as HHH (1a) and later as MMM (1c); in (1b), while the tones of /ké ŋꜜkút/ are phonetically similar to those in (1c), they are best treated as HHH, since they represent a conditioned lowering. Further, the successive lowering in (1d) and (1e) would require a number of mid tones (i.e. Ms of different heights), with /édí/ 'came' bearing a different, lower, M in (1e) than in (1d).

Winston examines briefly a different solution, which contrasts non-low and low, with non-low tones divided into H and M, and with restrictions on the distribution of M, but finds this equally awkward. He instead proposes "two distinct systems of contrasting tonal units" (1960: 187): first, H *vs.* L, which accounts for sentence (1a) *vs.* (1b), and second, "a unit of 'downstep'," which operates only in the context of HH, and accounts for sentence (1a) *vs.* (1c). Winston draws attention to the fact that not only does downstep distinguish sentences, but it is phonemic in its own right, as the words /ɔ́bɔ́ŋ/ 'mosquito', /ɔꜜbɔ́ŋ/ 'chief', and /ɔ̀bɔ̀ŋ/ 'cane' demonstrate. His analysis is also insightful in that it focuses attention not on the nature of the tones themselves – e.g. the last four tones of (1a)–(1c) – but on the relation between these tones and preceding tones; the drop is the realization of downstep.

2.3 *Automatic and non-automatic downstep*

As mentioned, the Efik data included instances of "non-automatic" downstep, in which a H is lowered by an underlying or floating L, and the lowering of a surface H tone under the influence of a preceding surface L, termed "automatic" downstep. These terms were introduced in Stewart (1965) and continue to be used, though many writers use (simply) downstep to refer to "non-automatic" downstep, and "downdrift" when referring to automatic downstep. As discussed below in §7, however, there are in fact different tonal processes grouped together as downdrift, and the use of downdrift in referring to automatic downstep as well

(1979) proposed an alternative view, which helped to lay the foundation for most current views of downstep. Clements's proposal saw terracing as "the result of intonational processes applying to the tone level frame itself, rather than directly to individual tones" (1979: 358). The occurrence of actual downstepped tones was restricted to initial position in sequences in which pitch was lowered, but they have the effect of precipitating a (downward) register shift, re-establishing the levels at which subsequent tones within a given prosodic unit are realized. Whether this shift affects all tones (i.e. H and L, and M in a three-tone system) within a given prosodic unit, or just H, became an empirical and language specific question.

2.5 *Downstep in autosegmental phonology*

The development of autosegmental phonology (Leben 1973; Goldsmith 1976) provided the rest of the foundation for our current understanding of downstep and related tonal phenomena. Perhaps the key contribution of autosegmental phonology to our understanding of downstep was its ability to represent tone on a separate tier, and consequently different tone types on different tiers. Several authors have contributed to this debate and its development, exploiting this insight in different ways and to different degrees. Among them are Hyman (1979, 1993), Clements (1983, 1990), Stewart (1983, 1993), Pulleyblank (1986), Yip (1989, 1993), Snider (1990, 1999), and Clark (1993). The most recent and most detailed of these contributions is that of Snider (1999), who provides a proposal for an understanding of downstep that lays considerable emphasis on the incorporation of upstep as a phenomenon to be accounted for by the same means as downstep, as well as a critique of the related approaches, at least where they differ from his own.

Snider's (1999) proposal is presented within the theoretical framework of Register Tier Theory (RTT), which exploits the mechanisms of autosegmental phonology and feature geometry, and their tiered and hierarchical representations. RTT incorporates a Register tier admitting two features, h and l, a Tonal tier with two features, H and L, a Tonal Root Node tier (TRN), and a Tone-Bearing Unit (TBU) tier. Features on the register and tonal tiers are linked to a node on the TRN tier, and each TRN node is in turn linked to a mora on the TBU tier. This permits the specification of four level tones, which Snider labels Hi (= h, H), Mid2 (= h, L), Mid1 (= l, H), and Lo (= l, L). In a two-tone language (i.e. with H and L, as is frequently found in sub-Saharan Africa), the register feature associated with a particular TBU permits a register shift relative to the preceding TBU's register: h = higher than the previous register setting, l = lower than the previous register setting. Automatic downstep, then, is represented by the spread of the register feature l, which effects a downward shift, realizing the following high tone a step lower than on the preceding register setting.

Snider's tonal features H and L equate well with Yip's (1980) [±High], Pulleyblank's (1986), and Yip's (1989) [±Raised]. However, his h and l register features provide an advantage over Yip's and Pulleyblank's [±Upper], in that they are relative features, whereas [±Upper] is non-relative; it is either high or low. Both Snider's system and the Yip/Pulleyblank system, then, account well for phonemic tone levels, but the relative nature of Snider's register features permits description of the cumulative nature and the terracing associated with downstep.

3 Distributional characteristics of downstep

3.1 *Downstep type*

Non-automatic downstep (often equated with phonologically distinctive downstep) is usually found only in systems that also have automatic (or non-distinctive) downstep, though the two need not co-occur. Automatic downstep is not uncommon in the absence of non-automatic downstep; however, the reverse – i.e. cases of non-automatic downstep in the absence of automatic downstep – is rare.

Automatic downstep has been reported in languages such as Hausa (Chadic, Nigeria) by Leben (1984), Lindau (1986), Inkelas and Leben (1990), and Leben *et al.* (1989), and in Yoruba (Benue-Congo, Nigeria) by Connell and Ladd (1990), Laniran (1992), Akinlabi and Liberman (1995), and Laniran and Clements (2003). In neither of these languages is non-automatic downstep found, and downstep is neither lexically nor grammatically distinctive. Both automatic and non-automatic downstep occur in a great many languages, including not only the paradigm examples of Akan, Efik, and Igbo, but also languages such as Baule (Kwa, Ivory Coast; Ahoua 1996), Bimoba (Gur, Togo; Snider 1998), Chumburung (Kwa, Ghana; Snider 1999), Yala-Ikom (Benue-Congo, Nigeria; Armstrong 1968), and Zande (Adamawa-Ubangi, Democratic Republic of the Congo; Boyd 1981).

For only very few languages has non-automatic downstep been reported in the absence of automatic downstep. Three such languages are Dschang (Grassfields Bantu, Cameroon), Ikaan (Benue-Congo, Nigeria; Salffner 2009), and Kikuyu (Bantu, Kenya; Clements and Ford 1980), and each of these present analytic complexities for which consensus has yet to be reached.

3.2 *Geographical distribution*

All of the above mentioned languages are geographically located in sub-Saharan Africa, the region best known for downstep, and representative of several different language families and phyla. Downstep is also found in the Americas, the best-known examples being in Central America. Isthmus Zapotec (Oto-Manguean, Mexico; Mock 1981, cited in Yip 2002) shows downstep functioning in a manner expected from research on African languages, i.e. it is triggered by a floating L. Other languages of this region show both downstep and upstep; varieties of Mixtec (Oto-Manguean, Mexico) are discussed in §6 below.

Only very rarely has downstep been reported among Asian languages. One such language is Kuki-Thaadow (Tibeto-Burman; Hyman 2007), spoken in north-east India and Burma. Contrary to the well-reported differences between Asian and African tone languages, Kuki-Thaadow's tone system behaves very much like those found in Africa. Hyman analyzes Kuki-Thaadow as having three underlying tones: HL, H, and L. Downstep occurs when HL precedes H, with L being realized as a downstep on the following H, as shown in (2) (Hyman 2007: 6):

(2) /mêeŋ vóm thúm hí/ → méeŋ ꜜvóm thúm hí [¯ – – –]
'these three black cats'

According to Hyman's report, downstep in Kuki-Thaadow is realized phonetically by raising of the preceding H, with the amount of raising being determined by the number of downsteps to follow; though no instrumental data is presented, this is reminiscent of Rialland's (2001) findings for Dagara (see §6).

3.3 *Downstep in languages with more than two tones and with tones other than H*

The discussion thus far has centered on languages with just two tones, H and L, and downstepping of H tones. This is clearly the most commonly attested situation, though downstep does occur in languages with more than two tones, and with tones other than H, including in some languages with just two underlying tones.

The first language with more than a basic two-tone system recognized as having downstep, and as downstepping tones other than H, was Yala-Ikom (Benue-Congo, Nigeria; Armstrong 1968). Yala-Ikom has three contrastive tones, H, M, and L. H is lowered after both L and M, and M is lowered following L; this occurs regardless of whether the tone causing the lowering is underlying or surface (i.e. floating or associated), and terracing results in both situations. The example below (Armstrong 1968: 53) illustrates downstepping of M following a M (Armstrong used the diacritic ' to indicate downstep); note that the downstepped M is followed by H:

(3) ɔ́ tābɔ̄l↓ɔ̄nɛ̄ ní [¯ – – _ _ –] HMMMMH
'It did not begin in the evening.'

Downstep begins with the third M, a result of the underlying ("latent" in Armstrong's terms) floating L of /là/ 'in', which remains after vowel elision. Evidence for M triggering downstep of H is provided in examples such as the derived verbal noun, /òré↓ré/ 'eating'; the presence of a floating M is confirmed by cross-dialectal comparison, where M surfaces in Yala-Ogoja /òrɔ́ɔ̄ré/ 'eating'.

In Vute (Bantoid, Cameroon), also a three-tone language with H, M, and L, H apparently undergoes both automatic and non-automatic downstep (Guarisma 1978; Thwing and Watters 1987). Guarisma describes the second H of a HLH as being lowered, but does not comment on whether subsequent Hs are terraced, nor does she comment on the interesting situation of M, and the possibility of overlap of ↓H and M, or whether Ms and Ls lower correspondingly. However by Guarisma's examples, non-automatic downstep does extend beyond the first tone of the sequence and so terracing exists. Guarisma describes non-automatic downstep as marking associative constructions, but her examples of both non-automatic and automatic downstep are with associatives.

(4) ŋgwè sèhí → [¯ _ –] 'head (of) abscess'
ɓár jémi → [¯ – –] 'spots (of) leopard'

Bamileke-Dschang (Grassfields Bantu, Cameroon) is a language with two tones, H and L. Dschang has received considerable attention in the literature on downstep (Tadajeu 1974; Hyman 1985; Pulleyblank 1986; Clark 1993; Manfredi 1993; Stewart 1993; Bird and Stegen 1995; Snider 1999) as the first language (and one of very few languages) analyzed as having downstep affecting L as well as H; it has a four-way surface contrast, H, $^{\downarrow}$H, L, $^{\downarrow}$L. There is considerable divergence among the views of these authors as to the nature of downstep in Dschang; the downstepped H is attributable to a *following* floating L, with leftward spreading, and this seems largely agreed (Tadajeu 1974; Pulleyblank 1986; Snider 1999). Snider attributes the downstepped L to a floating H, which, when inserted between two Ls, results in downstepping of the second L. Clark (1993), however, prefers to see Dschang with a basic four-tone system and only relatively few occurrences of downstep.

So, while there are languages like Yala-Ikom with more than two tones in which both H and M may be downstepped, and others like Vute in which only H is affected, there are no languages reported where (non-automatic) downstep affects M (or L) but not H. Similarly, like Dschang, there are no languages reported where $^{\downarrow}$L occurs but not $^{\downarrow}$H.

Interestingly, for Yala-Ikom, discussed above, Armstrong (1968) reports that the effect of downstep triggered by M is indistinguishable from that triggered by L; i.e. $^{\downarrow}$H lowers the same degree regardless of whether L or M is the trigger. On the other hand, in Yala-Ikom $^{\downarrow}$H remains higher than M. This is contrary to reports for other three tone languages in which H is downstepped and for which it is claimed $^{\downarrow}$H is indistinguishable from M (e.g. Supyire, Gur, Mali; Carlson 1983; Moba, Gur, Togo; Russell 1996, cited in Snider 1999; Bimoba, Gur, Ghana; Snider 1998). Snider (1998) provides instrumental evidence for the phonetic equivalence of $^{\downarrow}$H and M in Bimoba.

3.4 *Downstep in "non-tonal" languages*

The possibility that downstep could account for pitch phenomena in non-tonal languages was first introduced in Pierrehumbert's (1980) work on English, in which it was proposed that declination could, largely, be accounted for as the successive lowering of pitch accents (CHAPTER 116: SENTENTIAL PROMINENCE IN ENGLISH). The term "catathesis" rather than downstep was adopted for a time (Poser 1984; Beckman and Pierrehumbert 1986), in order to avoid the terminological conflicts inherent in the use of downdrift *vs.* downstep and automatic *vs.* non-automatic downstep (e.g. as mentioned briefly in §2.3, and more on which in §7.1.3); usage has since reverted to, and settled on, downstep. Pierrehumbert's model has evolved considerably, both in her own work and that of others following, broadly speaking, the same tack. It has been applied to several other languages, e.g. Japanese (Japonic, Japan; Pierrehumbert and Beckman 1988; Kubozono 1989), and Dutch (Germanic, The Netherlands and Belgium; van den Berg *et al.* 1992), with discussion on the nature of downstep in these languages largely being separate from the debate on its functioning in African and other tone languages. One of the key issues has been the extent to which downstep is distinguishable from declination; in other words, can or does downstep account for all downward movement of pitch across a phrase or utterance? While there has been some effort to include both downstep and declination in models of intonation in such languages, resulting in a move away from the restrictive position in Pierrehumbert (1980), Ladd (2008) points out the methodological/empirical difficulty in separating the two. (In tone languages this is less problematic; see §7.1.2.) Ladd's general view of downstep as part of intonation (e.g. 1992, 1993, 2008) sees (intonational)

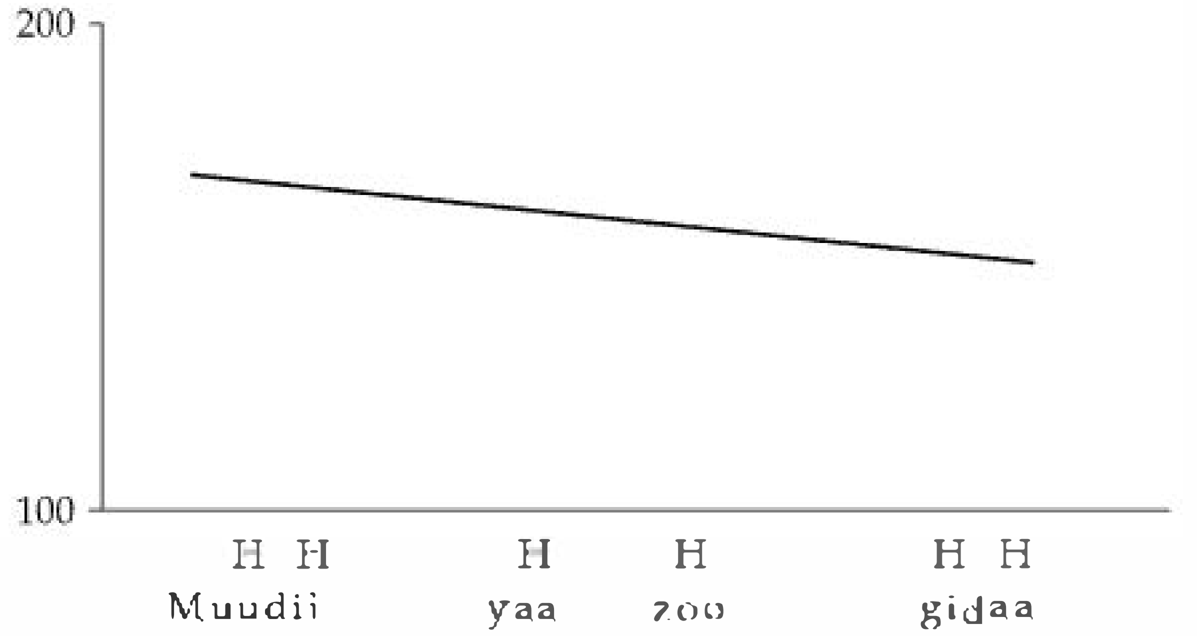


Figure 35.1 Declination illustrated in a sequence of Hausa High tones in the sentence *Muudii yaa zoo gidaa* 'Muudii came home', adapted from Lindau (1986)

7.1.2 Downdrift

Downdrift is somewhat more difficult to characterize, precisely because, as mentioned earlier, the term has been used in different senses. Most commonly, it has been used synonymously with Stewart's automatic downstep. Hombert (1974: 171), for example, describes downdrift as "the progressive lowering of a high tone after a low tone," and in a footnote explicitly equates it with automatic downstep. Similar views are expressed by a range of authors including, more recently, Snider and van der Hulst (1993) and Hyman (2001). However, Hombert also attributes an intonational element to downdrift, observing that Ls also descend, and then suggests that the term downdrift refer to "the lowering of like tones (consecutive or not)" (1974: 172, fn. 6). Given this proviso concerning like tones (or, when it applies to consecutive like tones), it appears to be the same phenomenon as that described earlier as declination; however as we have seen, declination is clearly something quite independent of automatic downstep. Both of these views of downdrift – i.e. that it involves a local assimilation between Ls and Hs, and that it is a phrase or sentence level effect – are found elsewhere in the literature, though it is most frequently characterized as being equivalent to automatic downstep. An illustration of downdrift of a presumably assimilatory nature, through the alternation of Hs and Ls, can again be taken from Lindau's (1986) study of Hausa (Chadic, Nigeria), and is shown in Figure 35.2. In this figure, as in Figure 35.1, a trend line for the H tone only is represented (in Lindau's representation, the second sentence also shows a downtrend affecting the Ls, and both map the actual F0 trace). What is important in comparing the two is that the slope in Figure 35.2 is steeper than that in Figure 35.1, with a decline of 33 percent per second; i.e. the slope shown in Figure 35.2 combines the declination of Figure 35.1 with the effect of downdrift, the local assimilation of Hs to Ls.[6]

7.1.3 Downstep

An important, indeed defining, feature of downstep, in addition to its lowering of a H relative to a preceding H (or lowering of other tones relative to preceding

[6] Gussenhoven (2004: 101) cites similar evidence for Japanese from Poser (1984).

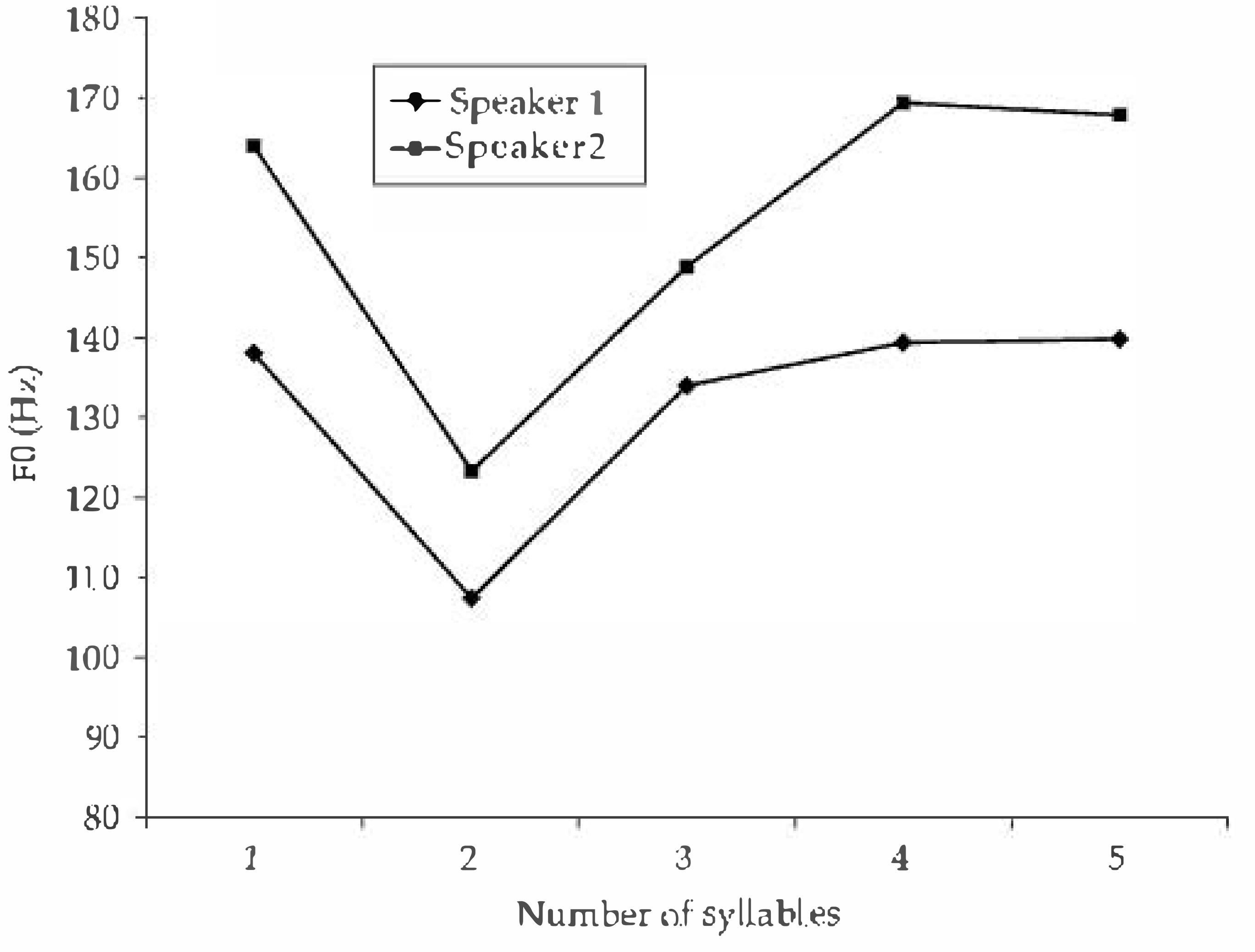


Figure 35.8 Pitch traces of T1T4T1T1T1 (HLHHH) for two Mambila speakers, showing the absence of automatic downstep

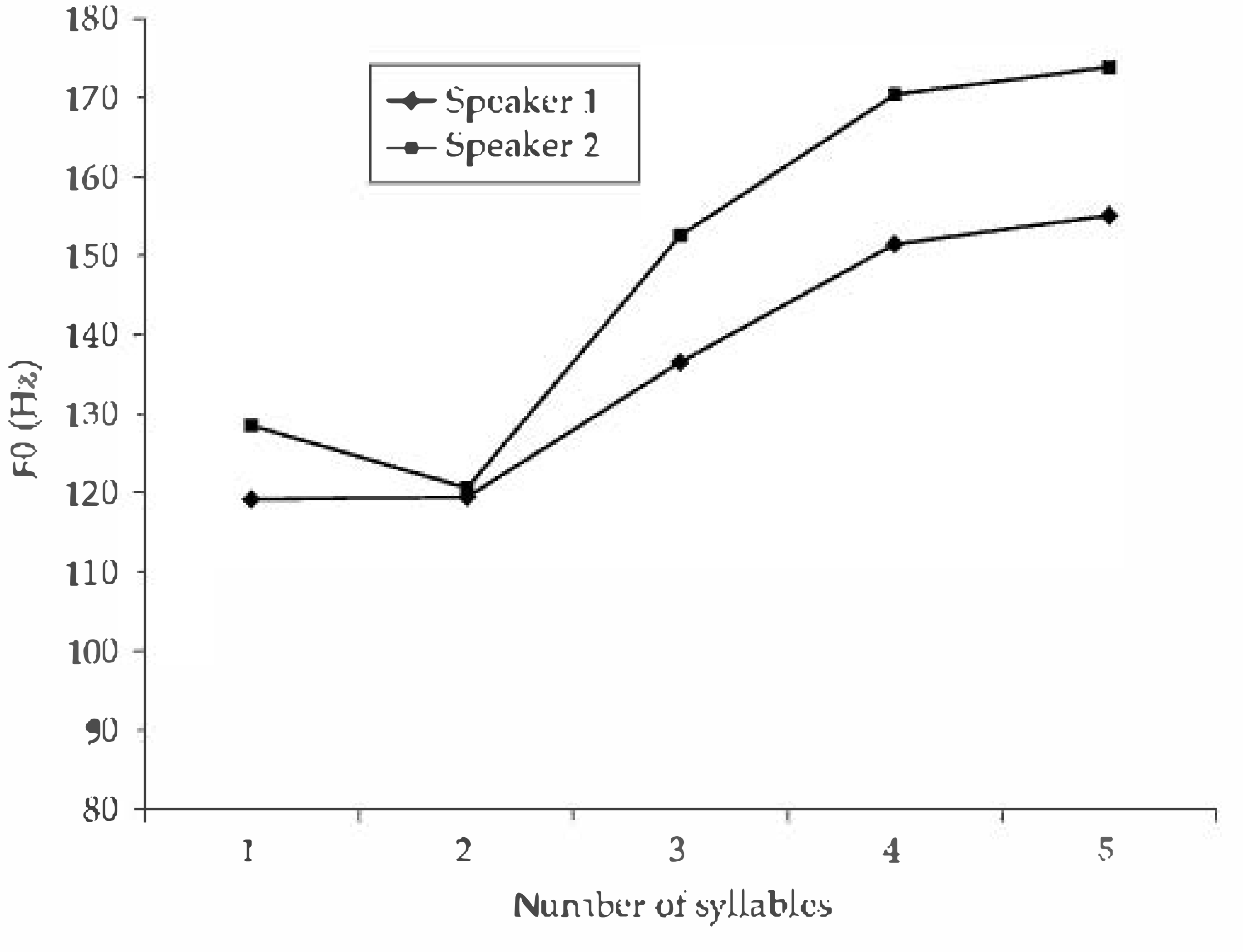


Figure 35.9 Pitch traces of T4T4T1T1T1 (LLHHH) for two Mambila speakers, showing the absence of automatic downstep

Rather than downstep, Mambila has a local interaction, a lowering of a high(er) tone following a low(er) tone, that appears to be simply a phonetic effect that is corrected once the low tone is not involved, allowing the H to regain its former height. That the effect of this interaction is cumulative, i.e. in a

Armstrong, Robert G. 1968. Yala (Ikom): A terraced-level language with three tones. *Journal of West African Languages* 5. 49–58.
Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.
Bendor-Samuel, John (ed.) 1974. *Ten Nigerian tone systems*. Jos: Institute of Linguistics.
Berg, Rob van den, Carlos Gussenhoven & Toni Rietveld. 1992. Downstep in Dutch: Implications for a model. In Docherty & Ladd (1992), 335–359.
Bird, Steven & Oliver Stegen. 1995. *The Bamileke Dschang associative construction: Instrumental findings*. Edinburgh: Centre for Cognitive Science, University of Edinburgh.
Boyd, Raymond. 1981. À propos de la notion de faille tonale. In Gladys Guarisma (ed.) *Tons et accents dans les langues africaines*, 39–64. Paris: SELAF.
Carlson, Robert. 1983. Downstep in Supyire. *Studies in African Linguistics* 14. 35–45.
Christaller, J. G. 1875. *A grammar of the Asante and Fante language*. Basel: Basel Evangelical Missionary Society.
Clark, Mary. 1993. Representation of downstep in Dschang Bamileke. In van der Hulst & Snider (1993), 29–73.
Clements, G. N. 1979. The description of terraced-level tone languages. *Language* 55. 536–558.
Clements, G. N. 1983. The hierarchical representation of tone features. In Ivan R. Dihoff (ed.) *Current approaches to African linguistics*, vol. 1, 145–176. Dordrecht: Foris.
Clements, G. N. 1990. The status of register in intonation theory: Comments on the papers by Ladd and by Inkelas and Leben. In Kingston & Beckman (1990), 58–71.
Clements, G. N. & Kevin C. Ford. 1980. On the phonological status of downstep in Kikuyu. In Didier Goyvaerts (ed.) *Phonology in the 1980s*, 309–357. Ghent: E. Story-Scientia.
Connell, Bruce. 1999. Four tones and downtrend: a preliminary report on pitch realization in Mambila. In Paul Kotey (ed.) *New dimensions in African linguistics and languages: Trends in African linguistics*, vol. 3, 75–88. Trenton, NJ: Africa World Press.
Connell, Bruce. 2002. Downdrift, downstep and declination. In Gut & Gibbon (2002), 3–12.
Connell, Bruce. 2003. Pitch realization and the four tones of Mambila. In Shikeki Kaji (ed.) *Cross-linguistic studies of tonal phenomena: Historical development, phonetics of tone, and descriptive studies*, 181–197. Tokyo: ILCAA.
Connell, Bruce & D. Robert Ladd. 1990. Aspects of pitch realization in Yoruba. *Phonology* 7. 1–29.
Daly, John P. 1993. *Representation of tone in Peñoles Mixtec*. Dallas: Summer Institute of Linguistics.
Docherty, Gerard J. & D. Robert Ladd (eds.) 1992. *Papers in laboratory phonology II: Gesture, segment, prosody*. Cambridge: Cambridge University Press.
Gandour, Jackson, Siripong Potisuk & Sumalee Dechongkit. 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22. 477–492.
Gibbon, Dafydd. 1987. Finite state processing of tone systems. In *Proceedings of the 3rd Conference of the European Chapter of the Association for Computational Linguistics*. 291–297.
Gibbon, Dafydd. 2001. Finite state prosodic analysis of African corpus resources. In Paul Dalsgaard, Børge Lindberg, Henrik Benner & Zheng-hua Tan (eds.) *Proceedings of Eurospeech 2001*, 83–86. Aalborg: ISCA Archive.
Goldsmith, John A. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.
Guarisma, Gladys. 1978. *Etudes vouté (langue bantoïde du Cameroun)*. Paris: SELAF.
Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
Gut, Ulrike & Dafydd Gibbon (eds.) *Typology of African prosodic systems*. Bielefeld: University of Bielefeld.
Hombert, Jean-Marie. 1974. Universals of downdrift: Their phonetic basis and significance for a theory of tone. *Studies in African Linguistics*. Supplement 5. 169–183.
Hulst, Harry van der & Keith L. Snider (eds.) 1993. *The phonology of tone: The representation of tonal register*. Berlin & New York: Mouton de Gruyter.

Pike, Eunice V. & Kent Wistrand. 1974. Step-up terrace tone in Acatlán Mixtec (Mexico). In Ruth M. Brend (ed.) *Advances in tagmemics*, 81–104. Amsterdam: North Holland.
Poser, William J. 1984. The phonetics and phonology of tone and intonation in Japanese. Ph.D. dissertation, MIT.
Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology*. Dordrecht: Reidel.
Rialland, Annie. 2001. Anticipatory raising in downstep realization: Evidence for preplanning in tone production. In Shigeki Kaji (ed.) *Cross-linguistic studies of tonal phenomenon: Tonogenesis, typology, and related topics*, 301–322. Tokyo: ILCAA.
Rialland, Annie & P. A. Somé. 2000. Dagara downstep: How speakers get started. In Vicki Carstens & Fredrick Parkinson (eds.) *Advances in African linguistics: Trends in African linguistics*, 251–263. Trenton, NJ: Africa World Press.
Russell, Jann. 1996. Some tonal sandhi rules in Moba. Togo: Summer Institute of Linguistics.
Salffner, Sophie. 2009. Tone in the phonology, lexicon and grammar of Ikaan. Ph.D. dissertation, University of London.
Schachter, Paul. 1965. Some comments on John M. Stewart's "The typology of the Twi tone system." *Bulletin of the Institute of African Studies* 1. 28–42.
Snider, Keith L. 1990. Tonal upstep in Krachi: Evidence for a register tier. *Language* 66. 453–474.
Snider, Keith L. 1998. Phonetic realization of downstep in Bimoba. *Phonology* 15. 77–101.
Snider, Keith L. 1999. *The geometry and features of tone*. Dallas: Summer Institute of Linguistics & University of Texas at Arlington.
Snider, Keith L. & Harry van der Hulst. 1993. Issues in the representation of tonal register. In van der Hulst & Snider (1993), 1–27.
Stewart, John M. 1965. The typology of the Twi tone system. Preprint from the *Bulletin of the Institute of African Studies* 1. 1–27.
Stewart, John M. 1983. Downstep and floating low tones in Adioukrou. *Journal of African Languages and Linguistics* 5. 57–78.
Stewart, John M. 1993. Dschang and Ebrie as Akan-type total downstep languages. In van der Hulst & Snider (1993), 185–244.
Tadajeu, Maurice. 1974. Floating tones, shifting rules, and downstep in Dschang-Bamileke. *Studies in African Linguistics*. Supplement 5. 283–290.
Thomas, Elaine. 1974. Engenni. In Bendor-Samuel (1974), 13–26.
Thwing, Rhonda & John Watters. 1987. Focus in Vute. *Journal of African Languages and Linguistics* 9. 95–121.
Urua, Eno-Abasi E. 1996–97. A phonetic analysis of Ibibio tones: A preliminary investigation. *Journal of West African Languages* 26. 15–25.
Urua, Eno-Abasi E. 2002. The tone system of Ibibio. In Gut & Gibbon (2002), 65–85.
Ward, Ida C. 1933. *The phonetic and tonal structure of Efik*. Cambridge: Heffer.
Welmers, William E. 1959. Tonemics, morphotonemics, and tonal morphemes. *General Linguistics* 4. 1–9.
Welmers, William E. 1965. Some comments on John M. Stewart's The typology of the Twi tone system. Preprint from the *Bulletin of the Institute of African Studies* 1. 28–42.
Welmers, William E. 1973. *African language structures*. Berkeley: University of California Press.
Winston, F. Dennis. 1960. The "mid" tone in Efik. *African Language Studies* 1. 185–192.
Xu, Yi. 1994. Production and perception of coarticulated tones. *Journal of the Acoustical Society of America* 95. 2240–2253.
Yip, Moira. 1980. The tonal phonology of Chinese. Ph.D. dissertation, MIT.
Yip, Moira. 1989. Contour tones. *Phonology* 6. 149–174.
Yip, Moira. 1993. Tonal register in East Asian Languages. In van der Hulst & Snider (1993), 245–268.
Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

36 Final Consonants

Marie-Hélène Côté

1 Introduction

Final consonants, in the stem, the word, or the phrase, often display properties that set them apart from consonants in other positions. Basic principles of syllabification predict that final consonants are codas (see also CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE) and, as such, are expected to pattern like non-final codas. Final consonants thus pose an analytical challenge when this expectation is not fulfilled.

Languages in which final consonants simply mirror internal codas are referred to as "symmetrical." In Manam, only nasals appear in both positions, and stress is regularly attracted to closed syllables, internal and final (1a); default stress is penultimate in the absence of closed syllables (1b). So final consonants in Manam display the same segmental profile and stress-attracting power as internal codas (Buckley 1998; Piggott 1999).

(1) a. ['embegi] 'sacred flute'
[ʔu'laŋ] 'desire'
[ura'pundi] 'I waited for them'
b. [wa'bubu] 'night'

Spanish and Selayarese offer other illustrations of the correspondence between internal codas and final consonants. As shown by Harris's (1983: 14–15) list of word-medial and final rimes in Spanish, the set of permissible codas is the same in both positions and includes any consonantal category, possibly followed by [s]. Final syllables closed by consonants also attract stress, which is consistent with their contributing weight to the final syllable, as coda consonants regularly do cross-linguistically (CHAPTER 57: QUANTITY-SENSITIVITY). In Selayarese (Mithun and Basri 1986), medial codas are restricted to homorganic nasals, the first parts of geminates, and [ʔ]; final consonants are limited to [ʔ] and [ŋ]. Assuming that [ʔ] and [ŋ] lack place specification (e.g. Paradis and Prunet 1993; Lombardi 2002), all codas can be characterized by the absence of independent place features (CHAPTER 7: FEATURE SPECIFICATION AND UNDERSPECIFICATION; CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION): they are placeless or acquire the place of the following onset.

In many other languages, final consonants pattern differently from internal codas. First, the right edge of constituents regularly hosts more consonants than internal codas may accommodate. Second, final consonants may be ignored in the application of metrical processes, while internal codas cannot be. These two tendencies, formulated in (2), define final consonant exceptionality.

(2) a. *Segmental immunity*
Final consonants escape segmental constraints that apply to internal codas.
b. *Metrical invisibility*
Final consonants are ignored in the application of metrical processes.

Both patterns occur in Cairene Arabic. While only one consonant is allowed in phrase-internal codas, two may appear phrase-finally (Wiltshire 2003); the additional final consonant is said to escape the coda conditions applicable elsewhere in the phrase. In addition, word stress is attracted to non-final CVC syllables and final CVCC/CVVC, as opposed to CV and final CVC, as if the final consonant were invisible to the stress assignment algorithm (Hayes 1995).

Final consonant exceptionality has attracted considerable attention in the development of modern phonological theory. Segmental immunity (2a), clearly the most widely discussed aspect of final exceptionality, is treated in §2–§4, starting in §2 with a review of various representative patterns (generalizations). Analyses of the special behavior of final consonants have almost exclusively relied on special accommodations to syllable structure in final position; different representational devices are examined and compared in §3 (representations). Representations, however, provide only part of the story: they offer a formal frame for the expression of the specificity of final consonants, but no explanation for it. §4 (motivations) is concerned with the formal, grammatical, or functional factors that have been called upon to account for the freedom of occurrence of right edge consonants. Metrical invisibility (2b) and its relationship to segmental immunity are addressed in §5.

Final consonants are also implicated in other processes, which are not reviewed in this chapter, since they fall under different topics. First, if final consonants appear with greater freedom, they are also regularly subject to deletion processes (CHAPTER 68: DELETION). Final clusters variably simplify in many languages. The factors that govern simplification, however, appear to be relevant to final and non-final clusters alike (e.g. Côté 2004), and I have chosen not to address this topic. Single final consonants also delete, giving rise to various types of C/Ø alternations. Examples include French liaison (see CHAPTER 112: FRENCH LIAISON), linking [r] in non-rhotic dialects of English (Hay and Sudbury 2005; among many others), and Maori verbal forms (Blevins 1994). Interestingly, such cases may involve a re-analysis of historically word-final consonants as epenthetic consonants (Vennemann 1972). Finally, final consonants are subject to resyllabification with a following initial segment. Both C/Ø alternations and resyllabification fall under the scope of external sandhi phenomena.

2 Segmental immunity: Generalizations

The immunity of final consonants emerges in static segmental distributions in the lexicon (§2.1) and in the asymmetrical application of segmental processes (§2.2).

More consonants are licensed in final than in internal coda position, allowing additional segmental slots (size effects) or a wider range of place, manner, or laryngeal contrasts (feature effects).

2.1 In the lexicon

Eastern Ojibwa (Piggott 1991, 1999) and Tojolabal (Supple and Douglass 1949; Lombardi 1995) exemplify increased licensing possibilities in final position in manner and laryngeal features. In Ojibwa, while nasals and fricatives are permissible codas in all positions in the word (3a), stops are only allowed word-finally (3b). In Tojolabal, the contrast between plain/aspirated and laryngeal stops and affricates is neutralized in word-internal codas, where only plain segments appear (4a), but remains active in onsets and word-finally (4b).

(3) a. [baŋgisin] 'it falls'
[moːʃkineː] 'it is full'
[wiːjaːs] 'meat'
b. [nindib] 'my head'
[ninik] 'my arm'

(4) a. [hutp'in-] 'to push'
[ʔatnija] 'you bathed'
b. [potot'] 'class of plant'
[k'ak] 'flea'

French illustrates the presence of additional consonantal slots word-finally. While it admits a large variety of final clusters of up to four consonants (5), all morpheme-internal clusters may be analyzed with codas limited to one consonant (Dell 1995). Final clusters include sequences of rising sonority, in violation of the Sonority Sequencing Principle (e.g. Clements 1990; CHAPTER 49: SONORITY). This can be taken as a further indication that final consonants are not regular codas: they exceed the possibilities offered by the syllable template applicable elsewhere in the word not only in terms of the number of segments, but also in their relative autonomy with respect to general syllabic principles.

(5) [adɔpt] 'adopt'
[sɛrkl] 'circle'
[ɑ̃bidɛkstr] 'ambidextrous'

English offers different kinds of final exceptionality effects. As in French, more consonants are found finally than in internal codas: up to three in monomorphemic words (e.g. *next*) and four with the addition of word-level suffixes (e.g. *thousandths*) *vs.* only one internally (exceptionally two, as in *empty*; see Borowsky 1986). Unlike French, however, English does not tolerate word-final sequences of an obstruent followed by a non-sonorant. In addition, English displays asymmetries in vowel + consonant combinations. Word-finally, long vowels are followed by any consonant (6a); morpheme-internally, long vowels in closed syllables appear in restricted contexts: before fricative + stop (6b) or a sonorant homorganic with the following onset (6c), often with additional combinatorial constraints. Coronal obstruents enjoy

(12)	a.	/siok-ta/	[siokta]	'our (INCL) chicken'
		/bantil-kon/	[bantilkon]	'inform (BENEFACTIVE)'
	b.	/laigan-ku/	[laiganku]	'my house'
		/wuruŋ-ta/	[wuruŋta]	'our (INCL) language'

Patterns so far have been described in terms of more consonants being *allowed* at the right edge. Another language type has been put forward, which *requires* constituents to end in a consonant. This possibility is instantiated in Yapese, the case most commonly discussed (Piggott 1991, 1999; Broselow 2003; Wiltshire 2003). This language has no internal codas, but a generalized final short vowel deletion process, which results in words ending in a consonant on the surface. (Final long vowels shorten but do not delete.) The status of Yapese as a distinct type is questionable. As in Kayardild (8), vowel deletion applies finally but not internally (at the word level rather than the phrase), leading to the same generalization as other cases of final consonant immunity: consonants are more easily tolerated in final position. The Yapese pattern may be interpreted as favoring vowel deletion to the extent that it results in phonotactically acceptable forms, rather than actively requiring words to end in a consonant. Menominee is another language in which words end in a consonant, lexically or as a result of final vowel deletion (Bloomfield 1962).[2]

3 Segmental immunity: Representations

If final consonants escape the conditions applying to codas in other positions, their identity as codas is called into question or must be qualified. At least four directions have been explored to account for the internal–final asymmetry. (i) One consists in admitting position-specific syllable well-formedness conditions, for example by defining different coda constraints for final and non-final syllables. This approach is often taken at a descriptive level but it has not been favored in analytical work. (ii) Uniform syllabic conditions may be maintained across positions but violated at edges under pressure from independent constraints. Recent Optimality Theory (OT) analyses have often relied on this type of reasoning; see §4 for a discussion of some relevant factors. (iii) Another line of research has explored the idea that syllabic structure is irrelevant in all or some of the final immunity effects, which arise through sequential generalizations. It has been argued, for instance, that final clusters in English and other languages are accounted for with a constraint limiting sequences of consonants to only one place of articulation (with coronals unspecified for place in English) (Iverson 1990; Yip 1991; Lamontagne 1993; see also Burzio 2007). Such a sequential generalization allows a unified account of consonant clusters in all positions in the word. Côté (2000) takes a more radically non-syllabic approach to consonant phonotactics in general, and final edge effects in particular, which are defined in terms of segment sequencing and adjacency to

2 If words may be required to end in a consonant, we should find cases of systematic final consonant epenthesis, instead of vowel deletion. Interestingly, I am aware of no such cases at the word level. However, many examples of *phrase*-final consonant epenthesis are reported (see some examples in Trigo 1988). This asymmetry between the word and the phrase needs to be investigated further, but it suggests that phrase-final epenthesis corresponds to an articulatory closure effect that is not relevant word-finally.

Appendix (13b): It has been argued that final consonants belong to a separate constituent that hosts consonants that do not fit into the coda. This constituent has been variably called "appendix"[4] (e.g. Halle and Vergnaud 1980; Mohanan 1982; Charette 1984; Borowsky 1986; Goldsmith 1990; Iverson 1990; Wiltshire 1994; Booij 1995; Kraehenmann 2001), "affix" (Fujimura and Lovins 1982), and "termination" (Fudge 1969). By stipulation, the appendix is available only in word-final position. This constituent is usually attached to the syllable node, as in (13b); it is alternatively part of the word structure, as a sister to the syllable. Two types of affixes may be distinguished, for non-suffixal and suffixal consonants (Goldsmith 1990; Duanmu 2008).

Defective syllables (13c): Final consonants are by default taken to be part of the syllable headed by the closest preceding vowel. This assumption is regularly challenged by claims that these consonants in fact belong to a separate syllable, one without a pronounced nucleus (e.g. McCarthy 1979; Selkirk 1981; Iverson 1990; Burzio 1994; Dell 1995; Bye and de Lacy 2000; Cho and King 2003). Representational and terminological details abound here. These special syllables have been termed "degenerate," "empty-headed," "minor," "defective," "semi-syllables," and "catalectic." They may or may not contain a nucleus position; the consonants may be onsets, rimes, or segments attached directly to the syllable node. Different types of degenerate syllables may even be distinguished, for example moraic *vs.* non-moraic (Nair 1999), or syllables whose nucleus position is empty *vs.* those whose nucleus is occupied by segmental material shared by the onset (Goad 2002; Goad and Brannen 2003). Final consonants have been considered to be universally onsets, notably in the model of Government Phonology (Kaye 1990; Harris and Gussmann 2002). Others advocate a mixed coda *vs.* onset approach to final consonants, depending on their segmental profile and behavior, and determined on a language-specific basis or even varying within the same language (Piggott 1991, 1999; Goad 2002; Rice 2003).

Attachment to higher prosodic constituents (13d): Final consonants may attach directly to prosodic constituents higher than the syllable, usually the prosodic word (PWd), but also phrasal constituents (e.g. Rubach and Booij 1990; Rialland 1994; Rubach 1997; Wiltshire 1998, 2003; Auger 2000; Spaelti 2002). As a variation on this theme, Piggott (1999) considers that final consonants are codas or onsets of empty-headed syllables *licensed* by, rather than attached to, the prosodic word. Attachment to higher prosodic constituents implies that the relevant domains for final consonant exceptionality are prosodic in nature. This proposal does not directly account for additional contrasts or slots at the end of morphosyntactic constituents, such as the stem (Broselow 2003).

Extraprosodicity (13e): The most prevalent approach to final consonant exceptionality involves the concept of extraprosodicity (or extrametricality; see CHAPTER 43: EXTRAMETRICALITY AND NON-FINALITY). Originally designed to exclude final syllables from stress assignment algorithms (Liberman and Prince 1977), extraprosodicity has been extended to final consonants by Hayes (1980 (citing a presentation by K. P. Mohanan 1982)) for stress and Steriade (1982) for syllabification. Designating final consonants extraprosodic makes them invisible for the

[4] The term "appendix" has also been used to refer to non-moraic "coda" consonants (13f) (Sherer 1994; Zec 2007) or consonants attached directly to the syllable or prosodic word nodes (13d) (Rosenthall and van der Hulst 1999).

purposes of syllabification, stress assignment, and other processes. The consonants are later adjoined to prosodic structure, in conformity with the principle of Prosodic Licensing (Itô 1986), at a stage in the derivation when syllabic constraints are no longer applicable and metrical structure has already been built. Extraprosodicity is subject to the Peripherality Condition, which restricts it to edges of constituents.

Here again, the theme of extraprosodicity allows for numerous variations, regarding its universality and the level at which it operates. Final consonants are claimed to be universally extraprosodic at the lexical level (Borowsky 1986; Itô 1986). At the word level, extraprosodicity is parametrized (Itô 1986) or universal (Piggott 1991). Itô argues that it is turned off post-lexically, but cases of final consonant exceptionality at the phrasal level motivate its possible extension to post-lexical phonology (Rice 1990).

Non-moraic "coda" consonants (13f): In languages with moraic codas, additional final consonants may be represented as non-moraic (Lamontagne 1993; Sherer 1994; Hall 2002; Kiparsky 2003).

The merits and disadvantages of each of these approaches depend in large part on theory-internal considerations. Each must give up on at least one established principle or generalization of phonological theory. The idea of enriching syllable structure with an appendix constituent has encountered some resistance, since it involves position-specific syllable architecture. Little evidence has been adduced to motivate the appendix as a constituent, which would be expected to act as a trigger or target of some phonological processes; the only case known to me is Mohanan's (1982) suggestion that [r] depalatalizes in the appendix position in Malayalam. Attachment to higher prosodic constituents violates the principle of exhaustive syllabification, as well as strict layering of prosodic constituents. Extraprosodicity requires multiple levels of syllabification, in itself a contentious issue, and may be interpreted as a weakening of the principles of prosodic phonology (Piggott 1999). On the other hand, it avoids syllabic constituents that are otherwise unnecessary (Steriade 1982). Degenerate syllables imply a higher level of abstractness; empty syllabic positions are either viewed as going against the "uncontroversial assumption that syllables must have nuclei" (Rubach 1997: 570–571) or, more positively, as a natural consequence of the phonological architecture in which the segmental and suprasegmental structures are independent of each other (Harris and Gussmann 2002).

Beyond conceptual considerations, at least three issues must be addressed by all approaches relying on specific representations for final consonants. One issue concerns the featural or combinatorial restrictions that additional consonants allowed at the right edge may themselves be subject to, and how these should be expressed. Final obstruent + sonorant sequences occur in French, but not in English. Germanic languages are also well known for allowing word-final strings of voiceless coronal obstruents. The representations in (13b)–(13f) do not make explicit predictions as to the range of consonants they may host, with the exception of the onset approach, according to which final consonants are expected to display onset-like properties. Final consonants or clusters in many languages do have an onset or coda-onset profile (e.g. French; Dell 1995). But many other patterns appear more challenging for the onset approach, as final consonants are regularly much more limited than onsets. In particular, the claim that final consonants are universally onsets is not readily compatible with languages in which

(20) a. No codas, no final empty nuclei Type (15c)
b. No codas, final empty nuclei Type (15a)
c. Codas, no final empty nuclei Type (15e)
d. Codas, final empty nuclei Type (15b)

Crucially, this typology excludes the other symmetrical pattern (15d), in which internal codas and final consonants obey the same constraints. This appears too restrictive, as argued by Piggott (1991). In response, Piggott (1999) proposes another parameter based on the notion of remote licensing. All segments must be licensed by a higher prosodic category, either directly by the syllable, or indirectly (remotely) by the PWd or a phrasal constituent. Piggott, unlike Kaye, allows final consonants to be either codas or onsets, depending on their segmental profile; final onsets, which escape coda restrictions, are always licensed remotely, final codas may be licensed by the syllable or a higher constituent, and vowels must be licensed directly by the syllable. Languages vary in whether remote licensing is excluded (all final segments are either vowels or codas), possible (final segments are vowels, codas, or onsets), or obligatory (final segments must be consonants). This last option, reminiscent of McCarthy's (1993) FinalC, derives Yapese generalized apocope. Unlike Coda Licensing, Piggott's parametric approach does not provide for languages of type (20c); it also predicts that final consonants that exceed the coda template display onset-like properties, which is not always the case, for example with final coronal obstruents in Germanic languages.

An OT account similar in spirit to that of Piggott has been proposed by Spaelti (2002). According to his WeakEdge family of constraints, the right edge of a constituent should contain as little prosodic structure as possible. These constraints favor the attachment of final segments to constituents higher than the syllable in the prosodic hierarchy. Since only consonants may be so attached, WeakEdge establishes consonants as the preferred segment type in final position.

Piggott's and Spaelti's proposals rely on the idea that constituents *should* end in a consonant, echoing the constraint Final-C mentioned above, and specifically in a non-coda consonant. Goad (2002) argues that final non-codas are indeed advantageous from a processing viewpoint. Final consonants that are not possible internal codas signal the right edge of words more clearly than final vowels or codas do, since they cannot appear syllable-finally inside words. Likewise, codas signal the right edge of syllables better than vowels do. This parsing argument needs to be tested; for now, two questions arise. First, if it is a desirable thing for words to end in onsets, one should expect to find more cases of generalized word-final vowel deletion or epenthesis of a non-coda consonant. As noted above, many languages actually require that words, but not syllables, end in a vowel. Second, if codas are the best indicators of the right edge of syllables, why are they considered marked in syllable typology?

4.4 Perception and adjacency to prosodic boundaries

The syllabic basis of consonant phonotactics has been questioned in the last decade or so (Steriade 1999a, 1999b; Blevins 2003), in particular by proponents of the "licensing by cue" approach, according to which the likelihood that a feature or segment occurs in a given context is a function of its relative perceptibility in that context (chapter 98: speech perception and phonology). Côté (2000) applies this idea to segmental immunity at the right edge, arguing that the additional licensing possibilities in peripheral positions are motivated by perceptual factors. More consonants are tolerated at edges, because their perceptibility is enhanced by a number of phonetic processes: lengthening, articulatory strengthening, and reduction of the amount of overlap with adjacent segments.

The formal architecture is based on two constraint families that require consonants to be followed by a vowel (21a) or adjacent to a vowel (21b), contexts where they benefit from optimal transitional cues. But consonants with stronger internal or contextual cues are less dependent on vocalic transitions and vowel adjacency. This includes final consonants, which are subject to the more specific constraints in (21c) and (21d), where *i* ranges over the set of prosodic boundaries at the word level and above, including Ø for word-internal consonants (which are not adjacent to any boundary).

(21) a. $C \rightarrow V$ A consonant is followed by a vowel.
b. $C \leftrightarrow V$ A consonant is adjacent to a vowel.
c. $C]_i \rightarrow V$ A consonant followed by a prosodic boundary i is followed by a vowel.
d. $C]_i \leftrightarrow V$ A consonant followed by a prosodic boundary i is adjacent to a vowel.

It is assumed that the higher the prosodic boundary a consonant is adjacent to, the more easily it surfaces without the support of an adjacent vowel; consonants not adjacent to any prosodic boundary are the weakest. This is expressed in the rankings in (22), which follows the three-way distinction between phrase-final, word-final, and (word-)internal consonants established in §2.2. Syllable well-formedness and extraprosodicity are irrelevant concepts in this framework, but $C \rightarrow V$ and $C \leftrightarrow V$ obviously bear similarity to, but are not equivalent to, NoCoda and NoComplexCoda/Onset, respectively.

(22) a. $C]_{\emptyset} \rightarrow V >> C]_{PW} \rightarrow V >> C]_{Ph} \rightarrow V$
b. $C]_{\emptyset} \leftrightarrow V >> C]_{PW} \leftrightarrow V >> C]_{Ph} \leftrightarrow V$

This approach derives the Kamaiurá and Cairene Arabic patterns. In Kamaiurá, NoDeletion is ranked between $C]_{\emptyset} \rightarrow V$ and $C]_{PW} \rightarrow V$; word-internal consonants have to be followed by a vowel, but word-final ones survive (23). It also directly accounts for cumulative immunity effects, as in Basque (10), with the rankings in (22). Note that C in these constraints may be restricted to specific categories or features (e.g. stops or [coronal] consonants).

(23)

o-mo-kon-mo-kon	$C]_{\emptyset} \rightarrow V$	NoDeletion	$C]_{PW} \rightarrow V$
a. omokonmokon	*!		*
b. omokomoko		**!	
☞ c. omokomokon		*	*

Broselow (2003) interprets the existence of exceptionality effects involving the stem – a morphosyntactic constituent – as contradicting the prosodic basis of the perceptual account. Moreover, enhancement effects are strongest at the phrase

4.6 *Summary and discussion*

The proposals in §4.1–§4.5 can be compared along two empirical dimensions: (i) do they adequately account for all types of final immunity?; (ii) are they compatible with other phonotactic patterns involving the internal–final relationship? Concerning the first question, alignment constraints, due to their variety, are capable of deriving the full range of immunity effects; parameters such as Coda Licensing and Remote Licensing also account for additional consonantal slots, but their onset status appears at odds with the limited range of final consonants often tolerated beyond syllabic possibilities. Positional faithfulness struggles with final consonant immunity obtained though final vowel deletion; ANCHOR-based analyses also fail to protect final clusters by targeting only the last segment. Prosodic approaches raise the issue of immunity effects at the end of morphosyntactic constituents. Finally, it remains to be seen how reasonably morphology can embrace all cases of final consonant immunity.

Beyond final immunity effects, Coda and Remote Licensing are integrated into parametric systems that claim to account for the full typology of internal–final phonotactic patterns, but the former fails to provide for symmetric languages of type (15d), where final consonants display a typical coda profile, and the latter ignores languages that require words to end in a vowel (15e). Alignment constraints probably offer the most flexible framework and derive the full range of patterns in (15). In fact, the flexibility of alignment constraints is such that final immunity effects enjoy no special status: consonant sequences are as likely to be more complex internally as to be more complex finally. This position might be argued to lack restrictiveness or explanatory power or, conversely, better reflect the range of formally possible patterns, depending in part on the status of type (15f), with complex codas productively allowed only inside constituents. Approaches based on positional faithfulness, perception, or morphology make no specific claims with regard to phonotactic patterns other than final immunity, especially those involving more complex internal codas, which need to be derived by independent constraints or factors. The requirement that words end in vowels might be interpreted as a morphological constraint, but type (15f) would seem to be more challenging. Progress in the analysis of final (and internal) immunity effects rests on a deeper understanding of the patterns in (15), how they arise diachronically, and what factors they are sensitive to.

5 Metrical invisibility

As noted in (2b), final exceptionality also manifests itself prosodically, final CVC syllables patterning like CV ones in stress assignment and vowel length alternations. While regularly noted, the metrical invisibility of final consonants has not given rise to the same analytical diversity as phonotactic immunity. Whether or not metrical invisibility and phonotactic immunity are amenable to a unified approach is also unclear: despite some attempts at a common analysis, there is evidence that the segmental and metrical manifestations of final exceptionality should be kept separate. §5.1 presents the relevant generalizations underlying metrical invisibility and its relationship with segmental immunity; §5.2 addresses its functional motivation.

5.1 Generalizations

In quantity-sensitive stress systems, syllable weight is normally determined by the segmental make-up of the rime. In some cases weight assignment also depends on the position of the rime in the word. Of interest here is the pattern where CVC syllables count as light in final position but as heavy elsewhere. See Hayes (1995: 57), Lunden (2006: 1), and Gordon *et al.* (2010: 142) for lists of languages in which final CVC patterns as light.

Various dialects of Arabic illustrate this effect, among them Cairene (Hayes 1995: 67–71; see also CHAPTER 124: WORD STRESS IN ARABIC). Cairene has three syllable types: light CV, heavy CVC and CVː, and superheavy CVCC and CVːC, found only in final position. Stress falls on final superheavy syllables (24a), otherwise on heavy penults (24b), otherwise on the penult or antepenult, according to a complex algorithm not described here. This pattern reveals an equivalence between final CVCC/CVːC and penultimate CVC/CVː, which regularly attract stress, and between final CVC and penultimate CV, which do not. This is straightforwardly accounted for if word-final consonants are ignored in the computation of weight and stress.

(24) a. [ka'tabt] 'I wrote'
[haʤ'ʤaːt] 'pilgrimages'
b. [ka'tabta] 'you (MASC SG) wrote'
[haː'ðaːni] 'these (MASC DUAL)'
[mu'darris] 'teacher'

As shown in (9) and (24), final consonants are simultaneously invisible to syllabic restrictions and stress assignment, which makes Cairene Arabic a textbook example of final consonant exceptionality, regularly discussed since Broselow (1976) and McCarthy (1979). Such data make it tempting to interpret segmental immunity and metrical invisibility as two effects of a single phenomenon. Yet the two must be distinguished. In Cairene, stress assignment applies at the word level, but the CVC syllable template is enforced at the phrasal level, after resyllabification of word-final consonants with a following vowel, as in (25a). Crucially, stress remains on the second syllable in [ʃi'ribt] and the first in ['katab] (25b), even though on the surface the second syllable is assumed to be CVC in both cases.

(25) a. /ʃiribt ahwa/ [ʃi.'rib.'tah.wa] 'you drank coffee' (Selkirk 1981: 222)
b. /katab lilmalik/ ['ka.tab.lil.'ma.lik] 'he wrote to the king' (Broselow 1976: 16)

Other evidence of the independence of segmental immunity and metrical invisibility can be found in the stress pattern of Hindi-Urdu (Hayes 1980; Hussain 1997). The relevant facts resemble those of Cairene Arabic, with stress falling on final superheavy syllables, otherwise on the rightmost heavy. Unlike Cairene, however, superheavy syllables are not restricted to the final position. Final consonants are metrically invisible, since CVC attracts stress only in non-final syllables. But segmental immunity is not involved, since complex codas occur in all positions.

Like stress, vowel length is sensitive to syllable shape, vowel shortening typically occurring in closed syllables and lengthening in open syllables. Again, final consonants appear to be ignored in some languages, with lengthening applying in final CVC syllables or shortening applying only in CVCC ones. Icelandic (Gussmann 2002) regularly stresses the initial syllable of the word. The stressed vowel lengthens in open syllables (26a) and in monosyllables closed by only one consonant (26b), but no lengthening is observed in non-final closed syllables (26c) and in monosyllables closed by two or more consonants (26d). This is a straightforward case of final consonant invisibility: final CVC patterns like non-final CV, and final CVCC like non-final CVC. Similar length alternations are observed in Swiss German (Spaelti 2002) and Menominee (Milligan 2005).

(26)	a.	['puː]	'estate'
		['staːra]	'stare'
	b.	['pruːn]	'edge'
		['θaːkʰ]	'roof'
	c.	['sɛnta]	'send'
		['flaska]	'bottle'
	d.	['tʰjalt]	'tent'
		['rixs]	'rich (GEN SG MASC)'

English shows both stress and length effects of final metrical invisibility. In verbs, final CVCC attracts stress (*u'surp, tor'ment*), but CVC does not (*'edit, de'velop*). The stress-attracting power of internal CVC is, however, visible in nouns (*a'genda, a'malgam*) (Hayes 1982). With respect to length, long vowels in final CVːC syllables regularly correspond to short vowels in final CVCC or non-final CVC, after the addition of a consonantal or syllable-size suffix (*keep–kept, wide–width, five–fifth–fifty, wise–wisdom, intervene–intervention*). This suggests that shortening applies in closed syllables (internal CVC and final CVCC) but spares final CVC, treated as open by virtue of final consonant invisibility.

The representations specific to final consonants discussed in §3 have also been used to derive their metrical invisibility, in particular extraprosodicity (13e) and non-moraic coda consonants (13f). The latter directly accounts for invisibility in the context of stress assignment, if stress depends on syllable weight and weight on moraic structure. Extraprosodicity for metrical purposes is generally kept distinct from phonotactically motivated extrametricality (e.g. Hayes 1995: 106), echoing the remarks above on the non-equivalence between metrically invisible and segmentally immune consonants. Iverson (1990) in fact argues that extraprosodicity should be restricted to stress and excluded from the segmental domain, with cases of segmental immunity and vowel length alternations re-analyzed as involving some of the other devices mentioned in §3: appendices, empty-headed syllables, and sequential cluster constraints (see also Lamontagne 1993).

5.2 *Motivation*

The lightness of final CVC syllables is motivated by the avoidance of final stress, embodied in OT by the constraint Non-Finality (Prince and Smolensky 2004), which excludes final stressed syllables or head feet (see CHAPTER 43: EXTRAMETRICALITY AND NON-FINALITY). Non-Finality, in conjunction with other constraints, correctly derives stresslessness on final CVC and stress on final CVCC and CVVC in Arabic dialects (Rosenthall and van der Hulst 1999).

If NON-FINALITY generates the facts, the question remains why final stress should be avoided or why final CVC is treated as light. A number of proposals functionally related to the special status of final CVC have recently been put forward. Ahn (2000), Lunden (2006), Hyde (2009), and Gordon *et al.* (2010) offer explanations that, although distinct, are all related to final lengthening.[10] Ahn suggests that the increased vowel duration resulting from final lengthening jeopardizes the contrast between short and long vowels by making final short vowels comparable in duration to non-final long vowels. Stressing the final vowel would weaken the length contrast even further. Lunden and Gordon *et al.* develop a duration-based account of syllable weight, according to which syllables count as heavy (and attract stress) if their rime is sufficiently longer than that of light syllables. The relative difference in duration between internal CVC and CV is sufficient for CVC to be categorized as heavy, but this may not be the case finally, where final lengthening reduces significantly the durational ratio of CVC to CV. Gordon *et al.* (2010) also reveal that the languages that asymmetrically treat final CVC as light lack vowel length contrasts in final syllables. It is proposed that phonetic final lengthening tends to be more pronounced when no length contrasts need to be maintained, making it more likely that CVC be interpreted as light.

Hyde (2009) focuses on certain properties of final lengthening rather than on duration itself. He notes that, unlike initial lengthening, final lengthening is typically associated with tempo deceleration and declining intensity. These characteristics make final position less compatible with stress, either because diminished intensity makes stress more difficult to perceive, or because the intensity that typically accompanies stress makes it more difficult to decelerate. See Hyman (1977) and Gordon (2001) for related ideas. Gordon invokes intonational factors, final stress being avoided because it would result in the high tone associated with stress and the low final boundary tone being realized on the same syllable. Note that these different factors – duration, final lengthening, length contrasts, tones, deceleration, and intensity – are potentially complementary rather than contradictory in explaining the distinction between final stressless CVC and stressed CVCC/CVVC.

6 Conclusions

Final consonants are implicated in a multiplicity of data and analytical approaches, involving a variety of representations, constraints, and parameters. Among the relevant empirical domains, consonant phonotactics has largely dominated the debates, and the stresslessness of final CVC has drawn some attention, while vowel length alternations have been relatively neglected. Analytically, no unified conception of final consonant exceptionality has really emerged, despite attempts based on extraprosodicity, which have been challenged by evidence for the independence of metrical invisibility and segmental immunity. In a changing theoretical landscape, discussions have tended to shift over time from issues of representation to motivations, ranging from abstract parameters (coda or remote

10 Final lengthening may also relate to length alternations that treat final CVC as open, as in Icelandic (26). I leave this issue open.

licensing) and OT constraints (alignment, positional faithfulness) to the role of morphology and functional explanations (perceptual factors, final lengthening). These approaches are probably to some extent complementary.

Focusing on phonotactic patterns, several issues remain to be clarified, concerning the typology of final immunity effects and their relationship to internal immunity effects, whereby more consonants are allowed in internal codas than in final position, and the prosodic or morphosyntactic nature of the constituents involved (phrases, words, stems). To what extent can patterns that require constituents to end in a consonant be analyzed along the same lines as those that merely allow additional consonants? What is the status of patterns displaying more complex codas inside constituents? Should internal and final immunity effects be accounted for in a unified and symmetrical fashion or do they involve distinct factors? One difficulty in answering such questions stems from the fact that positional asymmetries in the complexity of consonant sequences may relate to many factors other than the internal–final contrast, including stressed *vs.* unstressed syllables, initial *vs.* non-initial syllables, morpheme boundaries, and coda–onset linking.

ACKNOWLEDGMENTS

Thanks to two anonymous reviewers and the *Companion* editors for helpful comments. I also acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada.

REFERENCES

Ahn, Mee-Jin. 2000. Phonetic and functional bases of syllable weight for stress assignment. Ph.D. dissertation, University of Illinois, Urbana.

Auger, Julie. 2000. Phonology, variation, and prosodic structure: Word-final epenthesis in Vimeu Picard. In Josep M. Fontana, Louise McNally, M. Teresa Turell & Enric Vallduví (eds.) *Proceedings of the 1st International Conference on Language Variation in Europe*, 14–24. Barcelona: Universitat Pompeu Fabra.

Beckman, Jill N. 1998. Positional faithfulness. Ph.D. dissertation, University of Massachusetts, Amherst.

Blevins, Juliette. 1994. A phonological and morphological reanalysis of the Maori passive. *Te Reo* 37. 29–53.

Blevins, Juliette. 1995. The syllable in phonological theory. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 206–244. Cambridge, MA & Oxford: Blackwell.

Blevins, Juliette. 2003. The independent nature of phonotactic constraints: An alternative to syllable-based approaches. In Féry & van de Vijver (2003), 375–403.

Bloomfield, Leonard. 1962. *The Menomini language*. New Haven: Yale University Press.

Booij, Geert. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.

Borowsky, Toni. 1986. Topics in the lexical phonology of English. Ph.D. dissertation, University of Massachusetts, Amherst.

Broselow, Ellen. 1976. The phonology of Egyptian Arabic. Ph.D. dissertation, University of Massachusetts, Amherst.

Broselow, Ellen. 1980. Syllable structure in two Arabic dialects. *Studies in the Linguistic Sciences* 10. 13–24.

Broselow, Ellen. 2003. Marginal phonology: Phonotactics on the edge. *The Linguistic Review* 20. 159–193.

Buckley, Eugene. 1998. Alignment in Manam stress. *Linguistic Inquiry* 29. 475–496.

Burzio, Luigi. 1994. *Principles of English stress*. Cambridge: Cambridge University Press.

Burzio, Luigi. 2007. Phonology and phonetics of English stress and vowel reduction. *Language Sciences* 29. 154–176.

Bye, Patrik & Paul de Lacy. 2000. Edge asymmetries in phonology and morphology. *Papers from the Annual Meeting of the North East Linguistic Society* 30. 121–135.

Charette, Monik. 1984. The appendix in parametric phonology. *Studies in African Linguistics*. Supplement 9. 49–53.

Cho, Young-mee Yu & Tracy Holloway King. 2003. Semisyllables and universal syllabification. In Féry & van de Vijver (2003), 183–212.

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 283–333. Cambridge: Cambridge University Press.

Clements, G. N. 1997. Berber syllabification: Derivations or constraints? In Roca (1997), 289–330.

Côté, Marie-Hélène. 2000. Consonant cluster phonotactics: A perceptual approach. Ph.D. dissertation, MIT.

Côté, Marie-Hélène. 2004. Syntagmatic distinctness in consonant deletion. *Phonology* 21. 1–41.

Dell, François. 1995. Consonant clusters and phonological syllables in French. *Lingua* 95. 5–26.

Derbyshire, Desmond C. 1979. *Hixkaryana*. Amsterdam: North-Holland.

Duanmu, San. 2008. *Syllable structure: The limits of variation*. Oxford: Oxford University Press.

Evans, Nicholas D. 1995. *A grammar of Kayardild*. Berlin: Mouton de Gruyter.

Everett, Daniel L. & Lucy Seki. 1985. Reduplication and CV skeleta in Kamaiurá. *Linguistic Inquiry* 16. 326–330.

Féry, Caroline & Ruben van de Vijver (eds.) 2003. *The syllable in Optimality Theory*. Cambridge: Cambridge University Press.

Fudge, Erik C. 1969. Syllables. *Journal of Linguistics* 5. 253–286.

Fujimura, Osamu & Julie B. Lovins. 1982. Syllables as concatenative phonetic units. Bloomington: Indiana University Linguistics Club.

Goad, Heather. 2002. Markedness in right-edge syllabification: Parallels across populations. *Canadian Journal of Linguistics* 47. 151–186.

Goad, Heather & Kathleen Brannen. 2003. Phonetic evidence for phonological structure in syllabification. In Jeroen van de Weijer, Vincent J. van Heuven & Harry van der Hulst (eds.) *The phonological spectrum*, vol. 2: *Suprasegmental structure*, 3–30. Amsterdam & Philadelphia: John Benjamins.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Oxford & Cambridge, MA: Blackwell.

Gordon, Matthew. 2001. The tonal basis of weight criteria in final position. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 36. 141–156.

Gordon, Matthew, Carmen Jany, Carlos Nash & Nobutaka Takara. 2010. Syllable structure and extrametricality: A typological and phonetic study. *Studies in Language* 34. 131–166.

Gussmann, Edmund. 2002. *Phonology: Analysis and theory*. Cambridge: Cambridge University Press.

Hall, T. A. 2002. Against extrasyllabic consonants in German and English. *Phonology* 19. 33–75.

Halle, Morris & Jean-Roger Vergnaud. 1980. Three-dimensional phonology. *Journal of Linguistic Research* 1. 83–105.

Harris, James W. 1983. *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: MIT Press.

Harris, John & Edmund Gussmann. 2002. Word-final onsets. *UCL Working Papers in Linguistics* 14. 1–42.

Hay, Jennifer & Andrea Sudbury. 2005. How rhoticity became /r/-sandhi. *Language* 81. 799–823.

Hayes, Bruce. 1980. A metrical theory of stress rules. Ph.D. dissertation, MIT. Published 1985, New York: Garland.

Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13. 227–276.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hussain, Sarmad. 1997. Phonetic correlates of lexical stress in Urdu. Ph.D. dissertation, Northwestern University.

Hyde, Brett. 2009. The rhythmic foundations of Initial Gridmark and Nonfinality. *Papers from the Annual Meeting of the North East Linguistic Society* 38(1). 397–410.

Hyman, Larry M. 1977. On the nature of linguistic stress. In Larry M. Hyman (ed.) *Studies in stress and accent*, 37–82. Los Angeles: Department of Linguistics, University of Southern California.

Itô, Junko. 1986. Syllable theory in prosodic phonology. Ph.D. dissertation, University of Massachusetts, Amherst. Published 1988, New York: Garland.

Iverson, Gregory K. 1990. The stipulation of extraprosodicity in syllabic phonology. *Language Research* 26. 515–552.

Kaye, Jonathan. 1990. "Coda" licensing. *Phonology* 7. 301–330.

Kiparsky, Paul. 2003. Syllables and moras in Arabic. In Féry & van de Vijver (2003), 147–182.

Kraehenmann, Astrid. 2001. Swiss German stops: Geminates all over the word. *Phonology* 18. 109–145.

Krämer, Martin. 2003. What is wrong with the right side? Edge (a)symmetries in phonology and morphology. Unpublished ms., University of Tromsø (ROA-576).

Lamontagne, Greg. 1993. Syllabification and consonant cooccurrence conditions. Ph.D. dissertation, University of Massachusetts, Amherst.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.

Lombardi, Linda. 1995. Laryngeal neutralization and Alignment. *University of Massachusetts Occasional Papers in Linguistics* 18. 225–247.

Lombardi, Linda. 2002. Coronal epenthesis and markedness. *Phonology* 19. 219–251.

Lunden, Anya. 2006. Weight, final lengthening and stress: A phonetic and phonological case study of Norwegian. Ph.D. dissertation, University of California, Santa Cruz (ROA-833).

McCarthy, John J. 1979. On stress and syllabification. *Linguistic Inquiry* 10. 443–465.

McCarthy, John J. 1993. A case of surface constraint violation. *Canadian Journal of Linguistics* 38. 169–195.

McCarthy, John J. & Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology 1993*. 79–153.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.

Milligan, Marianne. 2005. Menominee prosodic structure. Ph.D. dissertation, University of Wisconsin, Madison.

Mithun, Marianne & Hasan Basri. 1986. The phonology of Selayarese. *Oceanic Linguistics* 25. 210–254.

Mohanan, K. P. 1982. Lexical Phonology. Ph.D. dissertation, MIT. Distributed by Indiana University Linguistics Club.

Nair, Rami. 1999. Syllables and word-edges. Ph.D. dissertation, Northwestern University.

Nash, David. 1980. Topics in Warlpiri grammar. Ph.D. dissertation MIT. Published 1986, New York: Garland.

Nelson, Nicole. 2003. Asymmetric anchoring. Ph.D. dissertation, Rutgers University.

Paradis, Carole & Jean-François Prunet. 1993. A note on velar nasals: The case of Uradhi. *Canadian Journal of Linguistics* 38. 425–439.

Piggott, Glyne L. 1991. Apocope and the licensing of empty-headed syllables. *The Linguistic Review* 8. 287–318.

Piggott, Glyne L. 1999. At the right edge of words. *The Linguistic Review* 16. 143–185.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Rialland, Annie. 1994. The phonology and phonetics of extrasyllabicity in French. In Patricia Keating (ed.) *Phonological structure and phonetic form*, 136–159. Cambridge: Cambridge University Press.

Rice, Keren. 1990. Predicting rule domains in the phrasal phonology. In Sharon Inkelas & Draga Zec (eds.) *The syntax–phonology connection*, 289–312. Chicago: University of Chicago Press.

Rice, Keren. 2003. On the syllabification of right-edge consonants: Evidence from Ahtna (Athapaskan). In Stefan Ploch (ed.) *Living on the edge*, 427–448. Berlin & New York: Mouton de Gruyter.

Roca, Iggy (ed.) 1997. *Derivations and constraints in phonology*. Oxford: Clarendon Press.

Rosenthall, Sam & Harry van der Hulst. 1999. Weight-by-position by position. *Natural Language and Linguistic Theory* 17. 499–540.

Rubach, Jerzy. 1997. Extrasyllabic consonants in Polish: Derivational Optimality Theory. In Roca (1997), 551–581.

Rubach, Jerzy & Geert Booij. 1990. Edge of constituent effects in Polish. *Natural Language and Linguistic Theory* 8. 427–463.

Scheer, Tobias. 2008. Why the prosodic hierarchy is a diacritic and why the interface must be direct. In Jutta M. Hartman, Veronika Hegedüs & Henk van Riemsdijk (eds.) *Sounds of silence: Empty elements in syntax and phonology*, 145–192. Amsterdam: Elsevier.

Selkirk, Elisabeth. 1981. Epenthesis and degenerate syllables in Cairene Arabic. *MIT Working Papers in Linguistics* 3. 209–232.

Selkirk, Elisabeth. 1982. The syllable. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part II, 337–383. Dordrecht: Foris.

Sherer, Timothy. 1994. Prosodic phonotactics. Ph.D. dissertation, University of Massachusetts, Amherst.

Spaelti, Philip. 2002. Weak edges and final geminates in Swiss German. Unpublished ms., University of California, Santa Cruz (ROA-18).

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT.

Steriade, Donca. 1999a. Alternatives to syllable-based accounts of consonantal phonotactics. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.) *Item order in language and speech*, 205–245. Prague: Karolinum Press.

Steriade, Donca. 1999b. Phonetics in phonology: The case of laryngeal neutralization. *UCLA Working Papers in Linguistics* 2: *Papers in Phonology* 3. 25–146.

Supple, Julia & Celia M. Douglass. 1949. Tojolabal (Mayan): Phonemes and verb morphology. *International Journal of American Linguistics* 15. 168–174.

Trigo, Loren. 1988. On the phonological behavior and derivation of nasal glides. Ph.D. dissertation, MIT.

Vennemann, Theo. 1972. Rule inversion. *Lingua* 29. 209–242.

Wiese, Richard. 1996. *The phonology of German*. Oxford: Clarendon Press.

Wiltshire, Caroline. 1994. Alignment in Cairene Arabic. *Proceedings of the West Coast Conference on Formal Linguistics* 13. 138–153.

Wiltshire, Caroline. 1998. Extending ALIGN constraints to new domains. *Linguistics* 36. 423–467.

Wiltshire, Caroline. 2003. Beyond codas: Word and phrase-final alignment. In Féry & van de Vijver (2003), 254–268.

Yip, Moira. 1991. Coronals, consonant clusters, and the coda condition. In Carole Paradis & Jean-François Prunet (eds.) *The special status of coronals: Internal and external evidence*, 61–78. San Diego: Academic Press.

Zec, Draga. 2007. The syllable. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 161–194. Cambridge: Cambridge University Press.

37 Geminates

Stuart Davis

1 Introduction

The term "geminate" in phonology normally refers to a long or "doubled" consonant that contrasts phonemically with its shorter or "singleton" counterpart (see also CHAPTER 47: INITIAL GEMINATES). Such contrasts are found in languages like Japanese and Italian, as exemplified by the minimal pairs in (1) and (2), respectively.[1] Languages such as English and Spanish do not have geminates.

(1) *Japanese geminate contrast* (Tsujimura 2007)[2]

a. [saka] 'hill'
b. [sakka] 'author'

(2) *Italian geminate contrast*

a. [fato] 'fate'
b. [fatto] 'fact'

The issue of the phonological representation of geminates has engendered much controversy over the past thirty years. The main issue revolves around how to distinguish formally a geminate consonant from its singleton counterpart in a way that captures the cross-linguistic phonological patterning of geminate consonants. The featural representation of geminate consonants posited in Chomsky and Halle

[1] Languages with geminates vary considerably with respect to the durational difference between the geminate and its singleton counterpart. Idemaru and Guion (2008) report a 3:1 ratio in the duration of geminates to singletons in Japanese but only a 1.8:1 ratio for Italian. They further note that there may be other phonetic cues to geminates besides consonantal duration. These include pitch and intensity differences that may provide secondary acoustic cues to a geminate. However, this chapter will not focus on the phonetic properties of geminates, nor on the issue of which types of consonants are more likely to be geminated (but see Pycha 2007, 2009 and Kawahara 2007 for discussion on these issues). Instead, this article will focus on the phonological behavior of geminates and the matter of their representation in phonology.

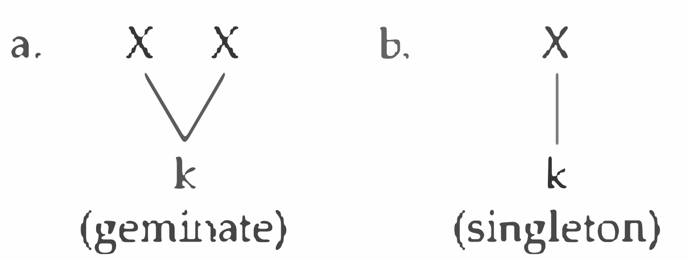
[2] In this chapter, geminate consonants are transcribed by a sequence of two identical letters; long vowels are represented either as a sequence of two identical vowel symbols or with the IPA length mark.

(1968) as being a single consonant possessing the distinctive feature [+long] has long been considered insufficient, since, as noted by researchers such as Leben (1980), long consonants can behave like a sequence of two consonants for certain phenomena. Leben posited an autosegmental representation of geminates in which a single phoneme is linked to two slots on a skeletal tier that encodes the prosody of the word. This skeletal tier is also referred to as a CV-tier, an X-tier, or a length tier, depending on the specific conception of the researcher. Important earlier works that incorporate a CV-tier include McCarthy (1979, 1981), Halle and Vergnaud (1980), Clements and Keyser (1983) and Hayes (1986), while Levin (1985) posited that the tier consisted of X-slots (see CHAPTER 54: THE SKELETON). Geminate representation on this view is exemplified by the geminate [kk] of the Japanese word in (1b), as is illustrated in (3).

(3) a. *CV-tier representation* b. *X-tier representation*

As seen in (3), a geminate consonant has one set of features indicated by the single consonant "k" on the phoneme (or melody) tier, whereas it is linked to two slots on a prosodic tier. In (4), we make clear the distinction between a geminate and a singleton using an X-tier that encodes prosody.

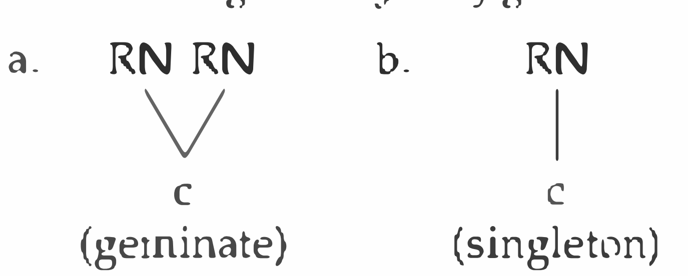
(4) *Prosodic length analysis of geminates*



While the proposals for the representation of geminates in (4) go back thirty years, this representation is specifically argued for by Ringen and Vago (2010), who refer to (4) as the segmental length analysis of geminates.

A different representation of geminates from that in (4) is the two-root node analysis of geminates posited by Selkirk (1990) shown in (5). The root node in a feature-geometric framework indicates the major class features of a sound (McCarthy 1988) and it dominates the rest of the specified features. Every phoneme has a root node, but a geminate under this view has two root nodes (RN = root node, c = consonant).

(5) *Prosodic length analysis of geminates*



There are at least two main differences between the two-root node analysis of geminates in (5) and the segmental length analysis in (4). First, unlike the X-slots epenthesis pattern. On the other hand, if a geminate is represented as moraic, as in (6a) and (7d), epenthesis might not be predicted to occur with a word ending in a geminate, since the consonantal length of a geminate is not segmentally encoded. That is, there would not be two C-slots or two consonantal elements at the end of the word to trigger the epenthesis. Ringen and Vago point out that the Hungarian epenthesis pattern poses a problem for the moraic view. Further, given the weight analysis of geminates in (7d), geminate consonants are predicted to play a role in processes that are sensitive to syllable weight even when singleton (coda) consonants do not. Much of the recent research on geminates has focused on whether geminates display weight properties that are independent of other consonants. This will be discussed shortly.

Over the past twenty years, a wide variety of phonological evidence has been brought to bear on the correct representation of geminates. The issue is still controversial.[3] All three views of geminate representation presented in this section, namely the prosodic length view in (4), the two-root node view in (5) and the moraic weight view in (6), have been argued for on the basis of the phonological patterning of geminates. Some composite views have even been proposed that combine aspects of the above representations, such as those of Schmidt (1992), Hume *et al.* (1997) and Curtis (2003). In §2 we will present specific evidence from a variety of phenomena to argue for the inherent weight representation of geminates. In §3, we will examine the behavior of geminates with respect to stress processes, cross-linguistically. In these sections, we will try to maintain a consistent view for the weight analysis in (6a) even when the data presented seem problematic for such a view. In §4 we will reconsider the representational issue and suggest that a composite view of the representation of geminates under a constraint-based approach can account for the patterning of geminates in the world's languages.

2 The weight analysis of geminates

The underlying weight analysis of geminate consonants, as proposed in Hayes (1989), views a geminate consonant as being underlyingly moraic, as shown in (6a), whereas a non-geminate consonant is underlyingly non-moraic, as in (6b). The weight representation of geminates in (6a) has a number of implications, which will be discussed in this section. One such implication is that if geminates are inherently moraic, they should count as moraic in considering minimal word effects: that is, the cross-linguistically common requirement that content words be at least bimoraic. In §2.1 we show that this is the case for Trukese. A specific structural aspect of the weight representation in (6a) is that geminates do not entail a double linking to two C-slots as in the length representation. This implies that there should be cases in which geminates do not pattern with a sequence of consonants. §2.2 discusses cases of the asymmetrical patterning of geminates and consonant

[3] The controversy over geminates has fostered a number of dissertations with a focus on the phonology of geminates. Some of the important ones include Sherer (1994), Ham (1998), Keer (1999), Morén (1999), Kraehenmann (2001), Muller (2001), Curtis (2003), and Topintzi (2006). Although space does not allow me to discuss the wide variety of interesting issues and proposals that are raised in these dissertations and the different positions that are taken, some issues raised in these dissertations will be brought up in the course of this chapter.

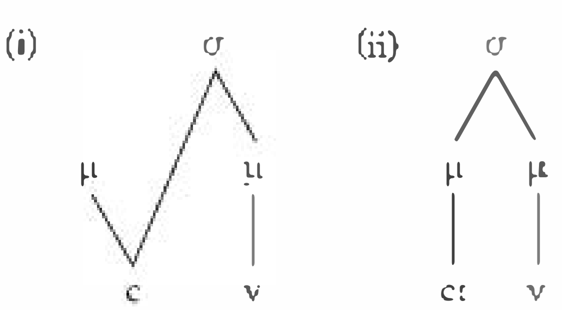
sequences. A third implication that emerges from the weight representation in (6a) is the prediction that there should be languages that treat syllables closed by a geminate (CVG) as heavy but do not otherwise treat syllables closed by a (coda) consonant (CVC) as heavy. In §2.3, we will provide evidence for this prediction by discussing languages that avoid long vowels in syllables closed by a geminate (CVVG), but do not generally avoid long vowels in closed syllables (CVVC). We hold off until §3 the discussion of geminate behavior in weight-sensitive stress systems.

2.1 *Trukese initial geminates*

One type of evidence for the underlying moraic nature of geminates as in (7d) comes from the bimoraic minimal word requirement in Trukese (also called Chuukese) and the behavior of word-initial geminates with respect to it. Although word-initial geminates are rare, they are attested in a number of languages. (Indeed, the dissertations of Muller 2001 and Topintzi 2006, 2010 are exclusively on initial geminates; see also CHAPTER 47: INITIAL GEMINATES.)[4] Muller (2001), whose study incorporates acoustic analyses of word-initial geminates in a variety of languages, including Trukese, concludes that initial geminates are moraic in some languages but not in others, while Topintzi (2006, 2008, 2010), focusing on languages where initial geminates pattern as moraic, argues that such geminates constitute moraic onsets, thus providing examples in which onsets carry weight.[5] Trukese provides a clear example of a language where a word-initial geminate patterns as moraic. Consider the data in (8) and (9), which reflect a minimal word constraint on Trukese nouns. The data here are cited from Davis (1999b) and Davis and Torretta (1998), and are mainly taken from Dyen (1965) and Goodenough and Sugita (1980). The relevance of Trukese geminates for moraic phonology has previously been observed by Hart (1991) and Churchyard (1991).

[4] It is clear from typological surveys of geminate consonants such as Thurgood (1993) and from the discussion in Pajak (2009) that geminates are most commonly found in intervocalic position and least commonly found when not adjacent to any vowel (e.g. between two consonants). Languages that allow for geminates that are only adjacent to one vowel (e.g. word-initial or word-final geminates), although not common, are not as rare as languages that allow for geminates to occur not adjacent to any vowel. As noted by Pajak (2009), the typological facts correspond to perceptual saliency in that the contrast between a singleton and a geminate consonant is most perceptually salient in intervocalic position and least salient in a position not adjacent to any vowel.

[5] Following a suggestion in Hayes (1989), Davis (1999b) proposes that word-initial geminates are moraic but that the mora is not part of the syllable onset. His representation is in (i), while Topintzi's moraic onset representation is given in (ii) (where the vowel of the syllable is also shown).



One difference between (i) and (ii) is that the latter predicts that onset geminates could occur word-internally, not just at the beginning of the word. In support of (ii), Topintzi (2008) provides interesting evidence from Marshallese that word-internal geminates are syllabified as onsets and are not heterosyllabic as commonly assumed in Moraic Theory.

2.3 *Avoidance of CVVG syllables*

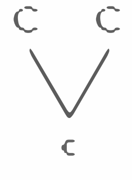
Hayes's theory of underlying moraic representation was presented in (7): geminate consonants (7d) differ from singleton consonants (7c) in being underlyingly moraic. Furthermore, Hayes distinguishes long vowels from short vowels by representing the former as underlyingly bimoraic and the latter as monomoraic. Moraic Theory has an implication for the patterning of geminates with respect to weight-sensitive processes – an implication not discussed by Hayes (1989), but taken up by other researchers such as Selkirk (1990), Tranel (1991), and Davis (1999a, 2003). That is, there should be languages in which syllables with a long vowel (CVV) and those closed by a geminate consonant (CVG) count as heavy since they would be bimoraic, while CV syllables and CVC syllables (i.e. syllable closed by a non-geminate consonant) would be considered light or monomoraic. This weight distinction is shown in (11) (G = geminate consonant, C = non-geminate consonant).

(11) *Syllable weight distinction based on geminates being underlyingly moraic*

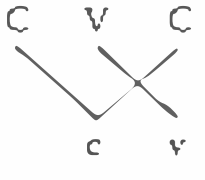
heavy	*light*
CVV	CV
CVG	CVC

The system in (11) is predicted to occur under Hayes's theory in any language with long vowels and geminate consonants that do not regard coda consonants as moraic. The moraic representation for syllables with the structure of (11) is given in (12).

(i) *Geminate in UR: length representation*



(ii) *Epenthesis into a geminate*



Geminate inalterability effects were handled by a condition on interpretation in segmental rules that association lines in structural description of rules had to be interpreted as exhaustive. A rule like spirantization in Tiberian Hebrew, which only applied to singleton consonants and not to geminates, would include in its rule environment a single C-slot linked to the phoneme. Since the rule environment did not explicitly show double linking as in (i), the rule would fail to apply to geminates. Kenstowicz (1994: 410–416) summarizes important criticisms of the CV account of both geminate integrity and inalterability. He points out that geminate integrity could be called into question if epenthesis is viewed as a two-stage process of inserting a V-slot followed by a late default spell-out rule. However, it is worth noting that geminate integrity effects follow automatically from the weight representation of geminates as in (10a). With respect to geminate inalterability, Kenstowicz specifically calls attention to work by Selkirk (1991), who noted that rules of inalterability tended to always involve spirantization processes, thus suggesting a more general explanation that does not involve the length representation of geminates. Along these lines, Kirchner (2000) approaches the issue of geminate inalterability from a general theory of lenition within a functionally based optimality-theoretic framework.

epenthesis apply between the two consonants to create open syllables, thereby avoiding any trimoraic syllable.

In the Arabic dialects, the rule (or constraint) of Weight-by-Position applies, resulting in a moraic coda. However, if we consider the weight division in (11) in which Weight-by-Position does not apply, we would expect to find a language where CVVG syllables but not CVVC syllables are avoided. A potential CVVG syllable would not surface, since it is trimoraic, while a CVVC could still occur, as it would only be bimoraic. Kiparsky (2008a) discusses Swedish dialects where vowel shortening occurs before a geminate but not before a single coda consonant. For example in West Swedish (Kiparsky 2008a: 191) /ruu-dde/ 'rowed' surfaces as [rudde], with its underlying long vowel shortened, but no shortening occurs when a long vowel is before a singleton coda consonant. The shortening before a geminate changes the potential trimoraic CVVG syllable to bimoraic CVG. There is no need for shortening in a CVVC syllable since it would be bimoraic, given that Weight-by-Position does not apply. Kiparsky specifically uses the Swedish data to argue for the moraic representation of geminate consonants.

Another language displaying this pattern of shortening is the Dravidian language Koya, brought up by Sherer (1994), based on Tyler (1969), and discussed in Davis (1999a, forthcoming). Koya has long vowels, coda consonants and geminate consonants. Sherer notes that there are words in Koya like those in (14a)–(14c), with a long vowel before a coda consonant. Crucially, as Tyler (1969: 6) observes, there are no words that contain a long vowel before a geminate. They are always short, as in (14d). All Koya data are cited from Tyler (1969), with the page numbers provided. (The transcription of the vowel quality is phonemicized and does not reflect the precise allophonic variant.)

(14)	a.	leːŋga	'calf' (p. 11)	b.	aːɳɖa	'female' (p. 8)
	c.	neːrs	'learn' (p. 76)	d.	ett	'lift' (p. 76)

Sherer additionally notes cases where a stem-final long vowel shortens before a geminate-initial suffix, as the examples in (15) show.

(15)	a.	keː + tt + oːɳɖu	[kettoːɳɖu]	'he told' (p. 39)
	b.	oː + tt + oːɳɖu	[ottoːɳɖu]	'he bought' (p. 38)

This shortening can be viewed as a way of avoiding trimoraic syllables. Shortening does not occur before a non-geminate consonant, as the examples in (16) illustrate.

(16)	a.	naːl + ke	[naːlke]	'tongue' (p. 47)
	b.	tuŋg + anaː + n + ki	[tuŋganaːŋki]	'for the doing' (p. 90)

In (16), a long vowel surfaces before a syllable-final singleton coda consonant. Since vowel shortening occurs before a geminate in (15), the Koya data in (14)–(16) are consistent with the weight system in (11), in which CVV and CVG syllables are bimoraic whereas CVC syllables are light.[8]

[8] Curtis (2003: 169–170) suggests that the lack of word-internal CVVG syllables in Koya may be due to a shortening effect that geminate consonants have on preceding vowels, since the perceptual cues for vowel length can be blurred in CVVG syllables; thus, Curtis maintains that vowel shortening before geminates is independent of the issue of the moraic status of geminates. However, this does not explain cases like Fula in (19) where avoidance of CVVG is achieved by degemination rather than vowel shortening.

While the above examples of Koya and Swedish are cases where vowel shortening occurs in syllables closed by a geminate, there are other languages where vowel-lengthening processes are prevented in CVG syllables but not in CVC syllables. This suggests that in such languages geminates are underlyingly moraic, although coda consonants in general are not. Vowel lengthening then does not apply before a geminate, since that would create a trimoraic syllable. This is illustrated by Seto (Southeastern Estonian), discussed by Kiparsky (2008b). According to Kiparsky, Seto has feet that are required to be trimoraic and such a restriction is normally implemented by foot-final vowel lengthening. As a result, a foot with the underlying sequence CV.CVC surfaces as CV.CVVC. However, given an input structure where the final consonant of the foot is part of a geminate, i.e. CV.CVG, no vowel lengthening occurs. This provides evidence that the geminate is underlyingly moraic: that is, foot-final vowel lengthening need not occur in CV.CVG since the foot is already trimoraic.

A different case that avoids the surfacing of CVVG syllables can be found in the West African language Fula as discussed by Paradis (1988) and Sherer (1994). Fula avoids CVVG syllables by degemination of the consonant but, importantly, it allows for CVVC syllables, as seen in (17).

(17) CVVC *syllables in Fula* (Sherer 1994: 176)

a. kaakt-ɛ 'spittle' b. caak-ri 'couscous'

This language has a suffixation process that triggers the gemination of a root-final consonant. Consider the singular/plural alternations in (18). Because of an active constraint that requires geminates to be [−continuant] in Fula, a root-final continuant segment changes to a stop when it geminates. (I thank Abbie Hantgan for help with the Fula data.)

(18) *Fula morphological gemination* (Paradis 1988: 78)

	stem (SG)	*suffixed form* (PL)	
a.	lɛw	lebb-i	'month'
b.	lɛf	lepp-i	'ribbon'

Of relevance here is that when a long vowel precedes the stem-final consonant, gemination fails to occur, but the stem-final consonant nonetheless is realized as a stop. This is illustrated by the singular/plural alternations in (19).

(19) *Lack of gemination after a long vowel* (Paradis 1988: 80)

	stem (SG)	*suffixed form* (PL)	*expected form*	
a.	laaw	laab-i	*laabb-i	'road'
b.	lɛɛs	lɛɛc-ɛ	*lɛɛcc-ɛ	'bed'

Given that gemination is part of this suffixing process, we note that the expected forms in (19), where the initial syllable would be CVVG, fail to surface as such. Rather, the nature of the occurring suffixed forms in (19) makes it appear that degemination has occurred. This can be understood as the avoidance of a trimoraic CVVG syllable. Since CVVC syllables are allowed as in (17), Fula seems to instantiate a language with the weight system of (11), where CVG syllables are heavy but not other CVC syllables. Thus, we see that a variety of languages have strategies to avoid CVVG syllables but not CVVC syllables. These languages can be understood as providing support for the inherent weight analysis of geminates.

3 The patterning of geminates in stress systems

One of the criticisms of the weight analysis of geminates proposed by Hayes (1989), discussed by Tranel (1991), comes from the observation that there do not seem to be quantity-sensitive stress systems that support the weight division in (11), repeated below as (20), where stress would be attracted onto a syllable with a long vowel or closed by a geminate consonant, but not on a syllable closed by a non-geminate.

(20) *Syllable weight distinction based on geminates being underlyingly moraic*

heavy	*light*
CVV	CV
CVG	CVC

The system in (20) is predicted to occur under Hayes's theory in any language that allows long vowels and geminate consonants, but in which Weight-by-Position does not generally apply to coda consonants. According to the division in (20), CVV and CVG syllables would syllabify as bimoraic, while CV and CVC syllables would syllabify as monomoraic, as was shown in (12). Since quantity-sensitive stress systems single out bimoraic syllables, it would be expected that at least some quantity-sensitive stress systems would reflect the weight division in (20) if the moraic representation of geminates were correct. Tranel suggests that weight systems like (20) do not exist and instead proposes a principle of equal weight for codas. Specifically, in languages in which codas pattern as moraic, geminates will be moraic; but in languages in which codas are not moraic, geminates would not be moraic. While our observation in §2.3 above – that in a variety of languages CVVG syllables are avoided but CVVC are not – can be taken as evidence for the weight division in (20), Tranel's observation is of importance. In this section, we will overview the behavior of geminates in quantity-sensitive stress systems. In §3.1, we will provide stress data from various languages which indeed support the division in (20) whereby CVV and CVG syllables pattern together, thus supporting the moraic weight analysis of geminates. These are languages whose stress patterns Tranel predicts not to occur. In §3.2, I will review the type of case mentioned by Tranel in which quantity-sensitive stress treats all closed syllables in the same manner whether they be CVG or CVC. These are the languages that motivated Tranel's principle of equal weight for codas and can be considered somewhat problematic for the weight analysis of geminates. In §3.3, I will present the case of the Australian language Ngalakgan (variably spelled Ngalagkan and Ngalakan; Baker 1997, 2008), in which CVC syllables can attract stress but apparently not CVG syllables. This seems to suggest that somehow geminates are resistant to carrying a mora. Thus, this section will identify three different types of geminate behavior with respect to quantity-sensitive language that requires exhaustive footing. The data item in (21e) is similar to the last two syllables in (21d) in that the syllable with a long vowel forms a foot on its own and the syllable immediately after it is at the beginning of a new trochaic foot, thereby receiving stress. We see then that bimoraic CVV syllables in Cahuilla are distinguished from monomoraic syllables, not only in bearing stress, but also by the presence of stress on the syllable immediately after them. CV and CVC syllables lack these two characteristics and function as monomoraic. It is interesting in this light to observe the patterning of syllables closed by a geminate, as in (21f), where the first syllable is CVG. This CVG syllable functions as bimoraic. It has stress, as would be expected of any initial syllable, but crucially it patterns exactly like a CVV syllable in that the syllable immediately after it also carries stress. This provides evidence that the CVG syllable comprises a bimoraic (trochaic) foot on its own. This constrasts with the initial CVC syllable in (21c) that forms a trochaic foot with the following syllable, suggesting the monomoraic nature of the CVC syllable. Cahuilla thus serves as a clear illustration for the weight distinction in (20) in which stress treats syllables with long vowels and those closed by geminates as bimoraic but not other types of syllables, be they CV or CVC.

San'ani Arabic (Yemen) presents a very interesting case, in which CVV and CVG syllables pattern together with respect to stress. Watson (2002: 81–82) specifically notes that they pattern together as opposed to CVC syllables (see also CHAPTER 135: WORD STRESS IN ARABIC). Consider the data in (22), which illustrate the stress pattern in words without geminates.

(22) *San'ani Arabic* (Watson 2002: 81–82)

a.	mak.'tu:b	'office'
b.	da.'rast	'I/you (MASC SG) learnt'
c.	'sa:.fa.rat	'she travelled'
d.	mak.'sa:.lih	'launderette'
e.	mi.'gam.bar	'sitting'
f.	'mad.ra.sih	'school'
g.	mak.'ta.ba.ti:	'my library'
h.	'li.bi.sat	'she wore/put on'
i.	'ka.tab	'he wrote'
j.	'ra.ga.ba.tih	'his neck'

Stress normally falls on one of the last three syllables of the word: it falls on a final superheavy syllable (CVVC or CVCC) if there is one, as in (22a)–(22b); it falls on the rightmost non-final heavy syllable (CVC or CVV) up to the antepenultimate, as in (22c)–(22f); otherwise, stress falls on the leftmost CV syllable, as in (22g)–(22j). The data in general show that the word-final segment does not play a role in the computation of weight so that the final syllable can only be stressed if it is superheavy. The word in (22g) illustrates two important aspects of the stress system. It shows that a word-final syllable ending in a long vowel does not attract the stress; it also indicates that a CVC syllable in pre-antepenultimate position fails to attract stress. The latter point is significant, since it suggests that Weight-by-Position, which assigns a mora to a coda consonant, is restricted to one of the last three syllables of the word. Now let us consider the data in (23) with words possessing geminate consonants.

(23) *San'ani Arabic* (Watson 2002: 81–82): *stress on words with geminate consonants*

a. ji.'ħib.bu 'they (MASC) love/like'
b. mit.'ʔax.xi.raːt 'late (FEM PL)'
c. mu.'sadʒ.dʒi.la.ti 'my recorder'
d. 'haː.ka.ða.haː 'like this'
e. 'daw.wart 'I/you (MASC SG) looked for'
f. 'sa.far.t 'I/you (MASC SG) travelled'

The comparison between (22) and (23) indicates the priority of CVG and CVV syllables for stress assignment, in that CVG and (non-final) CVV syllables always attract stress even when in pre-antepenultimate position as in (23c) and (22d). The word in (22g), in contrast, shows that a CVC syllable does not receive stress in pre-antepenultimate position. The difference between CVG and CVC syllables can be readily understood on the inherent weight analysis of geminates. If a geminate is underlyingly moraic, it contributes weight to the syllable regardless of its location in the word. Recall that Weight-by-Position does not apply here, because it is restricted to one of the last three syllables in San'ani Arabic. In pre-antepenultimate position, only CVV and CVG act as bimoraic. Moreover, (23e) shows that CVG syllables have a priority of stress over a final superheavy syllable and should be compared with (22a) where a regular CVC syllable is devoid of such priority. It could be argued that Weight-by-Position in San'ani Arabic only applies to words that would not otherwise have bimoraic syllables (CVV or CVG). That is, there is no necessity for Weight-by-Position to apply in (23e) or (23f). While we do not pursue a full analysis here (but see Watson 2002), the priority given to both CVV and CVG syllables in stress assignment, especially as seen by the comparison of (23c) and (23d) with (22g), provides an interesting argument for the underlying moraic weight analysis of geminates, and, in turn, against Tranel's (1991) claim of equal weight for codas.

We have detailed above two cases where CVV and CVG syllables pattern together with respect to stress systems as predicted by the inherent weight analysis of geminates. Further support for the weight analysis of geminates is found in other languages. For example, Gupta (1987) discusses a Hindi dialect in which stress is attracted to the leftmost heaviest syllable in the word. The dialect treats both CVV and CVG syllables as bimoraic, while CVC syllables behave as light, although, as noted by Curtis (2003), such a pattern appears unusual among Hindi dialects. Additional support may come from the stress system of Pattani Malay, discussed by Topintzi (2006, 2008, 2010) and references cited therein. Pattani Malay has geminates that are restricted to word-initial position and the language lacks long vowels. Although primary stress typically falls on the final syllable of a word, stress occurs on the initial syllable in words that begin with a geminate consonant. This can be taken as evidence for the moraification in (10a) where a geminate is underlyingly moraic. That is, stress is attracted onto a syllable that is bimoraic.

Despite the range of examples presented in this section, it remains rare to see languages that display the weight system in (20), grouping CVV and CVG syllables together as heavy. It is possible, on the other hand, that the rarity is due to the infrequent occurrence of the specific set of properties that is required for CVV and CVG to pattern together in stress assignment; namely, the language would have to have quantity-sensitive stress, long vowels, coda consonants, and geminates. Perhaps when such languages, San'ani Arabic for one, are examined closely, more instances of the special properties of CVG syllables will emerge. In this connection, it is worth noting that in most Arabic dialects CVG syllables are special: they always attract stress when in word-final position. This property separates them from other CVC syllables, which do not attract stress in word-final position. The difference thus finds a logical explanation in the underlying weight analysis of geminate consonants (especially in those dialects, such as Hadhrami Arabic, discussed earlier, which disallow final consonant clusters).[10]

3.2 *Languages in which stress treats all codas equally*

There are two types of languages in which stress assignment treats all codas equally. In the first type, stress is quantity-sensitive and is attracted to a heavy syllable, be it CVV, CVC, or CVG. Latin belongs to this group: any coda consonant makes a syllable heavy, so both CVC and CVG syllables are bimoraic. In the second type, which is more relevant for the representation of geminate consonants, both CVC and CVG syllables behave as light in a quantity-sensitive stress system. In such a language, stress is attracted to a CVV syllable, but both CVC and CVG syllables seem to pattern as monomoraic, treating CVG as light, just like other CVC syllables. As an illustration, consider the stress data from the Uralic language Selkup in (24). The data in (24a)–(24f) come from Halle and Clements (1983). The data items in (24g)–(24h) are reported in Ringen and Vago (2010) from the Selkup language scholar Eugene Helimski, and reflect the Taz Selkup dialect, which seems to have the same stress pattern as that in Halle and Clements (1983).

(24) *Selkup stress*

a.	qu'moːqi	'two human beings'
b.	'uːciqo	'to work'
c.	uː'cɔːmɨt	'we work'
d.	'quminɨk	'human being (DAT)'
e.	'amɨrna	'eats'
f.	'uːcikkak	'I am working'
g.	'ɛsykka	'(it) happens (occasionally)'
h.	ɛs'sɪːqo	'to happen (already)'

[10] In many Arabic dialects, word-final CVC syllables behave as extrametrical. Ham (1998) puts forward the very intriguing observation that final CVC syllables are always extrametrical in languages that possess word-final geminates. This is because a word-final geminate is moraic and would need to be distinguished in final position from a potential moraic coda. With the underlying moraic weight representation of geminates as in (10a), final extrametricality of CVC syllables is able to preserve the contrast between an underlying final geminate and the corresponding final singleton consonant. The geminate of a final CVG syllable would surface as moraic while the singleton coda of the final CVC would be non-moraic. This difference is found in Arabic dialects where a final CVG syllable attracts stress, making it distinct from a final CVC syllable (i.e. bimoraic), which is light (monomoraic) and does not attract the stress. In a variety of other languages having word-final geminates examined by Ham (1998), the same distinction is made between final CVG and CVC syllables. If Ham's observation holds up to further scrutiny, it constitutes an interesting argument for the underlying moraification of geminate consonants. (See also Topintzi 2008: 175 for discussion on this point.)

In Selkup, primary stress falls on the rightmost syllable with a long vowel (24a)–(24c) or on the initial syllable if there are no long vowels (24d). A CVC syllable does not count as heavy (24e), even if the CVC syllable is closed by a geminate, as seen in (24f) and (24g). As noted by Tranel (1991), if stress targets bimoraic syllables and geminates are underlyingly moraic, the second syllable in (24f) and (24g) would be the rightmost bimoraic syllable. Both the vowel and the geminate would contribute a mora to the second syllable. The fact that (24f) and (24g) do not receive stress on the second syllable, however, seems to provide evidence against geminates being underlying moraic, favoring a representation of geminates that is different from that in (10a).

The stress pattern of Selkup does not appear to be unique in ignoring geminate consonants. Davis (1999a: 41) notes the Altaic language Chuvash (Krueger 1961), which exhibits an almost identical stress pattern to that of Selkup: stress is attracted to the rightmost syllable with a full vowel (interpreted as being bimoraic), but CVC syllables are ignored. Thus, in both Chuvash and Selkup, CVC syllables do not function as bimoraic CVV syllables but instead act like monomoraic CV and CVC syllables.

Data from languages like Selkup have been used by Tranel (1991) and Ringen and Vago (2010) to argue against the underlying moraic weight representation of geminates. Ringen and Vago note that such languages are consistent with the length analysis of geminates as in (10b). In these languages, stress is sensitive to the presence of a long vowel, and ignores a coda consonant, whether the coda is part of a geminate or not. However, it is not that proponents of the weight representation are unaware of languages like Selkup. Topintzi (2008, 2010), who for the most part maintains the underlying moraic weight view of geminates, suggests that weightless geminates are represented by double consonants with two root nodes rather than as a single root node linked to a mora like (10a). But such a comment implies that there is language-specific variation in the representation of geminate consonants. On the other hand, Davis (2003) suggests that the stress pattern of languages like Selkup does not necessarily argue against the underlying moraic representation of geminates; viewed from an optimality-theoretic perspective, the pattern can be a consequence of certain high-ranking stress constraints that have the effect of ignoring the bimoraicity of any CVC syllable. As suggested by Steriade (1990: 275), there may be reasons in some languages to restrict the set of stress-bearing segments to those that are also tone-bearing, "for reasons that are clearly related to the fact that pitch is one of the main realizations of metrical prominence." Steriade's suggestion can be incorporated into an optimality-theoretic approach as a constraint that restricts pitch realization to vocalic elements: the constraint prefers to place stress on any CVV syllable over any syllable closed by a consonant, even if that consonant is part of a geminate. Thus, the lack of second syllable stress in (24f) and (24g) of the Selkup data need not reflect on the underlying moraicity of geminate consonants.

3.3 *Languages in which geminates repel stress*

A third type of geminate behavior is witnessed in languages where stress is attracted to a closed syllable, but not to one with a geminate. The Australian language Ngalakgan, discussed by Baker (1997, 2008), serves as a major example. Consider the pattern of primary stress in (25), taken from Baker (2008) who notes that other geographically proximate languages have a similar stress pattern.

(25) *Ngalakgan*

a.	pu'ruʈci	'water python'	k.	'calapir	'red ant (species)'
b.	ki'piʈkuluc	'frogmouth (bird)'	l.	'kupuj	'sweat (N)'
c.	mi'ɽarppuʔ	'crab'	m.	kaɳ'ʈalppuru	'plains kangaroo'
d.	pu'ʈolkoʔ	'brolga (bird)'	n.	'cakanta	'macropod sp.'
e.	ˌmaca'purka	'plant sp.'	o.	'ŋuɽuɳʈuc	'emu'
f.	'ɭaɽkurca	'vine sp.'	p.	'ŋoɭoŋkoʔ	'eucalyptus'
g.	'calpurkic	'fish sp.'	q.	'ŋamucˌculo	'subsection term'
h.	'ciwi	'liver'	r.	'capatta	'tortoise sp.'
i.	'ceraʈa	'women's ceremony'	s.	'moloppoɭ	'catfish sp.'
j.	'paɽaˌmunu	'sand goanna'	t.	'ŋanaʔpaj	'and moreover'

(25a)–(25l) show that primary stress in Ngalakgan falls on the leftmost (non-final) heavy syllable; if there is no heavy syllable, it falls on the initial syllable. From these data, it can be surmised that a coda consonant is moraic in making a syllable heavy, be it an obstruent (as in (25a) and (25b), or a sonorant (as in (25c)–(25g)). The data in (25m)–(25t) show that the leftmost closed syllable (underlined) fails to attract primary stress. Note that the coda in the leftmost closed syllable in (25m)–(25t) belongs to one of three types: in (25m)–(25p) the coda is a nasal homorganic with the following onset, in (25q)–(25s) it is the first part of a geminate consonant, and in (25t) the coda is a glottal stop. Key to our discussion is the fact that CVG syllables in (25q)–(25s) resist stress. However, the comparison of the stress-resistant nature of CVG syllables with the other instances of stress-resistant closed syllables in (25m)–(25p) and (25t) points to the fact that a common property these syllables have is that they do not possess their own place features: either the place features are shared with the following onset, or, in the case of the glottal stop in (25t), there is a lack of place features altogether. Thus, Ngalakgan seems to divide closed syllables into two types: those in (25a)–(25l), in which the coda has independent place features and attracts stress, and those in which the coda does not have its own independent place features and fails to attract stress. This suggests that Ngalakgan is best analyzed as having a requirement that moraic elements have independent place features (i.e. not shared with a following onset), as advocated in Baker (1997). It follows that the stressed closed syllables in (25a)–(25l) would be bimoraic and attract stress, whereas the CVG syllables in (25m)–(25p) would be monomoraic and not attract stress.

Languages like Ngalakgan seem to present a challenge for the underlying weight representation of geminates in (10a), since not only do syllables with geminates not attract stress, but they are not even equal in weight to CVC syllables, as in (25a)–(25l), which do attract stress. Baker (2008) offers an articulatory gestural analysis of the difference. He observes that the apparent CVC syllable attracts stress only if the postvocalic coda consonant has an articulatory gesture distinct from that of the following onset. That is, in a CVCCV sequence, the first syllable counts as heavy only if the two intervocalic consonants have distinct articulatory gestures. When the intervocalic sequence involves a geminate or a nasal that is homorganic with a following consonant (or a glottal stop, which does not have a distinct articulatory gesture) there is only one articulatory gesture, and stress is not attracted onto CVG (or CVN where N is homorganic with a following consonant). Baker (2009) adopts a composite view of geminate representation that incorporates a gestural tier along with root nodes and moras. Under this view, stress in Ngalakgan is characterized as sensitive to the gestural tier. Ringen and Vago (2010), in contrast, take the Ngalakgan data as supporting the length representation of geminates where the stress rule treats linked structures like those in (10b) as light. Davis (2003), who maintains the underlying weight representation of geminates, suggests that the Ngalakgan stress pattern in (25) does not necessarily argue against the underlying moraic representation of geminates as in (10a); rather, the language has a high-ranked constraint that requires moraic elements to have their own place features. Thus, while geminates may be underlyingly moraic, they do not surface as moraic.[11]

To conclude this section, we have surveyed languages demonstrating three types of behavior of CVG syllables in stress systems: (i) cases where CVG and CVV pattern together; (ii) cases where CVG patterns with other CVC syllables; and (iii) cases where CVG syllables are specifically resistant to stress. We have tried to maintain the underlying moraic weight representation of geminates despite apparent evidence to the contrary. In the concluding section we will further discuss representational issues.

4 Representational issues and conclusion

In this overview, we have focused on the cross-linguistic patterning of geminate consonants while trying to maintain the representational view of geminates in (10a), in which geminates are marked as being underlyingly moraic, over the length representation in (10b). In §2, we provided evidence for the underlying moraic representation of geminate consonants by considering a variety of phonological patterning pertinent to geminates. This included the moraic analysis of initial geminates in Trukese in §2.1 and the cross-linguistic avoidance of CVVG syllables in §2.3. We made it clear in §2.2 that geminates do not always behave like a sequence of two C-slots in prosodic patterning, thereby contradicting the length representation of geminates in (10b). In §3, we surveyed geminate patterning in stress systems identifying three types of behavior. Despite the differences, we still argued for the underlying moraic view of geminate consonants.

The issue of the representation of geminate consonants has been a controversial matter and will most likely remain so in future investigations. This is because geminates do not display uniform behavior, as we have illustrated. It seems that the very nature of the data under examination determines what type of representation must be appropriate. For example, the parallel patterning between final CC sequences and final geminates in Hungarian seems to be supportive of the length representation, while the difference between final CC sequences and final

[11] It should be noted that Baker (2008) actually considers the intervocalic geminates in (25q)–(25s) and the intervocalic homorganic nasal clusters in (25m)–(25p) to be syllabified completely as onsets rather than as heterosyllabic. This differs from his earlier work, Baker (1997), where a heterosyllabic parse of geminates and homorganic nasal clusters is maintained. One shortcoming of Baker's (2008) onset analysis is that geminates and homorganic nasal clusters do not occur word-initially. Nonetheless, even if geminates could be analyzed as syllabifying as onsets, they would not add weight to a syllable and thus would be different from the initial geminates of Trukese discussed earlier.

to different aspects of the representation. For example, in Trukese, the moraic aspect of the geminate representation is crucial, while it may or may not surface with two X-slots, depending on the constraints. In Hungarian, given the parallel behavior of word-final clusters and geminates, the X-tier representation is effective, while the geminate may or may not surface as moraic, depending on the constraints. In Ngalakgan, the behavior of geminates with respect to stress can be understood through the gestural tier, as proposed by Baker (2008, 2009). Geminates and homorganic nasal + consonant clusters have a single gesture. Consequently, they can pattern as single consonants despite having two X-slots, as long as there is a high-ranked constraint that requires a moraic element not to share place features. This would provide an explanation for the repulsion of geminates to stress. It follows that while geminates may have one underlying universal representation (as in (26)), its surface realization may vary cross-linguistically, e.g. as non-moraic in Ngalakgan, but moraic in Trukese.

The composite view, or some version of it, may ultimately be the best universal representation for an underlying geminate. For example, one criticism of a purely weight account of geminates is that it cannot distinguish between a geminate that syllabifies entirely in a coda from a single coda consonant in a language in which Weight-by-Position applies. This matter comes up in the Palestinian Arabic dialect described by Abu-Salim (1980) and mentioned in Rose (2000), where a coda singleton in a word like [bit.na] 'our house' is representationally indistinguishable from a coda geminate in a word like [sitt.na] 'our grandmother' on a strictly moraic view of geminates as in (10a). Unless a length tier (or two root nodes) is assumed, there is no obvious way to distinguish the two cases. Although such examples are probably rare, the occurrence of this type of contrast indeed favors a composite analysis, especially given the language-specific evidence for Arabic moraic structure presented in various parts of this chapter. That said, it may still be possible to argue for the underlying moraic weight representation as universal, but with the understanding that the surface realization of geminates may vary across languages because of the interaction of relevant constraints.

In conclusion, there is much about the phonology of geminates that remains to be investigated. Geminates do not all pattern the same way across languages. Consequently, geminate phonology will remain an area of theoretical controversy for the foreseeable future.

ACKNOWLEDGMENTS

I thank Brett Baker, Ken de Jong, Daniel Dinnsen, Tracy Alan Hall, Abbie Hantgan, Brian José, Michael Marlo, Paul Newman, Anne Pycha, Nathan Sanders, Robert Vago and Islam Youssef for discussion on various aspects of this paper. I especially thank Natsuko Tsujimura for her detailed comments. The usual disclaimers apply.

REFERENCES

Abu-Salim, Issam M. 1980. Epenthesis and geminate consonants in Palestinian Arabic. *Studies in the Linguistic Sciences* 10(2). 1–11.

Baker, Brett. 1997. Edge crispness: Segment to mora isomorphism. *Proceedings of the West Coast Conference on Formal Linguistics* 16. 33–47.

Baker, Brett. 2008. *Word structure in Ngalakgan.* Stanford: CSLI.

Baker, Brett. 2009. Monogestural clusters as onsets: The Australian evidence. Paper presented at the 83rd Annual Meeting of the Linguistic Society of America, San Francisco.

Bamakhramah, Majdi. 2009. Syllable structure in Arabic varieties with a focus on superheavy syllables. Ph.D. dissertation, Indiana University.

Broselow, Ellen. 1992. Parametric variation in Arabic dialect phonology. In Ellen Broselow, Mushira Eid & John J. McCarthy (eds.), *Perspectives on Arabic linguistics IV*, 7–45. Amsterdam & Philadelphia: John Benjamins.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Churchyard, Henry. 1991. Compensatory lengthening and "gemination throwback" in Trukese and Puluwat as evidence for Rime in moraic phonology. Unpublished ms., University of Texas.

Churma, Donald & Yili Shi. 1996. Glottal consonants and the "sonority" hierarchy. *Proceedings of the Eastern States Conference on Linguistics (ESCOL) 1995.* 25–37.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable.* Cambridge, MA: MIT Press.

Curtis, Emily. 2003. Geminate weight: Case studies and formal models. Ph.D. dissertation, University of Washington.

Davis, Stuart. 1994. Geminate consonants in moraic phonology. *Proceedings of the West Coast Conference on Formal Linguistics* 13. 32–45.

Davis, Stuart. 1999a. On the moraic representation of underlying geminates: Evidence from prosodic morphology. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.), *The prosody–morphology interface*, 39–61. Cambridge: Cambridge University Press.

Davis, Stuart. 1999b. On the representation of initial geminates. *Phonology* 16. 93–104.

Davis, Stuart. 2003. The controversy over geminates and syllable weight. In Féry & van de Vijver (2003), 77–98.

Davis, Stuart. Forthcoming. Quantity. In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.) *The handbook of phonological theory*, 2nd edn. Malden, MA & Oxford: Wiley-Blackwell.

Davis, Stuart & Gina Torretta. 1998. An optimality-theoretic account of compensatory lengthening and geminate throwback in Trukese. *Papers from the Annual Meeting of the North East Linguistic Society* 28. 111–125.

Dyen, Isadore. 1965. *A sketch of Trukese grammar.* New Haven: American Oriental Society.

Féry, Caroline & Ruben van de Vijver (eds.) 2003. *The syllable in Optimality Theory.* Cambridge: Cambridge University Press.

Goodenough, Ward & Hiroshi Sugita. 1980. *Trukese–English dictionary.* Philadelphia: American Philosophical Society.

Gupta, Abha. 1987. Hindi word stress and the obligatory branching parameter. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 23. 134–148.

Halle, Morris & G. N. Clements. 1983. *Problem book in phonology.* Cambridge, MA: MIT Press.

Halle, Morris & Jean-Roger Vergnaud. 1980. Three-dimensional phonology. *Journal of Linguistic Research* 1. 83–105.

Ham, William. 1998. Phonetic and phonological aspects of geminate timing. Ph.D. dissertation, Cornell University.

Hart, Michele. 1991. The moraic status of initial geminates in Trukese. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 17. 107–120.

Hayes, Bruce. 1986. Inalterability in CV phonology. *Language* 62. 321–351.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Hume, Elizabeth, Jennifer Muller & Aone van Engelenhoven. 1997. Non-moraic geminates in Leti. *Phonology* 14. 371–402.

Idemaru, Kaori & Susan Guion. 2008. Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association* 38. 167–186.

José, Brian & Julie Auger. 2005. Geminates and Picard pronominal clitic allomorphy. *Catalan Journal of Linguistics* 4. 127–154.

Kawahara, Shigeto. 2007. Sonorancy and geminacy. *University of Massachusetts Occasional Papers* 32. 145–186.

Keer, Edward. 1998. Icelandic preaspiration and the moraic theory of geminates. *Proceedings of the 10th Conference of Nordic and General Linguistics, Reykjavik.* (ROA-312.)

Keer, Edward. 1999. Geminates, the OCP and the nature of Con. Ph.D. dissertation, Rutgers University.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kiparsky, Paul. 2003. Syllables and moras in Arabic. In Féry & van de Vijver (2003), 147–182.

Kiparsky, Paul. 2008a. Fenno-Swedish quantity: Contrast in Stratal OT. In Bert Vaux & Andrew Nevins (eds.) *Rules, constraints, and phonological phenomena*, 185–219. Oxford: Oxford University Press.

Kiparsky, Paul. 2008b. Weight and length. Paper presented at the CUNY Conference on the Syllable, New York City.

Kirchner, Robert. 2000. Geminate inalterability and lenition. *Language* 76. 509–545.

Kraehenmann, Astrid. 2001. Quantity and prosody asymmetries in Alemannic: Synchronic and diachronic perspectives. Ph.D. dissertation, University of Konstanz. Published 2003, Berlin & New York: Mouton de Gruyter.

Krueger, John. 1961. *Chuvash manual: Introduction, grammar, reader, and vocabulary*. Bloomington: Indiana University Press.

Leben, William R. 1980. A metrical analysis of length. *Linguistic Inquiry* 11. 497–509.

Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.

McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12. 373–418.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

Morén, Bruce. 1999. Distinctiveness, coercion and sonority: A unified theory of weight. Ph.D. dissertation, University of Maryland at College Park.

Muller, Jennifer. 2001. The phonology and phonetics of word-initial geminates. Ph.D. dissertation, Ohio State University.

Pajak, Bozena. Forthcoming. Contextual constraints on geminates: The case of Polish. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 35 (ROA-105).

Paradis, Carole. 1988. On constraints and repair strategies. *Linguistic Review* 6. 71–97.

Prince, Alan. 1990. Quantitative consequences of rhythmic organization. *Papers from the Annual Regional Meeting, Chicago Linguistic* Society 26(2). 355–398.

Pycha, Anne. 2007. Phonetic vs. phonological lengthening in affricates. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 1757–1760. Saarbrücken: Saarland University.

Pycha, Anne. 2009. Lengthened affricates as a test case for the phonetics–phonology interface. *Journal of the International Phonetic Association* 39. 1–31.

Ringen, Catherine & Robert M. Vago. 2010. Geminates: Heavy or long? In Charles Cairns & Eric Raimy (eds.) *Handbook of the syllable*. Leiden: Brill.

Rose, Sharon. 2000. Rethinking geminates, long-distance geminates, and the OCP. *Linguistic Inquiry* 31. 85–122.

Schein, Barry & Donca Steriade. 1986. On geminates. *Linguistic Inquiry* 17. 691–744.

Schmidt, Deborah. 1992. Compensatory lengthening in a segmental moraic theory of representation. *Linguistics* 30. 513–534.

Selkirk, Elisabeth. 1990. A two-root theory of length. *University of Massachusetts Occasional Papers* 14. 123–171.

Selkirk, Elisabeth. 1991. On the inalterability of geminates. In Pier Marco Bertinetto, Michael Kenstowicz & Michele Loporcaro (eds.) *Certamen Phonologicum II: Papers from the 1990 Cortona Phonology Meeting*, 187–209. Turin: Rosenberg & Sellier.

Sherer, Timothy. 1994. Prosodic phonotactics. Ph.D. dissertation, University of Massachusetts, Amherst.

Steriade, Donca. 1990. Moras and other slots. *Proceedings of the Formal Linguistics Society of Midamerica* 1. 254–280.

Thurgood, Graham. 1993. Geminates: A cross-linguistic examination. In Joel Ashmore Nevis, Gerald McMenamin & Graham Thurgood (eds.) *Papers in honor of Frederick H. Brengelman on the occasion of the twenty-fifth anniversary of the Department of Linguistics, CSU Fresno*, 129–139. Fresno: Department of Linguistics, California State University, Fresno.

Topintzi, Nina. 2006. Moraic onsets. Ph.D. dissertation, University College, London.

Topintzi, Nina. 2008. On the existence of moraic onsets. *Natural Language and Linguistic Theory* 26. 147–184.

Topintzi, Nina. 2010. *Onsets: Suprasegmental and prosodic behaviour*. Cambridge: Cambridge University Press.

Tranel, Bernard. 1991. CVC light syllables, geminates and moraic theory. *Phonology* 8. 291–302.

Tsujimura, Natsuko 2007. *An introduction to Japanese linguistics*. 2nd edn. Cambridge, MA: Blackwell.

Tyler, Stephen. 1969. *Koya: An outline grammar*. Berkeley: University of California Press.

Watson, Janet C. E. 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

38 The Representation of sC Clusters

Heather Goad

1 Introduction

Clusters of the shape *s* + consonant (sC), exemplified by *stack*, have posed a challenge for theories of syllabification, as they defy many of the constraints holding of true branching onsets, as found in, for example, *track*. Accordingly, some researchers have proposed that *s* is organized outside the onset constituent that contains the following consonant. Others have proposed that this sort of analysis holds only for a subset of sC clusters; those that rise in sonority, for example in *slack*, are represented in the same fashion as branching onsets. Yet others have argued that some sC clusters, those of the shape s + stop, form complex segments.

This chapter will critique each of these proposals. An element shared by all of them is that phonological units are highly articulated: the burden of explanation is placed precisely on the structural relationships that adjacent segments enter into. Although this approach captures many peculiarities of sC clusters, there is little attempt to explain why the consonant that displays unorthodox behavior is typically *s*. Under the view that segments are ordered to maximize their perceptibility, the behavior of *s* becomes less puzzling: strident fricatives have robust internal cues, ensuring their perceptibility even in non-optimal contexts.

If the acoustic properties of *s* are of central importance, it behoves us to ask whether the differences between sC clusters and branching onsets can be explained solely by perceptual considerations. This, of course, would challenge the view that a structural approach to cluster well-formedness is necessary. In the final section of the paper I will argue that this position is too strong. I will conclude that an adequate understanding of sC clusters requires consideration of both perceptual and structural factors.

Much of the paper compares sC clusters with branching onsets. We will observe that they differ in several respects: phonotactic constraints, word-internal syllabification, allomorph selection, patterns of reduplication, options for cluster repair, etc. Although the general observations we will detail are likely to be accepted by most phonologists, there is little agreement on how these differences should be formally represented. In this context, there are three topics that will be addressed.

The first involves critical assessment of various proposals in the literature concerning the representation of sC clusters – as a single class – in contrast to obstruent + sonorant clusters. Henceforth, I will use the term "obstruent" to refer to obstruents other than *s*. "*s*" is itself a cover term for the sibilant(s) appearing in sC clusters; although this sibilant is usually /s/, in some languages, other sibilants pattern as *s* (e.g. German *s* is usually /ʃ/; in Russian, /s z ʃ ʒ/ all pattern as *s*).

Concerning obstruent + sonorant clusters, there will be nothing particularly special to say about their representation; they form branching onsets, and there is little controversy on this matter among those who accept a hierarchically organized syllable. For sC clusters, in contrast, several options will be considered. Most of these share the idea that *s* is an appendix, a segment which is not organized by any sub-syllabic constituent; one views *s* as a coda. We will see, in addition, that some researchers propose a single representation for sC in all languages; others argue for different representations across languages.

The second topic that must be addressed is whether, in a given language, all sC clusters are represented in the same fashion. *s* + sonorant clusters are phonotactically ambiguous: like obstruent + sonorant clusters, they rise in sonority, yet like *s* + obstruent clusters, they do not respect the place constraints holding of obstruent + sonorant clusters. Depending on the weight assigned to each of these, different conclusions will be arrived at concerning the analysis of *s* + sonorant. For those researchers who place most weight on sonority profile, *s* + sonorant clusters form branching onsets. This research itself falls into two categories. One body of work aims to show that s + sonorant patterns with branching onsets while *s* + obstruent patterns differently, and is organized with some type of appendix. Another body of work considers *s* + sonorant clusters to be branching onsets, but focuses on arguing that *s* + stop clusters form complex segments.

The proposals sketched above assume that the syllable is hierarchically organized. However, there is a growing literature that de-emphasizes the role of constituency and aims to provide phonetically grounded explanations for phonological behavior. The third topic therefore considers whether differences in the behavior of obstruent + sonorant, *s* + sonorant, and *s* + obstruent can be explained by perceptual considerations alone. This topic will be the focus of the final section of the paper. Until then, a structural approach will be assumed.

2 Cluster phonotactics

We begin by detailing the phonotactic constraints most commonly held of obstruent + sonorant clusters on the place and sonority dimensions, in turn, examining sC clusters on these same dimensions (see also CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE; CHAPTER 55: ONSETS). Our focus will be on the left word edge; other types of phonological behavior will be discussed in later sections, when we examine alternative representations for sC clusters.

Consider the inventories of two-member clusters found in word-initial position in English and Dutch in (1) and (2).[1] The data are organized by the place and manner values of C_1 for obstruent + sonorant clusters and of C_2 for sC clusters. (On cluster phonotactics for English, see Fudge 1969; Selkirk 1982; Clements and

[1] We restrict discussion of obstruents in clusters to those that are voiceless; some languages display fewer options for voiced obstruents (e.g. */vl vr/ in English). We avoid consonant+ glide clusters altogether, as there are more representational options available for glides than we have space to consider.

Keyser 1983; Goldsmith 1990; Harris 1994; for Dutch, see Trommelen 1984; van der Hulst 1984; Fikkert 1994; Booij 1995; van der Torre 2003.)

(1) *English*

a. *Obstruent + sonorant*

pl	*tl		kl
pr	tr		kr
fl	*θl	*ʃl	
fr	θr	ʃr	

b. *sC clusters*

sp	st	sk
sm	sn	
	sl	
	*sr	

(2) *Dutch*

a. *Obstruent + sonorant*

	*tn	kn
pl	*tl	kl
pr	tr	kr
fl		χl
fr		χr

b. *sC clusters*

sp	st	
		sχ
sm	sn	
	sl	
	*sr	

We first consider place identity, which forbids the consonants in obstruent + sonorant clusters from having the same place (CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION). This captures the ill-formedness of */tl θl ʃl/ in English (1a), and */tn tl/ in Dutch (2a).[2] Turning to (1b) and (2b), the well-formedness of /st sn sl/ indicates that sC clusters do not respect place identity, suggesting that sC clusters do not have the same representation as obstruent + sonorant clusters. However, before we can conclude this with certainty, we must consider place-sharing */sr/, which is ill-formed in English and Dutch.[3] Importantly, */sr/ is illicit even in Dutch dialects with dorsal /r/, suggesting that the ill-formedness of this cluster has nothing to do with place identity. We return to */sr/ later in the chapter.

A second, less commonly discussed, constraint on place concerns asymmetries that hold between C_1 and C_2 in obstruent + sonorant clusters (when $C_2 \neq$ glide). English is not very revealing, because C_2 is restricted to liquids, which are coronal. Dutch is potentially more illuminating, because it contains stop + nasal clusters, and nasals contrast for place.[4] As (2a) shows, when a nasal is in C_2 in an obstruent + sonorant cluster, it must be coronal: /kn/ is well-formed; */km/ is out. The broader generalization is thus that when C_2 is a contoid, it must be coronal (except /r/; see note 2).

[2] Place identity is not respected with /r/. In English and Dutch dialects with coronal /r/, coronal + /r/ clusters are well-formed, as are dorsal + /r/ clusters in Dutch dialects with dorsal /r/. Even in languages where /t/ and /r/ are articulated near-identically, the constraint is not respected (see Arvaniti 2007 on Greek). This may suggest that /r/ permanently lacks place (Rice 1992; Goad and Rose 2004; see also CHAPTER 30: THE REPRESENTATION OF RHOTICS).

[3] Concerning Dutch */sr/, some speakers realize /sχr/ as [sr] (Waals 1999). If this represents a re-analysis of /sχr/ (van der Torre 2003), then /sr/ is well-formed for these speakers. Concerning English /ʃr/, we have placed this cluster in the obstruent + sonorant category (Goad and Rose 2004), rather than treating it as an assimilated form of /sr/ (Clements and Keyser 1983; Goldsmith 1990).

[4] I say "potentially," because, as is undoubtedly evident, obstruent + sonorant clusters will be analyzed as branching onsets below. There is, however, dispute about the status of /kn/ in Dutch, as branching onset (Fikkert 1994; Booij 1995) or appendix-initial (van der Hulst 1984; Trommelen 1984; Kager and Zonneveld 1986). Notably, intervocalic /kn/ is syllabified as coda + onset, *contra* the branching onset analysis.

C_2 in sC clusters has a different profile: it can have any place of articulation. Indeed, in s + obstruent and s + nasal clusters, C_2 displays the same range of place contrasts attested for singleton onsets (e.g. /sp st sk/ alongside /p t k/ in English). Directly comparing nasal-final clusters in Dutch, we find the following: */km/, /kn/; /sm/, /sn/. The absence of */km/ alongside the presence of /sm/ is unexpected, if these clusters are represented identically.

Because of the disputed status of Dutch /kn/ (note 4), we turn to Modern Greek to better examine differences in place profile between C_2 in obstruent + sonorant *vs.* sC clusters. (On Modern Greek cluster phonotactics, see Joseph and Philippaki-Warburton 1987; Drachman 1990; Klepousniotou 1998; Morelli 1999; Tzakosta and Vis 2009; on Attic Greek, see Steriade 1982.) The data in (3) reveal that, although C_2 in an obstruent + sonorant cluster can have any manner, leading to a wider range of cluster profiles than in Dutch or English, C_2 must still be coronal (Klepousniotou 1998).[5] (3b) shows that sC clusters are more restricted on the manner dimension than they are in Dutch, but among clusters with obstruents in C_2, it can nevertheless be seen that C_2 can have any place.

(3) *Modern Greek*

a. *Obstruent + sonorant*

pt		kt
(pn)	*tn	(kn)
pl	*tl	kl
pr	tr	kr
ft	*θt	xt
fθ		(xθ)
*fn	(θn)	xn
fl	(θl)	xl
fr	θr	xr
(mn)		

b. sC *clusters*

sp	st	sk
sf	(sθ)	sx
(sm)	(sn)	
	*sl	
	*sr	

In sum, we have observed that C_2 in an obstruent + sonorant cluster does not parallel C_2 in an sC cluster on the place dimension. On the contrary, a closer parallel is observed between C_1 in an obstruent + sonorant cluster and C_2 in an sC cluster, which we return to below.

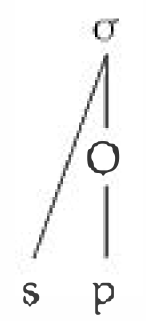
We consider finally the sonority constraints that hold between C_1 and C_2 in initial clusters.[6] Greek is not very revealing here, as both obstruent + sonorant and sC clusters can have a falling, flat, or rising sonority profile (although (3b) indicates that the productivity of the latter for sC clusters is questionable; we return to this below). We therefore focus on English and Dutch. (1) and (2) show that sC clusters need not rise in sonority, in contrast to obstruent + sonorant clusters. Although C_1 in an sC cluster is an obstruent and so the potential exists for these

[5] Stop + stop and fricative + stop are often considered to be archaic in spoken Modern Greek (Joseph and Philippaki-Warburton 1987; Morelli 1999). They do occur in higher registers, which is why they are included here. /ps ts ks/ are absent from (3a); I assume they are complex segments (following Tzakosta and Vis 2009). Clusters in parentheses (as well as /tm/) are not productive, although the number of /sm/-initial roots is somewhat larger than for the others.

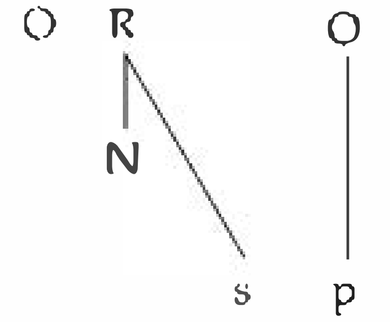
Thanks to Jenny Dalalakis and Katerina Klepousniotou for help with the Greek data.

[6] I assume the following sonority scale, which is roughly based on relative intensity: stop < fricative < nasal < liquid < vocoid. See also CHAPTER 49: SONORITY.

c. *Licensed by σ* (e.g. van der Hulst 1984)



d. *Coda* (e.g. Kaye 1992)



In (5a), *s* is extraprosodic. I am using this term in its narrowest sense, to refer only to the situation where an element is licensed but not organized into higher structure; compare (5b) and (5c). These two share with (5a) the idea that *s* is an appendix: *s* is not organized by any sub-syllabic constituent, in contrast to (5d), where *s* is a coda (technically in Kaye's model a rhymal dependent).

Two predictions follow from the difference in representation and headedness in (4) *vs.* (5). First, there should be languages that permit dependents in branching onsets but not sC clusters, and vice versa. This prediction holds true. Spanish is a language with branching onsets, but lacking initial sC clusters (Harris 1983). Acoma, spoken in New Mexico, has the opposite profile: initial clusters are restricted to sC (Miller 1965). Relatedly, Fikkert's (1994) study on the acquisition of Dutch reveals that some children acquire branching onsets first, parallel to Spanish, while others acquire sC clusters first, parallel to Acoma. Second, languages permitting both types of structures should not prevent them from being combined. To my knowledge, this prediction always holds. In languages that have branching onsets and sC clusters, three-member clusters of the shape *s* + branching onset are also well-formed. However, these clusters may be restricted to a subset of what would be expected from a free combination of sC clusters and branching onsets in the particular language (e.g. Greek /sx/, /xr/, */sxr/; English /sk/, /kl/, */skl/ (loans aside)).

Further, an explanation emerges under (5) for why the constraints against place identity and for rising sonority do not hold of sC clusters. It is not enough for two consonants to be adjacent; they must be sisters, as in (4).

Finally, we demonstrate that the difference in headedness between branching onsets and sC clusters can account for a commonly attested pattern of cluster reduction in acquisition, illustrated in (6) from two learners of German and English respectively, Annalena (Elsen 1991) and Amahl (Smith 1973).[7]

(6) a. *Annalena* (age 1;4–1;9)

	output	*target*	
obs + son	[daɯbɛ]	[tr]*aube*	'grape'
	[fɪkə]	[fl]*iege*	'fly'
s + obs	[pɪgəl]	[ʃp]*iegel*	'mirror'
	[daɪnə]	[ʃt]*ein*	'stone'
s + son	[mɪsən]	[ʃm]*eißen*	'to throw'
	[lɑːfə]	[ʃl]*afen*	'to sleep'

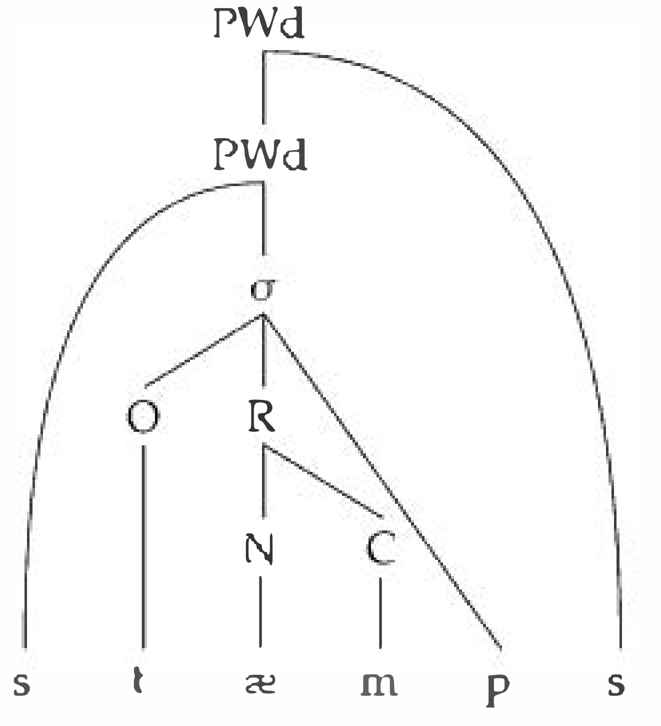
[7] [b̥ d̥ g̊] indicate voiceless unaspirated lenis stops in Amahl's data (Smith 1973).

The alternative that is typically adopted, therefore, is that these two types of extraprosodic segments are extraprosodic at different points in the derivation. Extraprosodic /s/ in *taxed* and /p/ in *tramps* are incorporated as codas in the post-lexical phonology before inflectional /t/ and /s/ are added (cf. Borowsky 1986). Initial /s/ in (7a) similarly loses extraprosodic status in the post-lexical phonology, where lexical constraints on sonority profile are no longer assumed to hold. The problem with this approach, however, is that it fails to show that extraprosodic elements ever function as members of the constituents that ultimately come to organize them, the onset in (7a) and the coda in (7b) (Piggott 1991).

This problem is resolved once the Strict Layer Hypothesis is abandoned – the requirement that elements be dominated by the immediately higher category in the prosodic hierarchy (Nespor and Vogel 1986). Extraprosodic segments on this view are technically not *extra*prosodic (unaffiliated); instead, they are organized by some higher constituent in the prosodic hierarchy.

One possible representation for /<s>tæm<p><s>/, consistent with this approach, is in (8): initial /s/ is organized by the PWd, extrarhymal /p/ is organized by the syllable, and inflectional /s/ is adjoined to the PWd. This representation is consistent with the Peripherality Condition: extrarhymal /p/ and inflectional /s/ are each at the right edge of a separate PWd (Goad and White 2006).

(8)



Returning specifically to sC clusters, the representation for /st/ in (8) involves s linking directly to the PWd (see e.g. Goldsmith 1990, drawing on evidence from English; Trommelen 1984 and Fikkert 1994 from Dutch; Goad and Rose 2004 from German). An alternative involves s linking directly to the syllable, the inverse of extrarhymal /p/ in (8) (see e.g. van der Hulst 1984 with evidence from Dutch; Levin 1985 and Kenstowicz 1994 from English; Barlow 1997 and Gierut 1999 from English in phonologically delayed children; Drachman 1990 and Tzakosta and Vis 2009 from Greek; Tzakosta 2009 from child Greek).[10] These two proposals were provided earlier as (5b) and (5c).[11]

[10] Note that Levin (1985) adjoins, rather than directly links, *s* to the syllable. To the list in the text we can add Giegerich (1992), Hall (1992), and Booij (1995), who analyze s as an onset-internal appendix, using data from English, German, and Dutch, respectively; and Ewen and Botma (2009), who organize *s* into the specifier position of the onset for Germanic.

[11] Other less commonly proposed licensers for *s* will not be considered due to space constraints: e.g. the Foot (Green 2003 on Munster Irish) and the Phonological Phrase (Vaux 1998 on Armenian).

The alternatives in (5b) and (5c) make different predictions concerning the distribution of sC clusters. I show that both options (or their equivalents) are needed, indicating that sC clusters cannot be represented identically across languages (Goad and Rose 2004; Vaux 2004; Ewen and Botma 2009). In languages with (5b), sC is only licensed PWd-initially. German has this profile (Goad and Rose 2004). (9) shows that sC clusters only occur stem-initially; word-internal tautosyllabic sC clusters are actually stem-initial, (9b). If stem-initial corresponds to PWd-initial, this restriction on sC distribution can be captured through (5b).

(9) *German*

a.	[ʃpɪnə]$_{PWd}$	'spider'
	[ʃteːən]$_{PWd}$	'to stand'
b.	[bə[ʃteːən]$_{PWd}$]$_{PWd}$	'to insist'
	[gə[ʃteːən]$_{PWd}$]$_{PWd}$	'to confess'
	*[CV.ʃCV]$_{PWd}$	

Hall (1992) argues against (5b) for German on grounds that it incorrectly predicts aspiration in s + stop clusters, as the stop is syllable-initial in this representation. However, Iverson and Salmons (1995), following Kim (1970), offer an explanation for the absence of aspiration in *s* + stop that holds independently of how *s* is organized: because *s* is voiceless, the peak of glottal width that characterizes aspiration is internal to *s*, not the following stop. Thus, we do not see the absence of aspiration in s + stop as reason to reject (5b).

In contrast to German, sC clusters in Dutch and English have a wider distribution, requiring (5c). Both languages contain monomorphemic examples where the rhyme preceding sC appears unable to accommodate *s* (e.g. Dutch [ɛkstər] 'magpie', English [ɛkstrə] *extra* (van der Hulst 1984; Levin 1985)). (5c), however, freely permits violations of the Peripherality Condition. Accordingly, before we definitively conclude that it is required for morpheme-internal sC, we must examine the following alternative: PWd-initial sC clusters involve appendices organized as in (5b); in word-medial clusters, *s* is a coda. If this analysis could be supported, we could dispense with (5c).

To show that (5c) is truly needed, we examine word-medial sC clusters in English in detail. Harris (1994) discusses the constraints governing three-position rhymes shaped VVC in this language. As (10a) reveals, coda sonorants in these superheavy syllables are confined to coronals which share place with the following onset (*[ʃowlbər], *[mawmpən]). PWd-internal VCC rhymes are not considered by Harris (they are not well-formed in Government Phonology, the framework in which he works). (10b) reveals that the onset is similarly constrained to coronal and the preceding consonants must be homorganic nasal + stop (*[vɪltnər], *[ʤʌlkʃən]).

(10) *English*

a.	VVC *rhymes*		b.	VCC *rhymes*	
	[ʃowldər]	*shoulder*		[æntlər]	*antler*
	[mawntən]	*mountain*		[vɪntnər]	*vintner*
	[kawnsəl]	*council*		[ʤʌŋ(k)ʃən]	*junction*

With these constraints in mind, we turn to cases where the consonant following VV/VC is s. Parallel to (10a), (11a) shows that only coronal s is permitted after VV (*[iːftər]) and the following consonant must be coronal (*[iːspər]).[12] This leads Harris to conclude that *s* in (11a) is syllabified as the coda of a three-position rhyme. The forms in (11b) are similarly parallel to those in (10b), seemingly leading to the same conclusion.

(11) a.	VVs *rhymes*		b. VCs *rhymes*	
	[iːstər]	*Easter*	[mɑnstər]	*monster*
	[ɔjstər]	*oyster*	[mɪnstrəl]	*minstrel*

Problems arise, however, in (12). (12a) shows that VCs does not always respect the constraints holding of (11b): the consonant preceding *s* is not restricted to place-sharing sonorants. (12b) reveals that the onset following *s* can be other than coronal, in contrast to (11a) and (11b). One could object on grounds that, *extra* aside, the words in (12) involve Latinate prefixes. However, these prefixes are not synchronically productive: they fall within the stress domain, and must therefore be contained inside the lower PWd (['ɑbstəkəl]$_{PWd}$, *['ɑb[stəkəl]$_{PWd}$]$_{PWd}$).

(12) *Appendix* s

a. *Non-coronal codas*
['ɛkstrə] *extra*
['ɑbstəkəl] *obstacle*

b. *Non-coronal onsets*
[ˌɛkspə'zɪʃən] *exposition*
['kɑnskrɪpt] *conscript*

If the forms in (12) truly involve appendixal s, we expect to find s occurring after three-position rhymes of the shape in (10). (13) shows that such words are well-formed, albeit rare. *Bolster* and *holster* are monomorphemic, and while *-ster* in *upholster* is historically a class 2 suffix (seventeenth-century *uphold-ster* 'small furniture dealer'), this analysis no longer holds, as revealed by the fact that *upholster* is now a verb and *-y* can attach outside (*upholstery*).[13]

(13) VVC *rhymes followed by* s

[bowlstər] *bolster*
[howlstər] *holster*
[ʌp(h)owlstər] *upholster*

In short, while some instances of medial sC in English may involve *s* as coda, appendixal s is required to capture the data in (12) and (13). This thereby supports the postulation of (5c), where *s* is licensed by the syllable.

Although languages like English appear to require (5c), this analysis cannot straightforwardly capture the fact that these same languages syllabify *s* as a coda

[12] Neither constraint holds in dialects with lengthened [ɑː]/[æː] in e.g. *after*, *basket* (Harris 1994).

[13] Hayes (2009: 210–211) considers the vowel in such words to be monophthongized (so presumably short) in some dialects. However, this does not hold of all dialects; witness, for example, RP [bəwlstə].

after short stressed vowels. If the appendix representation is available word-medially, why do native speakers judge sC as heterosyllabic ('*pes.ter*) rather than as appendix + onset ('*pe.ster*)? It cannot be due to sonority profile, as will be seen shortly for Dutch. Heterosyllabicity of medial sC is handled more elegantly in Kaye's (1992) Government Phonology approach to sC clusters, which we consider now.

Kaye (1992) proposes that sC clusters are syllabified as coda + onset sequences, shown earlier in (5d).[14] He provides support from Italian, Ancient Greek, European Portuguese, and British English; see also Brockhaus (1999) on German and Cyran and Gussmann (1999) on Polish. We consider first Italian, where the coda + onset pattern observed for '*pester*-type words is illustrated more concretely. In Italian, rhymes of stressed syllables must branch (Chierchia 1986). When a stressed syllable lacks a coda, the vowel lengthens; see (14a) and (14b). (14c) shows that sC clusters do not trigger lengthening as do branching onsets; instead, they pattern with coda + onset sequences (14d), revealing that medial *s* is a coda.

(14) *Medial* sC *in Italian*

a. ['faːto] 'fate'
b. ['kaːpra] 'goat'
c. ['pasta] 'pasta'
d. ['parko] 'park'

Turning to word-initial position, Kaye proposes that s in this position is also a coda; the difference between the initial and medial environments is that, in the former, *s* is the coda of an empty-headed syllable. Word-initial coda *s* follows from the Uniformity Principle in Government Phonology, which requires syllabification to be constant for a given string of segments, within and across languages. Kaye provides empirical support for Uniformity from Italian masculine definite article allomorphy and *raddoppiamento sintattico*. In the former, vowel- and sC-initial words pattern together, in contrast to words beginning with (branching) onsets; compare (15a) and (15b) with (15c). Since the representation for s in an sC cluster in (5d) includes a preceding nucleus, there is a structural parallel with vowel-initial words.

(15) *Masculine definite article allomorphy* (Davis 1990)

a. *l'est* /lo est/ 'the east'
b. *lo studente* 'the student'
c. *il burro* 'the butter'
il clima 'the climate'

In *raddoppiamento sintattico*, the first consonant in an onset geminates when the preceding word ends in a stressed vowel, while the first consonant in an sC cluster resists gemination; see (16). The pattern in (16b) follows directly from the view that s is a coda: as coda + onset, sC already has precisely the structure that holds of geminates.

[14] A precursor to this appears in Vennemann (1988). Vennemann proposes that initial *s* is "quasi-nuclear" in some languages, a type of degenerate syllable. He argues that this analysis of Latin *s* explains its development into a regular Vs syllable in some Romance languages, notably Spanish.

(16) *Raddoppiamento sintattico* (Chierchia 1986)

a.	*palto pulito*	[pal'toppu'lito]	'clean coat'
	citta triste	[ʧit'tat'triste]	'sad city'
b.	*citta straniera*	[ʧit'tastra'niera]	'foreign city'
		*[ʧit'tasstra'niera]	

Although it is evident from (15) and (16) that sC clusters in Italian cannot be analyzed in the same fashion as branching onsets, the proposal that they contain an initial appendix can also handle these data (see Chierchia 1986; Davis 1990). Of the cases Kaye considers, the construction that poses a particular challenge for appendixal *s* is European Portuguese vowel nasalization. We turn to this case now.

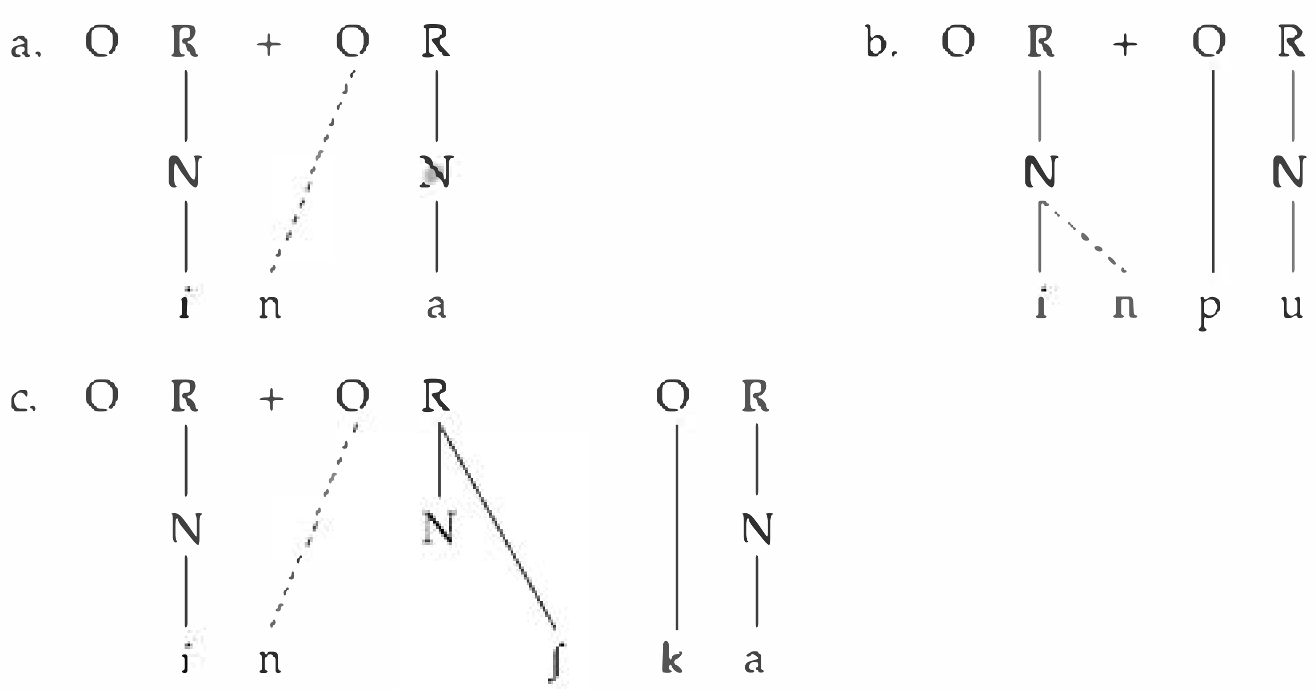
In European Portuguese, nasal consonants cannot close syllables. While /n/ is realized intact before vowel-initial bases (17a), before onset-initial bases, nasality surfaces on the preceding vowel (17b), (17c). Interestingly, sC-initial bases pattern as vowel-initial (17d).

(17) *European Portuguese*

a.	[in]*admissivel*	'inadmissable'
b.	[ĩ]*pureza*	'impurity'
	[ĩ]*satisfeito*	'dissatisfied'
c.	[ĩ]*tratavel*	'unsociable'
d.	[inʃk]*apavel*	'inescapable'

(17d) can be straightforwardly expressed under Kaye's view that sC clusters are coda + onset, because, for independent reasons, all syllables in Government Phonology contain an onset constituent. Consider the representations below.[15] In (18a), /n/ associates to the onset of the first syllable in the base. In (18b), this position is occupied, so nasality is preserved on the preceding vowel. The right result obtains in (18c) precisely because the syllable containing *s* includes an empty onset.

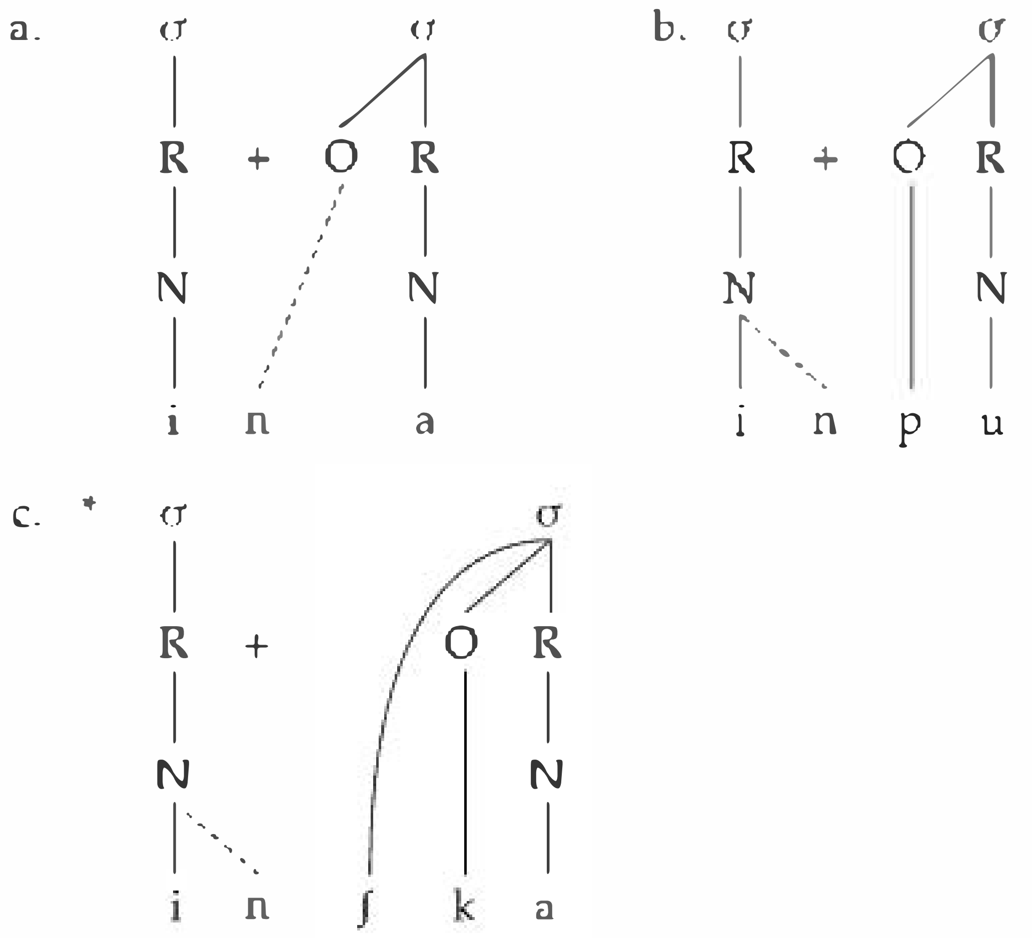
(18) *Coda analysis*



15 No representations are provided by Kaye. (18) reflects my best guess (minus X-slots), based on his discussion.

An appendix analysis of *s*, it seems, cannot formally capture (17d). See (19c), where bases with initial sC are incorrectly predicted to pattern with onset-initial bases because there is no empty constituent to host /n/.

(19) *Appendix analysis*



Kaye's paper compares the coda + onset analysis of sC clusters to the alternative that they form branching onsets; the option that *s* is analyzed as an appendix is not discussed. We have seen that appendixal *s* cannot straightforwardly capture European Portuguese. It also goes against the Uniformity Principle: Italian s is an appendix in *stra'niera* but a coda in *'pasta*; this would likely be considered a weakness by proponents of Government Phonology. However, there are contexts where word-internal s maintains its appendix status, in contrast to the pattern observed for Italian; sC can follow a rhyme that is already full, so *s* cannot be accommodated as an ordinary coda. We have already observed this for English, but we have also seen that English permits word-internal three-position rhymes under limited circumstances and most of the problematic sC data are in words that historically involve prefixes. To ensure that there is nothing unusual about English concerning the distribution of sC clusters, let us turn to Acoma (Miller 1965).

Two-position rhymes in Acoma are limited to VV and seemingly Vs (loans aside); see (20a) and (20b). If word-internal *s* were always a regular coda, as it appears to be in (20b), we would expect it to be restricted to occurring after short vowels, as in Italian. (20c) reveals that this is not the case.

(20) *Acoma*

a.	[spúuná]	'pottery'
	[jaˀái]	'sand'
b.	[sust'á]	'I took water'
	[ˀéská]	'rawhide'

Table 38.1 Jarmo's cluster development

Stage	*Pattern*		*Examples*		*Age*
1	stop	/'vlɪndər/	['kɪnʉ]	'butterflly'	2;2.6
		/'slaːpə(n)/	['taːpə]	'to sleep'	2;0.28
2	liquid	/'vlɪndər/	['lɪŋʉ]	'butterfly'	2;2.27
		/'slaːpə(n)/	['laːpə]	'to sleep'	2;3.9
3	fricative	/'χlɛiˌbaːn/	['χɛiχaːn]	'slide'	2;4.1
		/'slaːpə(n)/	['saːpə]	'to sleep'	2;3.9
4	stop + liquid	(skipped by Jarmo)			
5	fricative/s + liquid	/'flɛsjə/	['slɛsjə]	'bottle (DIM)'	2;4.1
		/slak/	[flak]~[slak]	'snail'	2;4.1

study: all follow the same developmental path for fricative + liquid and s + lateral clusters. The stages through which learners pass are exemplified in Table 38.1 with data from Jarmo.

Further, all children in Fikkert's study master *s* + stop clusters at a different point in time from rising sonority clusters. As all rising sonority clusters pattern together in Dutch acquisition, in contrast to s + obstruent, Fikkert concludes that *s* + sonorant clusters are represented in the same manner as branching onsets. For her, *s* + obstruent clusters involve *s* licensed by the PWd, as in (5b) above.

While sonority plays a decisive role in Fikkert's data, this is not the case for all children. Indeed, we observed in (6) that when only one member of a cluster is produced by Annalena and Amahl, it is the cluster head that survives, regardless of its relative sonority. To show that this pattern extends past the deletion stage, consider the developmental path for Amahl in Table 38.2.

Table 38.2 Amahl's cluster development

Stage	*obstruent + liquid*	*/sl/*	*/sm sn/*	*/sp sk/*	*/st/*	*Age*
1–8	reduction to head	reduction to head	reduction to head	reduction to head	reduction to head	2.60–2.175
13–14	branching onset acquired					2.233–2.256
15–19		fusion	fusion	vacuous fusion	vacuous fusion	2.261–2.333
20–22					fusion	2.345–3.38
24		appendix acquired				3.78–3.96
25			appendix acquired			3.104–3.128
26–29				appendix acquired	appendix acquired	3.133–3.355

approach to cluster behavior. The epenthesis site, in particular, is chosen to maximize perceptual similarity between the target (non-epenthesized) form and the output. In view of this, we consider in this section whether the differences that hold between sC clusters and true branching onsets can be explained by perceptual considerations alone; this, of course, would challenge the claim that a structural approach to cluster behavior is necessary.

We do not have space to examine Fleischhacker's proposal in detail, but the predictions she motivates are as follows: (i) anaptyxis is preferred to prothesis in stop + sonorant; (ii) prothesis is preferred to anaptyxis in s + stop; (iii) among s + sonorant, more anaptyxis is expected as C_2 increases in sonority; and (iv) more anaptyxis is expected in stop + sonorant than in fricative + sonorant. Concerning (iv), note that Fleischhacker's account does not distinguish s from other fricatives; that is, no explanation is provided for the observation that fricatives other than *s* pattern with stops in preferring anaptyxis to prothesis ((24) above). We return to this shortly. First, let us examine the role of perception in sC well-formedness in more detail (CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY).

As alluded to earlier, the acoustic properties of *s*, unlike other obstruents, enable it to appear in positions where it is not followed by a sonorant: strident fricatives have robust internal cues for both place and manner, ensuring their perceptibility in all contexts, even before stops (Wright 1996, 2004). Clearly, then, the view that segments are ordered to yield a rise in sonority toward the peak does not extend to s. Indeed, in spite of the sonority reversal, (strident) fricative + stop is superior to both stop + stop and stop + fricative,[20] even though the latter two contain a sonority plateau and minimal rise respectively, because (strident) fricatives are less dependent on formant transitions for their identification than stops (Wright 1996, 2004).[21]

However, while the acoustic properties of s explain why appendices are so often limited to *s* on the one hand and why these segments can be followed by stops on the other, they cannot, as far as I can tell, explain cross-linguistic preferences on sC profile. Table 38.3 shows that sC clusters have a rather unusual distribution across languages when viewed from the perspective of perceptual robustness. We focus on word-initial position. Since the perceptibility of all consonants in C_2 position in an initial sC cluster will be partly compromised by the preceding s,

Table 38.3 sC cluster profiles across languages

	Spanish	*French, Acoma*	*Greek*	*English*	*Dutch*	*German*	*Russian*
s + stop	*	✓	✓	✓	✓	✓	✓
s + fricative	*	*	✓	*	✓	*	✓
s + nasal	*	*	(*)	✓	✓	✓	✓
s + lateral	*	*	*	✓	✓	✓	✓
s + rhotic	*	*	*	*	(*)	✓	✓

20 Evidence that this observation is not restricted to *strident* fricative + stop comes from Greek: the stop + stop clusters in (3a) are often replaced by fricative + stop.

21 See Morelli (1999) for an alternative explanation of obstruent cluster well-formedness that appeals to markedness constraints on segment sequencing.

we would expect consonants that are most perceptible to be positioned after s. Masking should not be too severe in this context; as mentioned earlier, Byrd (1994) observes that #sk clusters involve less overlap than s#k and sk#. The problem may rather be one of duration: Byrd finds that /s/ is longer in #sk than in both s#k and sk#, while /k/ is shorter in #sk than in both s#k and sk#. If the relatively short duration of C_2 can be generalized to other #sC clusters, we would expect segments with robust internal cues to be favored in this position. Liquids should be optimal, since they have clear formant structure throughout. Nasals should be favored over stops, since their manner (and to a lesser extent their place) properties are present in the nasal spectrum. Stops, which have weak internal cues, should be the least optimal.

What we observe in Table 38.3, by contrast, is that s + stop is favored. French and Acoma do not permit s + sonorant clusters at all (French has s + sonorant in loanwords), and depending on the status of marginal s + nasal clusters, Greek may fall into this category as well. Otherwise, it permits s + sonorant clusters of lower sonority than those of higher sonority.

Although a larger typology of languages is required before firm conclusions can be drawn, Table 38.3 suggests that s + stop > s + nasal > s + lateral > s + rhotic (> = is more harmonic than). The favored profile in sC clusters is thus the opposite of that observed for branching onsets. This is not unexpected on a structural account if all sC clusters are head-final, in contrast to branching onsets. In sC clusters, C_2 is the onset head; thus it should respect the preferences holding of singleton onsets. Since obstruents are the optimal onsets (e.g. Clements 1990), a parallel should be observed between obstruents in C_1 position in branching onsets and stops in C_2 position in sC clusters (not fricatives more generally, because of the preceding s; see Wright 2004: 51).

While the C_1C_2 asymmetry in branching onsets *vs.* sC clusters follows from the status of s as an appendix, it is best captured, I suggest, under Kaye's proposal that s is a coda. Recall from (14) and (21) that medial sC clusters in Italian and Dutch are heterosyllabic. If sC clusters are *always* syllabified as coda + onset clusters, then their profile should respect cross-linguistic preferences for optimal syllable contact. Syllable contact will favor C_2 with lower sonority: Vs.TV > Vs.NV> Vs.lV > Vs.rV. As C_2 increases in sonority, the cluster prefers to be syllabified as a branching onset, but if this option is never available for sC clusters, then higher-sonority sC clusters will be forbidden, regardless of their position in the word.

The profile in Table 38.3 closely parallels Fleischhacker's typology in (25) for preferred epenthesis sites, in sC clusters. Prothesis occurs more commonly when C_2 has lower sonority. As the sonority of C_2 increases, prothesis will result in poor syllable contact. Note as well that the proposed syllable contact account of sC well-formedness leads to a distinction between s + sonorant and fricative + sonorant, as only the latter can form branching onsets. Thus, the fact that fricative + sonorant patterns with stop + sonorant in epenthesis follows, in contrast to under Fleischhacker's account (see (iv) above).

In sum, I contend that both perceptual and structural considerations must be factored into our understanding of cluster well-formedness. While perceptual considerations can explain why appendices are so often limited to s and why s + stop is well-formed in spite of its sonority profile, it is the structural differences between sC clusters and branching onsets that explain the preference for sC profile on the sonority dimension as well as some observed differences in epenthesis site.

7 Conclusion

In this chapter, we have examined several alternative analyses for sC clusters. On the empirical front, we have seen that s + stop clusters reliably pattern differently from true branching onsets. Not surprisingly, then, the proposals we have examined for s + stop all share an important property: s + stop clusters are head-final, whether *s* is unaffiliated, an appendix organized by some prosodic constituent above the onset, the coda, or the first member of a complex segment. Branching onsets, by contrast, are head-initial. This difference in headedness helps to explain parallels on the place dimension between C_2 in an s + stop cluster and C_1 in a branching onset, as well as syllabification preferences in word-medial contexts.

Beyond that, however, details of the proposals differ, and, following on this, each proposal is both supported and challenged by the available evidence. There are languages like Acoma, English, and Dutch where *s* can appear medially after rhymes that are "full," thereby providing support for the analysis that s is linked to the syllable and potentially challenging the coda analysis. However, in some of these same languages, namely English and Dutch, as well as in languages like Italian, the observation that sC clusters are heterosyllabic after stressed vowels supports the coda analysis and thereby questions the proposal that s is licensed by the syllable or that s + stop form a complex segment. The complex segment analysis, in turn, is supported by the reduplication pattern in Gothic which both the appendix and coda analyses fail to elegantly capture.

At present, then, it seems that multiple representations for sC clusters may be required. If the number of parametric options is limited and there is robust evidence available for learners to determine the appropriate representation for the language being acquired, this is far from problematic. For example, the fact that sC clusters have a more limited distribution in some languages (German) than in others (English, Dutch) can be captured if licensing by the PWd represents the least marked option and therefore the starting point for learners. There will then be positive evidence available to signal learners of some languages that sC clusters are licensed lower down, by the syllable. However, we have also seen that this type of scenario may not always work. If Dutch learners initially assume that s + sonorant clusters form branching onsets, the evidence available to undo this analysis in favor of one where all sC clusters pattern as a class is far from robust and may present a learnability challenge.

The problem with s + sonorant clusters more generally is that they are phonotactically ambiguous and, following from this, they pattern ambiguously across languages. While one could fail to be surprised by this, on the grounds that these clusters are both s-initial and rise in sonority, exactly how their ambiguous behavior should be formally expressed is far from clear. They appear to be analyzed as branching onsets in some languages (e.g. Jarmo's Dutch grammar) and as appendix/coda-initial in others (e.g. Amahl's English grammar), but in this particular case, this finding is surprising, in view of the otherwise high degree of similarity between the two target languages.

The solution that languages employ different analyses for s + sonorant is far from optimal and may lead some researchers to abandon a structural approach to the syllable altogether in favor of a perceptually grounded account of segmental contact. Indeed, the latter may find support in the observation that even within the class of *s* + sonorant, languages show different patterns of behavior; we have seen that the division between prothesis and anaptyxis can be drawn anywhere internal to this class. At the same time, however, a purely perceptually based account seems to be challenged by the finding that preferences for sC cluster profile are virtually the inverse of those observed for obstruent + sonorant clusters. While an appeal to syllable contact was made to capture both of these observations, the analysis follows most straightforwardly from the proposal that sC clusters are always syllabified as coda + onset strings. We have already seen that this proposal may be challenged by languages such as English, Dutch, and Acoma.

In spite of the quantity of research that has been undertaken on sC clusters, it is perhaps most evident that more needs to be done before the issue of their representation can be resolved (if ever). A sampling of questions at the two extremes includes the following. At one end of the spectrum, can a more detailed examination of perceptual factors capture differences in the behavior of fricative + sonorant and *s* + sonorant clusters, thereby further questioning the need for a structurally based approach to segmental contact and syllabification behavior? At the other end, if a structural account of behavior based on syllable contact proves fruitful to pursue, with judicious use of abstract representations, can the coda + onset analysis be motivated for all languages? I leave these and many other questions in between to future research.

ACKNOWLEDGMENTS

Many thanks to two anonymous reviewers and the editors for helpful comments. This work was supported by grants from SSHRC and FQRSC.

REFERENCES

Arvaniti, Amalia. 2007. Greek phonetics: The state of the art. *Journal of Greek Linguistics* 8. 97–208.

Barlow, Jessica. 1997. A constraint-based account of syllable onsets: Evidence from developing systems. Ph.D. dissertation, Indiana University.

Booij, Geert. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.

Borowsky, Toni. 1986. Topics in the lexical phonology of English. Ph.D. dissertation, University of Massachusetts, Amherst.

Brockhaus, Wiebke. 1999. The syllable in German: Exploring an alternative. In Harry van der Hulst & Nancy Ritter (eds.) *The syllable: Views and facts*, 169–218. Berlin: Mouton de Gruyter.

Broselow, Ellen. 1983. Nonobvious transfer: On predicting epenthesis errors. In Susan Gass & Larry Selinker (eds.) *Language transfer in language learning*, 269–280. Rowley, MA: Newbury House.

Byrd, Dani. 1994. Articulatory timing in English consonant sequences. Ph.D. dissertation, University of California, Los Angeles.

Cairns, Charles & Mark Feinstein. 1982. Markedness and the theory of syllable structure. *Linguistic Inquiry* 13. 193–225.

Chierchia, Gennaro. 1986. Length, syllabification and the phonological cycle in Italian. *Journal of Italian Linguistics* 8. 5–33.

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*, 283–333. Cambridge: Cambridge University Press.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Cyran, Eugeniusz & Edmund Gussmann. 1999. Consonantal clusters and governing relations: Polish initial consonant sequences. In Harry van der Hulst & Nancy Ritter (eds.) *The syllable: Views and facts*, 219–247. Berlin & New York: Mouton de Gruyter.

Davis, Stuart. 1990. Italian onset structure and the distribution of *il* and *lo*. *Linguistics* 28. 43–55.

Drachman, Gaberell. 1990. A remark on Greek clusters. In Joan Mascaró & Marina Nespor (eds.) *Grammar in progress: GLOW essays for Henk van Riemsdijk*. Dordrecht: Foris.

Elsen, Hilke. 1991. *Erstspracherwerb: Der Erwerb des deutschen Lautsystems*. Wiesbaden: Deutscher Universitäts-Verlag.

Ewen, Colin J. 1982. The internal structure of complex segments. In van der Hulst & Smith (1982: Part II), 27–67.

Ewen, Colin J. & Bert Botma. 2009. Against rhymal adjuncts: The syllabic affiliation of English postvocalic consonants. In Kuniya Nasukawa & Phillip Backley (eds.) *Strength relations in phonology*, 221–250. Berlin & New York: Mouton de Gruyter.

Fikkert, Paula. 1994. On the acquisition of prosodic structure. Ph.D. dissertation, University of Leiden.

Fleischhacker, Heidi. 2001. Cluster-dependent epenthesis asymmetries. *UCLA Working Papers in Linguistics 7: Papers in Phonology* 5. 71–116.

Fudge, Erik C. 1969. Syllables. *Journal of Linguistics* 5. 253–286.

Fujimura, Osamu & Julie B. Lovins. 1978. Syllables as concatenative phonetic units. In Alan Bell & Joan B. Hooper (eds.) *Syllables and segments*, 107–120. Amsterdam: North-Holland.

Giegerich, Heinz J. 1992. *English phonology: An introduction*. Cambridge: Cambridge University Press.

Gierut, Judith. 1999. Syllable onsets: Clusters and adjuncts in acquisition. *Journal of Speech, Language, and Hearing Research* 42. 708–726.

Goad, Heather & Yvan Rose. 2004. Input elaboration, head faithfulness and evidence for representation in the acquisition of left-edge clusters in West Germanic. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in phonological acquisition*, 109–157. Cambridge: Cambridge University Press.

Goad, Heather & Lydia White. 2006. Ultimate attainment in interlanguage grammars: A prosodic approach. *Second Language Research* 22. 243–268.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Oxford & Cambridge, MA: Blackwell.

Green, Antony Dubach. 2003. Extrasyllabic consonants and onset well-formedness. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*, 238–253. Cambridge: Cambridge University Press.

Hall, T. A. 1992. *Syllable structure and syllable-related processes in German*. Tübingen: Niemeyer.

Harris, James W. 1983. *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: MIT Press.

Harris, John. 1994. *English sound structure*. Oxford: Blackwell.

Hayes, Bruce. 1980. A metrical theory of stress rules. Ph.D. dissertation, MIT.

Hayes, Bruce. 2009. *Introductory phonology*. Malden, MA & Oxford: Wiley-Blackwell.

Hulst, Harry van der. 1984. *Syllable structure and stress in Dutch*. Dordrecht: Foris.

Hulst, Harry van der & Norval Smith (eds.) 1982. *The structure of phonological representations*. 2 parts. Dordrecht: Foris.

Iverson, Gregory K. & Joseph C. Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12. 369–396.

Joseph, Brian D. & Irene Philippaki-Warburton. 1987. *Modern Greek*. London: Croom Helm.

Kager, René & Wim Zonneveld. 1986. Schwa, syllables, and extrametricality in Dutch. *The Linguistic Review* 5. 197–221.

Karimi, Simin. 1987. Farsi speakers and the initial consonant cluster in English. In Georgette Ioup & Steven Weinberger (eds.) *Interlanguage phonology: The acquisition of a second language sound system*, 305–318. Cambridge, MA: Newbury House.

Kaye, Jonathan. 1990. "Coda" licensing. *Phonology* 7. 301–330.

Kaye, Jonathan. 1992. Do you believe in magic? The story of s+C sequences. *SOAS Working Papers in Linguistics and Phonetics* 2: 293–313. Reprinted 1996 in Henryk Kardela & Bogdan Szymanek (eds.) *A Festschrift for Edmund Gussmann*, 155–176. Lublin: University Press of the Catholic University of Lublin.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology* 7. 193–231.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kim, Chin-Wu. 1970. A theory of aspiration. *Phonetica* 21. 107–116.

Klepousniotou, Katerina. 1998. Onset clusters in Modern Greek. Unpublished ms., McGill University.

Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.

Lovins, Julie. 1974. Why loan phonology is natural phonology. *Papers from the Annual Regional Meeting, Chicago Linguistic Society: Parasession on natural phonology*, 240–250.

Miller, Wick R. 1965. *Acoma grammar and texts*. Berkeley & Los Angeles: University of California Press.

Morelli, Frida. 1999. The phonotactics and phonology of obstruent clusters in Optimality Theory. Ph.D. dissertation, University of Maryland at College Park.

Myers, Scott. 1987. Vowel shortening in English. *Natural Language and Linguistic Theory* 5. 485–518.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Piggott, Glyne L. 1991. Apocope and the licensing of empty-headed syllables. *The Linguistic Review* 8. 287–318.

Rice, Keren. 1992. On deriving sonority: A structural account of sonority relationships. *Phonology* 9. 61–99.

Selkirk, Elisabeth. 1982. The syllable. In van der Hulst & Smith (1982: Part II), 337–383.

Sigurd, Bengt. 1965. *Phonotactic structures in Swedish*. Lund: Uniskol.

Smith, Neil V. 1973. *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT.

Steriade, Donca. 1999. Alternatives to syllable-based accounts of consonantal phonotactics. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.) *Item order in language and speech*, 205–245. Prague: Karolinum Press.

Torre, Erik Jan van der. 2003. Dutch sonorants: The role of place of articulation in phonotactics. Ph.D. dissertation, University of Leiden.

Trommelen, Mieke. 1984. *The syllable in Dutch: With special reference to diminutive formation*. Dordrecht: Foris.

Tzakosta, Marina. 2009. Asymmetries in /s/ cluster production and their implications for language learning and language teaching. In Anastasios Tsangalidis (ed.) *Selected Papers from the 18th International Symposium of Theoretical and Applied Linguistics*, 365–373. Thessaloniki: Monochromia.

Tzakosta, Marina & Jeroen Vis. 2009. Phonological representations of consonant sequences: The case of affricates vs. "true" clusters. In Giorgos K. Giannakis, Mary Baltazani, Giorgos I. Xydopoulos & Anastasios Tsangalidis (eds.) *Proceedings of the 8th International*

Conference on Greek Linguistics, 558–573. Ioannina: Department of Linguistics, University of Ioannina.

Vaux, Bert. 1998. *The phonology of Armenian*. Oxford: Clarendon Press.

Vaux, Bert. 2004. The appendix. Paper presented at the Symposium on Phonological Theory: Representations and Architecture, City University of New York.

Vennemann, Theo. 1982. Zur Silbenstruktur der deutschen Standardsprache. In Theo Vennemann (ed.) *Silben, Segmente, Akzente*, 261–305. Tübingen: Niemeyer.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin: Mouton de Gruyter.

Vogt, Hans. 1942. The structure of the Norwegian monosyllable. *Norsk Tidsskrift for Sprogvidenskap* 12. 5–29.

Waals, Juliette. 1999. An experimental view of the Dutch syllable. Ph.D. dissertation, Utrecht University.

Weijer, Jeroen van de. 1996. *Segmental structure and complex segments*. Tübingen: Niemeyer.

Wiese, Richard. 1988. *Silbische und lexikalische Phonologie: Studien zum Chinesischen und Deutschen*. Tübingen: Niemeyer.

Wiese, Richard. 1996. *The phonology of German*. Oxford: Clarendon Press.

Wright, Richard. 1996. Consonant clusters and cue preservation in Tsou. Ph.D. dissertation, University of California, Los Angeles.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, 34–57. Cambridge: Cambridge University Press.

39 Stress: Phonotactic and Phonetic Evidence

MATTHEW GORDON

1 Introduction

Stress can be signaled through a number of different acoustic properties, including increased duration, greater intensity, and higher fundamental frequency. Stress may also affect segmental and syllable structure. Typically, stressed syllables trigger qualitative fortition and/or lengthening, whereas unstressed syllables are associated with lenition and/or shortening. To take an example of a stress-driven fortition process affecting syllable structure, Dutch (Booij 1995) inserts an intervocalic glottal stop as an onset to stressed vowels; epenthesis does not interrupt vowel sequences in which the second vowel is unstressed. We thus have pairs such as ['xa.ɔs] 'chaos' and [a.'ʔɔr.ta] 'aorta', in which the presence of glottal stop is predictable from stress. American English provides well-described cases of lenition in unstressed syllables. For example, post-vocalic coronal stops weaken to taps before unstressed syllabic sounds, e.g. /sɪti/ → ['sɪɾi] 'city'. Furthermore, most unstressed vowels reduce to schwa, e.g. ['kanˌtɛkst] *context vs.* [kən'tɛkstʃuəl] *contextual*, or may delete in certain contexts delete, e.g. ['tmeɪɾoʊ] ~ ['təmeɪɾoʊ] *tomato*, ['ksændɹə] ~ [kə'sændɹə] *Cassandra*.

While most segmental effects of metrical structure can be transparently linked to stress, there are others that are not predictable from stress, despite displaying properties typically associated with stress-induced alternations. For example, Nganasan, a Uralic language (Tereshchenko 1979; Helimski 1998; Vaysman 2009), has an alternation between strong and weak intervocalic consonants, termed "consonant gradation," whereby strong consonants, generally voiceless or prenasalized obstruents, alternate with weak consonants, typically voiced or not prenasalized. The appearance of strong and weak consonants is predictable from syllable count (1). In the onset of even-numbered non-initial syllables, the strong grade appears, while the weak grade appears in the onset of odd-numbered non-initial syllables. Long vowels interrupt the alternating syllable count and, as long as they are not word-initial, are always preceded by weak consonants.

(1) *Nganasan consonant gradation* (Vaysman 2009: 43)

ˌjamaˈða-tu	'his/her/its animal'
ˌŋoruˈmu-tu	'his/her/its copper'
suːˈðəː-ðu	'his/her/its lung'
ŋuˈhu-ðu	'his/her/its mitten'

As Vaysman shows, this pattern is explained if one assumes that words are parsed into binary feet starting at the left edge of words with long vowels forming monosyllabic feet and degenerate feet allowed word finally. Strong consonants occur foot medially and weak consonants occur in foot-initial syllables that are not also word-initial (2).

(2) *Nganasan consonant gradation as a reflex of foot structure* (Vaysman 2009: 43)

(ˌjama)(ˈða-tu)	'his/her/its animal'
(ˌŋoru)(ˈmu-tu)	'his/her/its copper'
(suː)(ˈðəː)-(ðu)	'his/her/its lung'
(ŋuˈhu)-(ðu)	'his/her/its mitten'

The interesting feature of the Nganasan data is that stress does not always fall on syllables predicted to be stressed by the metrical structure diagnosed by consonant gradation. Primary stress in Nganasan falls on the penultimate syllable in all the words in (2), with a secondary stress occurring on initial syllables that are not adjacent to the primary stress. The monosyllabic foot in the last two words is thus completely unstressed, as is the first foot in the penultimate word.

This chapter provides a typological overview of the phonetic correlates of stress and the various types of effects of stress and metrical structure on segment-level features, exploring how these effects can offer insight into the nature of stress and metrical structure and their formal representation (see CHAPTER 40: THE FOOT, CHAPTER 41: THE REPRESENTATION OF WORD STRESS and CHAPTER 57: QUANTITY-SENSITIVITY for related issues). The structure of the chapter is as follows. §2 examines supra-segmental correlates of stress including the phonetic parameters of duration, fundamental frequency, and intensity. §3 examines segmental alternations conditioned directly by the presence or absence of stress. §4 focuses on the role of foot structure in predicting fortition and lenition of vowels and consonants. §5 addresses segmental changes triggered by foot structures that conflict with metrical constituency as diagnosed by the stress system. §6 explores the role of history in shaping these mismatches between stress and the foot structure relevant for segmental alternations. Finally, §7 summarizes the chapter.

2 Suprasegmental phonetic correlates of stress

Fry (1955, 1958) pioneered research on the acoustic correlates of stress in his examination of the effect of stress in English on duration, intensity, and fundamental frequency. Focusing on the vowels in noun–verb minimal pairs such as *ˈconvert* (noun) *vs. conˈvert* (verb) and *ˈimport* (noun) *vs. imˈport* (verb), Fry found that stressed vowels were associated with greater duration, greater intensity, and higher fundamental frequency than their unstressed counterparts, with the last of these properties being most reliable as a cue to stress.

Since Fry's work, phoneticians have considerably broadened the typological database on stress correlates by examining other potential correlates of stress and by targeting a diverse set of languages for phonetic study. This research program has yielded many important results. For example, beyond the acoustic domain, stress is also associated with hyperarticulation of segments, which has ramifications for the segmental alternations discussed in §3. Furthermore, other potential acoustic correlates of stress have come to light, such as measurements of stress that are sensitive to spectral tilt (Sluijter and van Heuven 1996a, 1996b) or that integrate intensity over time (Lieberman 1960; M. Beckman 1986). Finally, typological study has shown that many languages are similar to English in using duration, intensity, and/or fundamental frequency to signal stress, e.g. Polish (Jassem *et al.* 1968), Tagalog (Gonzalez 1970), Mari (Baitschura 1976), Indonesian (Adisasmito-Smith and Cohn 1996), Pirahã (Everett 1998), Aleut (Taff *et al.* 2001), Chickasaw (Gordon 2004), Turkish (Levi 2005), and Kabardian (Gordon and Applebaum 2010). It has also become increasingly clear that the phonetic study of stress is a complicated matter for several reasons.

Languages differ in their relative reliance on different cues to stress where the relevance of certain properties is functionally constrained in many languages by the extent to which potential stress correlates are used to mark phonemic contrasts other than stress. For example, lexical tone languages – e.g. Thai (Potisuk *et al.* 1996) and Pirahã (Everett 1998) – are less reliant on fundamental frequency to cue stress, and languages with phonemic length contrasts, e.g. Finnish, may have phonetically longer unstressed vowels than stressed vowels.

There are also languages in which potential phonetic markers of stress do not converge on a single syllable but rather are shared between multiple, often, though not always, adjacent syllables. For example, in Welsh (Williams 1985) an unstressed final syllable often has higher fundamental frequency and longer vowel duration than an unstressed penultimate syllable in the same word. In such cases, lengthening of the consonant immediately following the stressed vowel seems to be the most reliable correlate of stress. A similar situation arises in Estonian, where the primary stressed initial syllable, if it contains a phonemic short vowel, will be shorter than the immediately following syllable and often have less intensity and lower fundamental frequency (Lehiste 1965; Eek 1975; Gordon 1995). Lengthening of the consonant in the onset of the stressed syllable serves as the most reliable cue to stress in Estonian (Lehiste 1966; Gordon 1997). Hyman (1989) discusses cases in Bantu of different diagnostics leading to different conclusions about the location of stress. For example, certain Eastern and Southern Bantu languages display evidence for metrical prominence on the penultimate syllable, such as vowel lengthening, attraction of high tone, and even phonetic stress. However, these properties may conflict with other properties that suggest stress on another syllable, e.g. high tone on the antepenult in Zulu, even though the penult conditions vowel lengthening. A similar pattern of high tone on the antepenult preceding a stressed penult is found in the Northern Iroquoian language Onondaga (Chafe 1970, 1977; Michelson 1988), the Polynesian language Tongan (Schütz 1985), and several Micronesian languages (Rehg 1993). In "split-cue" stress systems such as those described in this paragraph, determining the location of stress is potentially problematic.

3 The taxonomy of segmental correlates of metrical structure

Fortition and lenition effects associated with metrical structure may be broadly classified into three groups according to the property triggering these segmental alternations. The first type of segmental effect is well documented and involves stress (or lack thereof) directly as a trigger of fortition and/or lenition. A second type of segmental effect is predictable from constituent structure rather than stress, but the constituent structure motivating the segmental change accords with the metrical parse evinced by the stress system. A third type of segmental alternation, exemplified by Nganasan, is linked to metrical constituency, where the foot structure diagnosed by the segmental change is at odds with that suggested by the stress system. In the following sections we take a closer look at examples of each of these types of relationships between segmental properties and metrical structure.

3.1 Stress-driven segmental phenomena: Fortition and lenition

Many languages display segmental changes that are conditioned by stress or lack of stress. The typical pattern is for sounds to strengthen in stressed contexts and to weaken in unstressed positions. Fortition and lenition can target either consonants or vowels. In the case of consonants, unstressed position is usually associated with decreased resistance to coarticulatory effects and hypo-articulation (de Jong 1995), resulting in reduced constriction either temporally or in magnitude. Kirchner (2001), Lavoie (2001), Bye and de Lacy (2008), and Vaysman (2009) summarize a number of segmental alternations conditioned by stress, of which I mention a few here (see CHAPTER 66: LENITION for an overview of the typology of lenition). Post-vocalic coronal stops in American English reduce to flaps before an unstressed syllabic sound, and stops become aspirated in the onset of stressed syllables. In Kupia (Christmas and Christmas 1975), the stops /p ʈ/ have lenited variants in the onset of unstressed syllables: /p/ is realized as a fricative and /ʈ/ as a tap. West Tarangan (Nivens 1992) displays fortition in the onset of stressed syllables: /j/ affricates to /ʤ/, and /w/ occlusivizes to /g/, a change that also applies to word-initial consonants. In the development from Proto-Samurian to pre-Lezgian (Topuria 1974; Giginejshvili 1977; Yu 2004), voiced stops devoiced, a type of fortition, and geminated in the onset of stressed syllables.

Stress often also triggers lengthening of consonants. Thus, in Urubu Kaapor (Kakumasu 1986) and optionally in Tukang Besi (Donohue 1999) oral stops lengthen in the onset of primary stressed syllables. Lengthening is also employed as a strategy to beef up the rime of stressed syllables. Hayes (1995: 83) discusses several cases of lengthening in order to enhance the weight of stressed syllables. For example, in Munsee (Goddard 1979), a consonant geminates after metrically prominent short vowels, thereby converting the stressed syllable from light (CV) to heavy (CVC).

Vowels are also subject to fortition and lenition processes conditioned by stress. As in the case of consonantal alternations, vowels may be affected either qualitatively or quantitatively by the presence or absence of stress. Cross-linguistically, it is very common for vowels to lengthen in stressed syllables. Hayes (1995: 83) initial syllable in non-derived words containing only reduced vowels underlyingly (3b). (We abstract away from cases where an underlying schwa alternates with a full vowel on the surface.)

(3) *Eastern Mari stress* (Vaysman 2009: 62–64)

a.	koŋ'ga	'oven'
	ʃer'ge	'comb'
	køgør'ʧen	'dove'
	'teŋgəz	'sea'
	'olək	'meadow'
	'joŋaləʃ	'mistake'
	pu'ʃaŋgə	'tree'
b.	'βəɲər	'canvas'
	'əʃkəl	'step'
	'ləβə	'butterfly'

Rounding harmony is propagated rightward from the stressed vowel in a word. Thus, the 3rd person possessive suffix surfaces as [ʃe] when the stressed vowel is unrounded (4a) but as [ʃø] or [ʃo] when the stressed vowel is a rounded vowel (4b). (The backness of the rounded vowel is conditioned by a process of front–back harmony.)

(4) *Eastern Mari rounding harmony* (Vaysman 2009: 92)

a.	'ergəʃe	'his/her/its boy'
	y'reməʃe	'his/her/its street'
	pykʃer'ɯne-ʃe	'his/her/its walnut tree'
b.	'ʃyrə-ʃø	'his/her/its soup'
	kəʎ'mo-ʃø	'his/her/its shovel'
	'ʃoʃə-ʃo	'his/her/its spring'

Another type of harmony that is sensitive to stress involves the propagation of a feature from an unstressed syllable up through a stressed syllable, which blocks further spreading of the harmonizing feature. Tudanca Spanish (Penny 1978) instantiates this type of harmony. In Tudanca, underlying final high vowels, which surface as more centralized than their non-final counterparts, induce centralization (in the front/back and/or height dimension depending on the vowel) of preceding vowels up to and including the stressed vowel (5). (Centralized vowels are marked by the ˟ diacritic.) Stress is lexically governed and may fall on either the penultimate or antepenultimate syllable.

(5) *Tudanca laxness harmony* (Penny 1978: 54–55)

'pintu̽	'male calf'	cf. 'pinta	'female calf'
'se̽ku̽	'dry (MASC)'	cf. 'seka	'dry (FEM)'
'po̽rti̽ku̽	'portico'		
'pu̽lpi̽tu̽	'pulpit'		
anti'gwi̽si̽mu̽	'very old'		
o're̽ga̽nu̽	'oregano'		

Walker (2004, 2005) discusses several cases of metaphony in Romance languages involving harmonizing in height of a stressed vowel to a posttonic one (see also CHAPTER 110: METAPHONY IN ROMANCE for an overview of similar processes in Italian dialects). In some language varieties, as in Tudanca Spanish, harmony propagates left-wards from the triggering vowel to the stressed vowel through any intervening unstressed vowels. In other varieties, e.g. Asturian Lena Bable (Hualde 1989, 1998), the stressed vowel is transparent to the harmonizing feature, which propagates past the stressed vowel leftward to the pre-tonic vowel.

3.3 *Exceptional lenition in prominent syllables*

Despite the cross-linguistic tendency for stressed syllables to be associated with increased segmental strength, this pattern is not universal. Mokša Mordvin (Vaysman 2009) optionally lenites consonants in the onset of stressed syllables that are not word initial. Stress in Mokša is sensitive to a distinction between the low-sonority vowels [i u ə] and the high-sonority vowels [a æ o e]. In words containing vowels belonging to the same sonority class, stress falls on the first syllable of a word (6a). In words consisting vowels of different sonority classes, stress falls on the leftmost vowel belonging to the higher-sonority group (6b).

(6) *Mokša Mordvin stress* (Vaysman 2009: 135–137)

a.	'bʲanəʧ	'boat'
	'məkur	'buttocks'
	'kuʃin	'jug'
	'juʒə	'skin'
	'aka	'older sister, aunt'
	'lopæ	'leaf'
	'pango	'mushroom'
	'sʲeja	'goat'
b.	tsʲə'ræ	'son'
	vi'na	'alcohol'
	az'na	'older sister's husband'
	sʲi̥rək-'ka	'elm (PROLATIVE)'

Lenition in stressed onsets entails voicing of underlying voiceless obstruents, liquids, and glides, the spirantization of underlying voiced stops, the conversion of /m/ to /w/ and the deletion of /n/, with concomitant nasalization of the stressed vowel. Crucially, lenition does not target word-initial consonants, as the examples in (7) indicate.

(7) *Mokša Mordvin lenition in stressed onsets* (Vaysman 2009: 142–143)

kur'ka	~ kur'ga	'turkey'
sʲər'pe	~ 'sʲər'be	'heart'
bʲanʲə'ʧ-oze	~ bʲanʲə'ʤ-oze	'my boat'
pəjə'l̥ʲ-oze	~ pəjə'lʲoze	'my knife'
bu'jæ-ze	~ bu'jæ-ze	'my end'
tʲə̥'bʲe-ze	~ tʲe'βʲe-ze	'my work'
pin'gæ-ze	~ pin'ɣæ-ze	'my period of time'
pʲi'ma	~ pʲi'wa	'large cup, mug'
kə'nak-oze	~ kə'ãk-oze	'my guest'

This constraint interaction can be illustrated by considering the analysis of Guaraní vowel harmony developed in J. Beckman (1998). In Guaraní (Gregores and Suárez 1967), nasalized and oral vowels contrast in stressed syllables, but in unstressed syllables nasalized vowels may only surface before a nasal consonant. The [nasal] feature also spreads leftward from a prenasalized stop and from a phonemic nasalized vowel up to but not including a stressed vowel (8). Nasalization additionally spreads rightward (as the examples below indicate) although its phonetic properties are different, which has led certain researchers, e.g. Flemming (1994), to analyze it as phonetic rather than phonological. Beckman thus does not develop an analysis of rightward spreading of nasality.

(8) *Guaraní nasal harmony* (Gregores and Suárez 1967: 69)

/amaa'porõre'ju/	→	ʔãᵐbaʔa'porõrẽju	'if I work you come'
/je' intena/	→	je'intẽnã	'just once more!'
/ija,kāra'ku/	→	ʔĩnã,kãrã'ku	'is hot-headed'
/rojotopa'pamarõro'xova,rã/	→	rojotopa'pamãr̃õr̃õ'xoṽã,r̃ã	
		'if now we meet all of us, we will have to go'	

Two faithfulness constraints play a pivotal role in Beckman's analysis. First, a generic faithfulness constraint, IDENT(nasal), requires that segments underlyingly associated with a [nasal] feature preserve that feature on the surface. The second constraint is the positionally defined analog to IDENT(nasal), IDENT-σ́(nasal), which requires that surface segments in stressed syllables preserve their underlying [nasal] specification. The existence of contrastive nasality on stressed vowels but not on unstressed vowels follows from the ranking of a markedness constraint banning nasalized vowels, $*V_{nasal}$, above generic IDENT(nasal) but below position-specific IDENT-σ́(nasal). This ranking ensures that any underlyingly nasalized vowel will lose its nasality if it surfaces in an unstressed position. Critical to the analysis of nasal harmony in Beckman's analysis is an alignment constraint, ALIGN-L(nasal), requiring that all instances of the feature [nasal] be aligned with the left edge of a word. This constraint is honored in forms in which nasality either is underlyingly associated with a segment in the first syllable or has propagated to the first syllable through nasal spreading. One violation is incurred for each segment intervening between a nasal feature and the left edge of the word. By sandwiching ALIGN-L(nasal) above IDENT (nasal) but below IDENT-σ́(nasal), nasality is correctly predicted to spread as far leftward as the stressed syllable, where it is blocked from spreading any further (9).

(9)

/je'intenã/ 'just once more'	IDENT-σ́(nas)	ALIGN-L(nas)	IDENT(nas)
☞ a. je'ɨntẽnã		***	*
b. nẽ'ɨ̃ntẽna	*!		***
c. je'ɨntenã		***, *!*****	

Lenition in unstressed syllables can be handled similarly in the positional faithfulness approach. A constraint banning non-lenited segments is ranked above a generic faithfulness constraint, but below a positional faithfulness constraint

(1990) as a foot, in recited Japanese verse. Finally, foot-final lengthening may be viewed as the foot-level manifestation of the well-attested phenomenon of final lengthening observed at levels above the foot, such as the word and phrase (Wightman *et al.* 1992). Foot-final lengthening has also been appealed to by Revithiadou (2004) as a factor in promoting the cross-linguistically pervasive phenomenon of iambic lengthening discussed in the next section.

4.3 Iambic/trochaic length asymmetries

Certain asymmetries in quantitative alternations are also best explained with reference to foot structure rather than directly to stress. In particular, stressed syllables in iambic and trochaic feet appear to display fundamentally different length characteristics (see Hayes 1985, 1995 and CHAPTER 44: THE IAMBIC–TROCHAIC LAW). Stressed syllables in languages employing an iambic parse are often lengthened cross-linguistically, whereas those with trochaic feet characteristically either fail to lengthen the stressed syllable, or, in some cases, even shorten the stressed syllable. Chickasaw (Munro and Ulrich 1984; Munro and Willmond 1994, 2005; Munro 1996, 2005; Gordon *et al.* 2000; Gordon 2003, 2004; Gordon and Munro 2007) is an iambic language in which closed syllables and syllables containing long vowels are heavy, i.e. are parsed as monosyllabic feet word-initially or following a stressed syllable. The final syllable is also parsed as a foot even if it is light (CV). The rightmost stress, i.e. the one on the final syllable, is the primary one, except that a long (or lengthened) vowel in pre-final position attracts the primary stress from a final CV(C). As shown in (11), stressed vowels in open non-final syllables substantially lengthen, where the output of lengthening is a vowel that is either nearly neutralized or completely neutralized in length with a phonemic long vowel depending on the vowel and the speaker (see Gordon *et al.* 2000 for phonetic duration results).

(11) *Iambic lengthening in Chickasaw* (lengthened vowels indicated by ˑ)

(ʧiˌpiˑ)(saˈliˑ)(ˌtok)	'I looked at you.'
	cf. (piˈsaˑ)(liˌtok) 'He looked at you.'
(ʧiˌʧoˑ)(ˌkoʃ)(koˈmoˑ)(ˌʧi)	'He makes you play.'
	cf. (ʧoˌkoʃ)(koˈmoˑ)(ˌtok) 'He played.'
(ʧiˌkiˑ)(siˈliˑ)(ˌtok)	'He bit you.'
	cf. (kiˈsiˑ)(liˌtok) 'He bit it.'
(aˌsaˑ)(biˈkaˑ)(ˌtok)	'I was sick.'
	cf. (aˈbiˑ)(kaˌtok) 'He was sick.'

The process of iambic lengthening has an intuitive purpose, in that it enhances the prominence of the stressed syllable. In some languages, the beefing up of the stressed syllable in an iambic foot is achieved by lengthening a consonant rather than a vowel (see Hayes 1995: 82–83 for a survey of iambic lengthening; but see Bye and de Lacy 2008 for re-analyses of iambic consonant lengthening). This consonant can either be the coda consonant in a stressed CVC syllable, as in the Chevak dialect of Central Alaskan Yupik (Woodbury 1981) or the onset of the following syllable, the first half of which ends up closing the stressed syllable, as in Delaware (Goddard 1979, 1982).

(13) *Nganasan consonant alternation and foot structure* (Vaysman 2009: 43, 52)

(ˌjama)('ða-tu)	'his/her/its animal'
(ˌŋoru)('mu-tu)	'his/her/its copper'
(sɯː)('ðəː)-(ðu)	'his/her/its lung'
(ŋu'hu)-(ðu)	'his/her/its mitten'
(baːr)('pə-ⁿtə)(nu)	'master, chief (LOC SG NON-POSS)'
(hɨa)('ʧə-ⁿtə)(nu)	'thumb (LOC SG NON-POSS)'
(ku'bu)(tənu)	'skin, fur (LOC SG NON-POSS)'
('hʷaː)(tənu)	'tree (LOC SG NON-POSS)'

As these examples show, the strong variant of the 3rd person possessive suffix (beginning with [t]) and the locative singular non-possessive suffix (beginning with a prenasalized [ⁿt]) surface foot medially, whereas the weak allophone (beginning with [ð] and plain [t] in the two suffixes, respectively) occurs foot initially. The strong grade and the weak grade both occur in the onset of unstressed syllables, meaning that stress does not predict the alternation. Furthermore, the foot structure diagnosed by the consonant alternations does not accord with the foot structure that would be required to predict primary stress.

It is not the case, however, that stress assignment in Nganasan is completely blind to foot structure. Secondary stress falls on odd-numbered syllables counting from the left edge of a word in keeping with the footing predicted by consonant gradation (14a). Two provisions to this generalization, however, make the relationship even between secondary stress and foot structure opaque. First, secondary stress may not clash with an immediately following stress (14b) and, second, secondary stress skips over a light (CV) syllable in favor of a heavy (CVV) syllable (14c). In both situations, a final syllable is potentially footed but unstressed, and a word may display a mix of iambic and trochaic feet.

(14) *Nganasan secondary stress* (Vaysman 2009: 24)

a.	(ˌbaku)(ˌnumə)('numə)	'my salmon (PROLATIVE)'
	(ˌtiri)(ˌmɨmə)('numə)	'my caviar (PROLATIVE)'
b.	(ˌkaðar)(mə'nu)(mə)	'my light (PROLATIVE)'
	(ˌʧemɲi)(mə'nu)(mə)	'my salary (PROLATIVE)'
c.	(ky)(ˌmaː)(mə'nu)(mə)	'my knife (PROLATIVE)'
	(le)(ˌhua)(mə'nu)(mə)	'my board (PROLATIVE)'

A further context in which foot structure is relevant to stress arises in words in which both the penult and the antepenult contain a central vowel. We abstract away from cases in which the penult and antepenult contain different central vowels, a situation that gives rise to variability in the location of stress, and consider here only cases in which both the penultimate and antepenultimate syllables contain the same central vowel. In words of this profile, the penult attracts stress if it is foot-initial (15a) but the antepenult carries the stress if the penult is foot-final (15b).

(15) *Nganasan stress in words with the same central vowel in the penult and antepenult* (Vaysman 2009: 36)

a.	(ˌbiðɨ)('tɨrə)	'you (SG) are drinking (INTR)'
	(ˌɲilɨ)('tɨrə)	'you (SG) are living'

7 Summary

Stress or lack of stress is associated with both suprasegmental and segmental properties. On a suprasegmental level, stress typically triggers lengthening, higher fundamental frequency, and greater intensity, although there are many languages in which these properties do not converge on a single syllable but rather are distributed over multiple syllables. On a segmental level, stress characteristically, although not always, triggers consonant fortition or blocks lenition targeting unstressed syllables. Some metrically conditioned segmental alternations, on the other hand, are better explained with reference to foot structure. For example, boundary-driven processes such as foot-initial fortition and foot-final lengthening are plausibly the foot-level analogs of well-documented phenomena applying at the word level. Furthermore, stressed vowel lengthening in many languages is explicable in terms of stress, but its cross-linguistic bias toward applying in iambic stress systems suggests that it is sensitive to foot structure. A final type of segmental alternation cannot be accounted for with reference to stress but rather suggests the relevance of foot structure that is orthogonal to the stress system in certain languages. Examination of historical data potentially provides insight into these mismatches between foot structure diagnosed by fortition and lenition and foot structure diagnosed by stress by showing that the segmental changes became entrenched at a chronologically earlier stage, when stress and foot structure coincided.

ACKNOWLEDGMENTS

The author wishes to thank two anonymous reviewers and the editors for their many helpful comments and suggestions on an earlier draft of this chapter.

REFERENCES

Adisasmito-Smith, Niken & Abigail C. Cohn. 1996. Phonetic correlates of primary and secondary stress in Indonesian: A preliminary study. *Working Papers of the Cornell Phonetics Laboratory* 11. 1–16.
Baitschura, Uzbek. 1976. Instrumental phonetische Beiträge zur Untersuchung der Sprachmelodie und des Wortakzentes im Tscheremissischen. *Études Finno-Ougriennes* 3. 109–122.
Beckman, Jill N. 1998. Positional faithfulness. Ph.D. dissertation, University of Massachusetts, Amherst.
Beckman, Mary E. 1986. *Stress and non-stress accent*. Dordrecht: Foris.
Bereczki, Gábor. 1988. Geschichte der wolgafinnischen Sprachen. In Sinor (1988), 314–350.
Booij, Geert. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.
Buckley, Eugene. 2000. On the naturalness of unnatural rules. *Proceedings from the 2nd Workshop on American Indigenous Languages*, 1–14. Santa Barbara: Department of Linguistics, University of California, Santa Barbara.
Bye, Patrik & Paul de Lacy. 2008. Metrical influences on lenition and fortition. In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.) *Lenition and fortition*, 173–206. Berlin & New York: Mouton de Gruyter.

Byrd, Dani. 1994. Articulatory timing in English consonant sequences. Ph.D. dissertation, University of California, Los Angeles.

Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73. 493–533.

Chafe, Wallace. 1970. *A semantically based sketch of Onondaga*. Baltimore: Waverly Press.

Chafe, Wallace. 1977. Accent and related phenomena in the Five Nations Iroqouis languages. In Larry M. Hyman (ed.) *Studies in stress and accent*, 169–181. Los Angeles: Department of Linguistics, University of Southern California.

Cho, Taehong & Patricia Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29. 155–190.

Christmas, Raymond B. & J. Elisabeth Christmas. 1975. *Kupia phonemic summary*. Kathmandu: Tribhuvan University.

Colarusso, John. 1992. *A grammar of the Kabardian language*. Calgary: University of Calgary Press.

Crosswhite, Katherine. 1998. Segmental *vs.* prosodic correspondence in Chamorro. *Phonology* 15. 281–316.

Crosswhite, Katherine. 2001. *Vowel reduction in Optimality Theory*. New York: Routledge.

Crosswhite, Katherine. 2004. Vowel reduction. In Hayes *et al.* (2004), 191–231.

de Jong, Kenneth J. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97. 491–504.

de Lacy, Paul. 2001. Markedness in prominent positions. *MIT Working Papers in Linguistics* 40. 53–66. (ROA-282.)

Delgutte, Bertrand 1982. Some correlates of phonetic distinctions at the level of the auditory nerve. In Rolf Carlson & Björn Granström (eds.) *The representation of speech in the peripheral auditory system*, 131–150. Amsterdam: Elsevier Biomedical.

Delgutte, Bertrand. 1997. Auditory neural processing of speech. In W. J. Hardcastle & John Laver (eds.) *The handbook of phonetic sciences*, 507–538. Oxford & Cambridge, MA: Blackwell.

Delgutte, Bertrand & Nelson Y. S. Kiang. 1984. Speech coding in the auditory nerve. I: Vowel-like sounds. *Journal of the Acoustical Society of America* 75. 879–886.

Dilley, Laura C., Stefanie Shattuck-Hufnagel & Mari Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24. 423–444.

Donohue, Mark. 1999. *A grammar of Tukang Besi*. Berlin & New York: Mouton de Gruyter.

Eek, Arvo. 1975. Observations on the duration of some word structures II. *Estonian Papers in Phonetics* 1975. 7–51.

Everett, Keren. 1998. Acoustic correlates of stress in Pirahã. *Journal of Amazonian Languages* 1. 104–162.

Flemming, Edward. 1994. The role of metrical structure in segmental rules. *Papers from the Annual Meeting of the North East Linguistic Society* 24. 97–110.

Fry, D. B. 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27. 765–768.

Fry, D. B. 1958. Experiments in the perception of stress. *Language and Speech* 1. 126–152.

Giginejshvili, B. K. 1977. *Sravnitel'naja fonetika dagestanskix iazykov*. Tbilisi: Tbilisi University.

Goddard, Ives. 1979. *Delaware verbal morphology: A descriptive and comparative study*. New York: Garland.

Goddard, Ives. 1982. The historical phonology of Munsee. *International Journal of American Linguistics* 48. 16–48.

Gonzales, A. 1970. Acoustic correlates of accent, rhythm, and intonation in Tagalog. *Phonetica* 22. 11–44.

Gordon, Matthew. 1995. Acoustic properties of primary and secondary word-level stress in Estonian. Poster presented at the 130th Meeting of the Acoustical Society of America, St Louis.

Gordon, Matthew. 1997. Phonetic correlates of stress and the prosodic hierarchy in Estonian. In Jaan Ross & Ilse Lehiste. *Estonian prosody: Papers from a symposium*, 100–124. Tallinn: Institute of Estonian Language.

Gordon, Matthew. 1998. A fortition-based approach to Balto-Fennic-Sámi consonant gradation. *Folia Linguistica Historica* 18. 49–79.

Gordon, Matthew. 2001. Syncope induced metrical opacity as a weight effect. *Proceedings of the West Coast Conference on Formal Linguistics* 20. 206–219.

Gordon, Matthew. 2002. A phonetically-driven account of syllable weight. *Language* 78. 51–80.

Gordon, Matthew. 2003. The phonology of pitch accents in Chickasaw. *Phonology* 20. 173–218.

Gordon, Matthew. 2004. A phonetic and phonological study of word-level stress in Chickasaw. *International Journal of American Linguistics* 70. 1–32.

Gordon, Matthew. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory* 23. 595–653.

Gordon, Matthew & Ayla Applebaum. 2010. Acoustic correlates of stress in Turkish Kabardian. *Journal of the International Phonetic Association* 40. 35–58.

Gordon, Matthew, Pamela Munro & Peter Ladefoged. 2000. Some phonetic structures of Chickasaw. *Anthropological Linguistics* 42. 366–400.

Gordon, Matthew & Pamela Munro. 2007. A phonetic study of final vowel lengthening in Chickasaw. *International Journal of American Linguistics* 73. 293–330.

Gouskova, Maria. 2003. Deriving economy: Syncope in Optimality Theory. Ph.D. dissertation, University of Massachusetts, Amherst (ROA-610).

Gregores, Emma & Jorge A. Suárez. 1967. *A description of colloquial Guaraní*. The Hague: Mouton.

Gruzov, Leonid Petrovich. 1960. *Sovremenyi mariiskii yazyk: Fonetika*. Joshkar-ola: Mariiskoe Knizhnoe Izdatel'stvo.

Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 11. 429–446.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.

Helimski, Eugene. 1998. Nganasan. In David Abondolo (ed.) *The Uralic languages*, 480–515. London & New York: Routledge.

Horne, Elinor Clark. 1974. *Javanese–English dictionary*. New Haven & London: Yale University Press.

Hualde, José Ignacio. 1989. Autosegmental and metrical spreading in the vowel-harmony systems of northwestern Spain. *Linguistics* 27. 773–805.

Hualde, José Ignacio. 1998. Asturian and Cantabrian metaphony. *Rivista di Linguistica* 10. 99–108.

Hyde, Brett. 2002. A restrictive theory of metrical stress. *Phonology* 19. 313–359.

Hyman, Larry M. 1989. Accent in Bantu: An appraisal. *Studies in the Linguistic Sciences* 19. 115–134.

Hyman, Larry M. 2001. The limits of phonetic determinism in phonology: *NC revisited. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 141–185. San Diego: Academic Press.

Hyman, Larry M. 2006. Word-prosodic typology. *Phonology* 23. 225–257.

Jassem, Wiktor, John Morton & Maria Steffen-Batóg. 1968. The perception of stress in synthetic speech-like stimuli by Polish listeners. *Speech Analysis and Synthesis* 1. 289–308.

Jun, Sun-Ah. 1993. The phonetics and phonology of Korean prosody. Ph.D. dissertation, Ohio State University.

Jun, Sun-Ah & Cécile Fougeron. 1995. The accentual phrase and the prosodic structure of French. In Kjell Elenius & Peter Branderud (eds.) *Proceedings of the 13th International Congress of the Phonetic Sciences*, vol. 2, 722–725. Stockholm: KTH & Stockholm University.

Kakumasu, James. 1986. Urubu-Kaapor. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.) *Handbook of Amazonian languages*, vol. 3, 326–406. Berlin & New York: Mouton de Gruyter.

Keating, Patricia, Taehong Cho, Cécile Fougeron & Chai-Shune Hsu. 2003. Domain-initial articulatory strengthening in four languages. In John Local, Richard Ogden & Rosalind Temple (eds.) *Phonetic interpretation: Papers in laboratory phonology VI*, 145–163. Cambridge: Cambridge University Press.

Kirchner, Robert. 2001. *An effort-based approach to consonant lenition*. New York & London: Routledge.

Kirchner, Robert. 2004. Consonant lenition. In Hayes *et al.* (2004), 313–345.

Kryvitskij, A. A. & A. I. Podluzhni. 1994. *Uchebnik belorusskogo jazyka dlja samoobrazovanija*. Minsk: Vysheishaia Shkola.

Lavoie, Lisa. 2001. *Consonant strength: Phonological patterns and phonetic manifestations*. New York: Garland.

Leer, Jeff. 1985. Prosody in Alutiiq (the Koniag and Chugach dialects of Alaskan Yupik). In Michael Krauss (ed.) *Yupik Eskimo prosodic systems: Descriptive and comparative studies*, 77–133. Fairbanks: Alaska Native Language Center.

Lehiste, Ilse. 1965. Vowel quantity in word and utterance in Estonian. *Congressus Secundus Internationalis Fenno-Ugristarum*, 293–303. Helsinki: Societas Fenno-Ugrica.

Lehiste, Ilse. 1966. *Consonant quantity and phonological units in Estonian*. Bloomington: Indiana University Press.

Levi, Susannah V. 2005. Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association* 35. 73–97.

Lieberman, Philip. 1960. Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America* 32. 451–454.

Lombardi, Linda. 2001. Why Place and Voice are different: Constraint-specific alternations in Optimality Theory. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*, 13–45. Cambridge: Cambridge University Press.

Maiden, Martin. 1995. Evidence from the Italian dialects for the internal structure of prosodic domains. In John Charles Smith & Martin Maiden (eds.) *Linguistic theory and the Romance languages*, 115–131. Amsterdam & Philadelphia: John Benjamins.

Mellander, Evan W. 2001. Quantitative processes in trochaic systems. *Proceedings of the West Coast Conference on Formal Linguistics* 20. 414–427.

Michelson, Karin. 1988. *A comparative study of Lake-Iroquoian accent*. Dordrecht: Kluwer.

Munro, Pamela. 1996. The Chickasaw sound system. Unpublished ms., University of California, Los Angeles.

Munro, Pamela 2005. Chickasaw. In Heather Hardy & Janine Scancarelli (eds.) *Native languages of the southeastern United States*, 114–156. Lincoln: University of Nebraska Press.

Munro, Pamela & Peter John Benson. 1973. Reduplication and rule ordering in Luiseño. *International Journal of American Linguistics* 39. 15–21.

Munro, Pamela & Charles Ulrich. 1984. Structure-preservation and Western Muskogean rhythmic lengthening. *Proceedings of the West Coast Conference on Formal Linguistics* 3. 191–202.

Munro, Pamela & Catherine Willmond. 1994. *Chickasaw: An analytical dictionary*. Norman, OK: University of Oklahoma Press.

Munro, Pamela & Catherine Willmond. 2005. *Chikashshanompa' kilanompoli'*. Los Angeles: UCLA Academic Publishing.

Nivens, Richard. 1992. A lexical phonology of West Tarangan. In Donald Burquest & Wyn Laidig (eds.) *Phonological studies in four languages of Maluku*, 127–227. Arlington: Summer Institute of Linguistics.

Penny, Ralph J. 1978. *Estudio estructural del habla de Tudanca*. Tübingen: Niemeyer.

40 The Foot

Michael Hammond

1 Overview

The metrical foot organizes the syllables of words into higher-order units built around stressed syllables.

In this chapter, we review the evidence for, and structure of, the foot. Along the way, we treat some of the major issues that have arisen in the development of this notion.

The organization of this chapter is as follows. First, we review the background against which the foot was proposed: linear generative phonology and then early footless metrical phonology. We then turn to the earliest foot proposals and the arguments advanced at the time, including arguments from stress theory and prosodic morphology. We then go on to consider how the theory of the foot has changed in Optimality Theory (OT). (For a more general discussion of stress, see CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE.)

2 Background

In this section, we lay out the necessary background for understanding the earliest proposals about the foot. First and foremost is the background of generative phonology generally and Chomsky and Halle (1968) specifically. We then go on to consider the foundation for the foot laid in early metrical theory.

2.1 *Generative phonology and* SPE

Here we discuss Chomsky and Halle (1968; henceforth *SPE*) as the foundation for the foot.[1] The main contribution of *SPE* for our purposes is an explicit treatment of the regularities of stress in English. The analysis is comprised of a number of

[1] Some of the issues in this section are developed further in CHAPTER 51: THE PHONOLOGICAL WORD, CHAPTER 116: SENTENTIAL PROMINENCE IN ENGLISH and CHAPTER 41: THE REPRESENTATION OF WORD STRESS.

rules written in the specific formalism proposed there. Their full rule for main stress in English is given in (1) (*SPE*: 240).

(1)

$$V \rightarrow [1\ \text{stress}] \ / \left[X \_\_ C_0 \left(\begin{bmatrix} -\text{tense} \\ \gamma\text{stress} \\ V \end{bmatrix} C_0^1 \left(\begin{bmatrix} \alpha\text{voc} \\ \alpha\text{cons} \\ -\text{ant} \end{bmatrix} \right) \right) \right.$$

$$\left. / \_\_ \left\langle \left(\begin{Bmatrix} (_{\text{c}}fik)At \\ [+D]C_0 \end{Bmatrix} \right) \begin{Bmatrix} \langle_1 + C_0 \rangle_1 \begin{bmatrix} -\text{stress} \\ -\text{tense} \\ -\text{cons} \end{bmatrix} [+\text{cons}]_0 \\ \langle_1 \begin{bmatrix} -\text{seg} \\ \langle_2 -\text{FB} \rangle_2 \end{bmatrix} \rangle_1 C_0[\beta\text{stress}]C_0 \langle_2 V_0 C_0 \rangle_2 \end{Bmatrix} \right\rangle \right]_{\langle \text{NSP} \langle_1 \text{VA} \rangle_1 \rangle}$$

Conditions: $\beta = \begin{Bmatrix} 2 \\ 1 \end{Bmatrix}$

$\gamma \leq 2$

X contains no internal #

Setting aside morphological and diacritic variables, and focusing on nouns in particular, the main stress rule assigns main stress and secondary stress to nouns in the following way. First, a [+tense] vowel in a final syllable gets primary or secondary stress, e.g. *kangaroo* [ˌkæŋgə'ru] or *chickadee* ['ʧɪkəˌdi]. If the final syllable does not have a stress, and the penultimate vowel is followed by appropriate consonants, then it gets stress, e.g. *agenda* [ə'ʤɛndə]. Likewise, if the penultimate vowel is [+tense], it gets stress as well, e.g. *aroma* [ə'romə]. Finally, in other cases, the antepenult receives stress, e.g. *America* [ə'mɛrəkə] or *remedy* ['rɛmədi].

Abstracting away from the formalism of the time, what we see is a restriction of the primary stress to the last three syllables of the word and a pressure to stress syllables of appropriate weight.

The alternating stress rule in (2) (*SPE*: 240) is responsible for stresses further to the left, e.g. the first stress in *kangaroo* [ˌkæŋgə'ru]. Like the main stress rule, it places stress subject to a basically alternating pattern.

(2) $V \rightarrow [1\ \text{stress}] \ / \_\_ C_0 \ (=) \ C_0 V C_0 [1\ \text{stress}] C_0]_{NAV}$

Both rules alternate in a similar way; both rules assign stress with respect to a following landmark. The fact that both rules exhibit similar patterns and the fact that this kind of alternation is ubiquitous in other languages was a missed generalization in *SPE*, and one that found an explanation in the development of the metrical foot.

A second important aspect of the analysis proposed in *SPE* is that, unlike other phonological features, the feature [stress] exhibited more than two levels in the phonology. Thus, while a feature like [high] could have the values [+high] or [−high] in the phonology, the feature stress could have an *infinite* number of values: [0stress], [1stress], [2stress], [3stress], etc. These different numerical values corresponded to degrees of stress that were held to be contrastive. Moreover, the values had a rather odd interpretation, where [0stress] is the least stressed element and [1stress] has the most stress. Since the values can increase without bound, as the integer value increases, the degree of stress gets smaller, but never quite reaches [0stress].

Rischel (1972) was the earliest proposal to replace aspects of the *SPE* stress system with a hierarchical tree. Specifically, Rischel proposes that compound stress in Danish does not require a cycle and that degrees of stress can be easily read off the morphosyntactic tree.

Compare the following two compounds in Danish:

(6)	*fædrelandssang*	'patriotic song'	[father-land]song
	perlehalsbånd	'pearl necklace'	pearl[neck-band]

In the first case, the compound is left-branching and has the stress values 132. In the second case, the compound is right-branching and has the stress pattern 123. The *SPE* rules given for English here would actually accommodate these directly, as shown in the derivations below.

(7) $[_N [_N$ fædre lands $]_N$ sang $]_N$

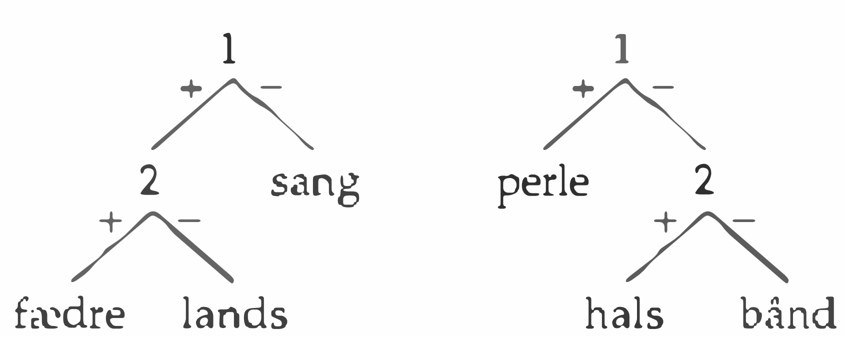
fædre	lands	sang	
1	1	1	NSR
1	2	–	CSR
1	3	2	CSR

(8) $[_N$ perle $[_N$ hals bånd $]_N]_N$

perle	hals	bånd	
1	1	1	NSR
–	1	2	CSR
1	2	3	CSR

Rischel proposes that cyclic effects can be gotten by reading stress levels directly off of trees. He gives trees like the following for the examples above. The pluses and minuses reflect the relative strength of left and right branches and the numbers on nodes reflect the relative effects of those strengths at different levels of the tree.

(9)



Rischel does not propose a specific algorithm for reading stress values off trees like these, but it is easy to see that various interpretations will produce what appear to be reasonable values. The gist is that reapplication of stress rules per se is not required to get the same kind of cyclic effects cited above from *SPE*.[2]

Liberman and Prince (1977) made a similar proposal a few years later, proposing a fairly complete analysis of English stress along similar lines. Basically, they propose that [stress] be treated as a binary feature, with the values [+stress] and

[2] Ultimately, Kiparsky (1979) argued that cyclicity is still necessary in a metrical theory of stress. The debate resurfaced again a few years later. See Hammond (1989), Halle and Kenstowicz (1991), and Cole and Coleman (1992) for more discussion.

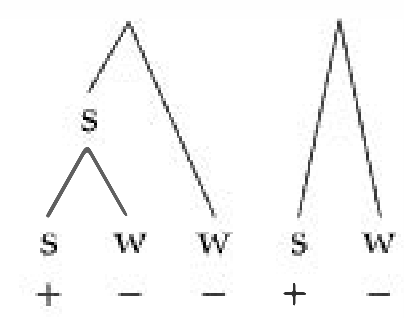
As noted above, the labeling of the lower-level trees is unambiguous, because of a general constraint against [+stress] in weak position. The higher-level trees are labeled in accord with the Lexical Category Prominence Rule (LCPR; 1977: 270).

(18) In the configuration $[N_1N_2]$, N_2 is strong *iff* it branches.

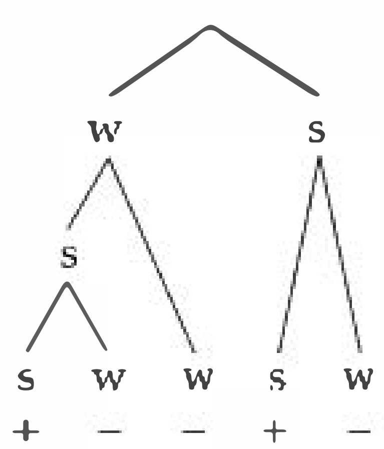
Let us take a look at an example: *Winnepesaukee* [ˌwɪnapəˈsɔki]. First, [+stress] values are assigned by the ESR, producing:

(19) + − − + −
Winnepesaukee

Syllables are gathered into feet as below:

(20)


Finally, the feet are gathered into a tree, the right node of which is labeled strong, since it is branching.

(21)


What is important about this entire tree-construction and tree-labeling procedure is that it explicitly recognizes two levels: a *foot* level and a higher *word* level. This is the first step toward an explicit theory of the foot. Liberman and Prince showed how the foot could be employed in a reanalysis of the basic stress facts of English that *SPE* introduced.

3 Why we need feet

The next step was the parametric elaboration of the foot. At around the same time as Liberman and Prince (1977), Hyman (1977) offered the first typological treatment of stress. While he was not able to go very far in terms of the technical analysis offered, this paper was an important catalyst in forcing phonologists interested in stress to look at the broader typological implications of their work.

The first parametric approaches to the metrical foot showed up in Halle and Vergnaud (1978) and McCarthy (1979), but the most influential early proposal was that of Hayes (1980). Let us look at the Hayes proposal in some depth.

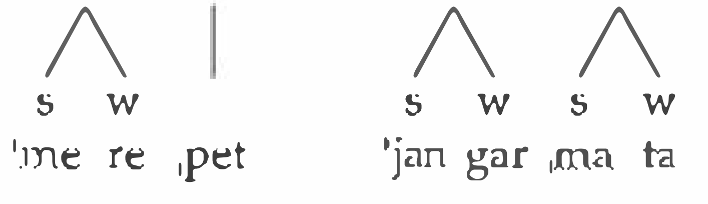
Hayes offered a theory of the foot based on the trees proposed in Liberman and Prince. In particular, feet were parameterized for the following:

(22) a. *Headedness*
Is the designated terminal element – the strongest element of the foot – on the left edge or the right?
b. *Boundedness*
Are feet binary or unbounded? Do feet contain only two syllables or as many as possible?
c. *Directionality*
Are feet built left-to-right or right-to-left?
d. *Iterativity*
Are feet constructed iteratively or not? That is, is only a single foot built on some edge or are as many feet built as possible?
e. *Quantity-sensitivity*
There are three choices here. First, feet can be quantity-sensitive (QS): weak nodes cannot dominate heavy syllables. Second, feet can be quantity-insensitive (QI): syllable weight is irrelevant. Last, feet can be obligatory-branching (OB): in OB feet, strong nodes must dominate heavy syllables and weak nodes may not.
f. *Syllable weight*
If feet are sensitive to syllable weight, are they sensitive to the weight of the syllable nucleus or the syllable rhyme?

Let us go through some of the examples Hayes cites in support of this theory. Maranungku (Tryon 1970) is cited as an example of left-headed binary left-to-right QI feet. Here, main stress falls on the first syllable of the word and secondary stresses fall on alternating syllables to the right.

(23)	ˈtiralk	'saliva'
	ˈmereˌpet	'beard'
	ˈjangarˌmata	'the Pleiades'
	ˈlangkaˌrateˌti	'prawn'
	ˈweleˌpeneˌmanta	'kind of duck'

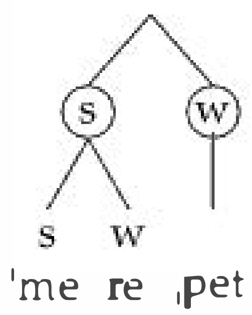
Here are two examples:

(24)


Notice how the left-to-right construction of feet is apparent from the fact that in words with an odd number of syllables, a monosyllabic, or *degenerate*, foot is built on the right.

The difference between primary and secondary stress is captured by positing a higher level of structure: the *word tree*. These are left- or right-headed unbounded trees built on the roots of feet (see CHAPTER 41: THE REPRESENTATION OF WORD STRESS). In Maranungku, the word tree is left headed.

(25)



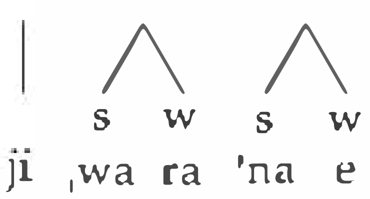
Note that in this and in subsequent diagrams we circle the roots of feet when the word tree is represented.[3]

Warao (Osborn 1966) provides an example of right-to-left footing with left-headed binary feet.

(26)	ˌjapuˌrukiˌtane'hase	'verily to climb'
	ˌnahoˌroaˌhaku'tai	'the one who ate'
	jiˌwara'nae	'he finished it'
	eˌnahoˌroaˌhaku'tai	'the one who caused him to eat'

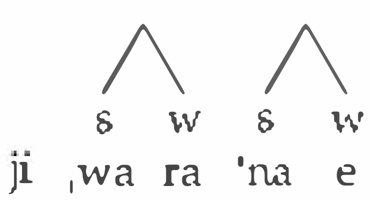
Warao differs from Maranungku in that, in words with an odd number of syllables, there is no initial degenerate foot; rather that foot is removed by an additional *destressing rule*. For example, [jiˌwara'nae] is first footed as follows:

(27)



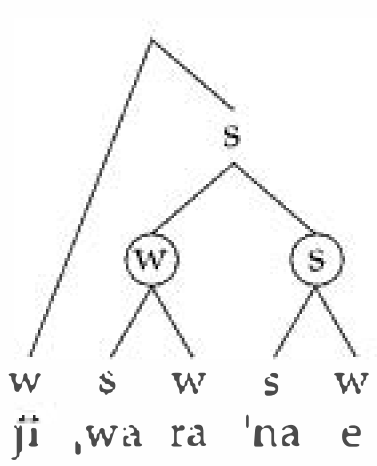
This intermediate representation is then converted to:

(28)



The word tree in Warao is right headed. Hayes assumes that unfooted syllables are adjoined as weak nodes to the word tree. For example:

(29)



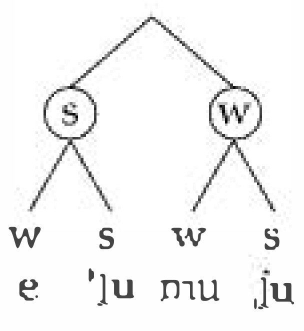
Hayes cites Weri (Boxwell and Boxwell 1966) as an example of binary right-headed feet constructed from right to left with a right-headed word tree.

[3] Hayes (1980) uses underlining, rather than circles.

(34) wu'le 'tomorrow'
ţi'panto 'year'
e'lumu,ju 'give us'
e'lua,enew 'he will give me'
ki'mufa,luwu,laj 'he pretended not to know'

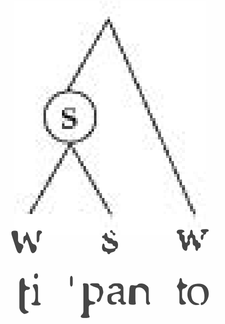
The analysis here is binary right-headed feet built left to right.

(35)



Monosyllabic feet are generally disallowed (or removed) in odd-syllabled cases:

(36)

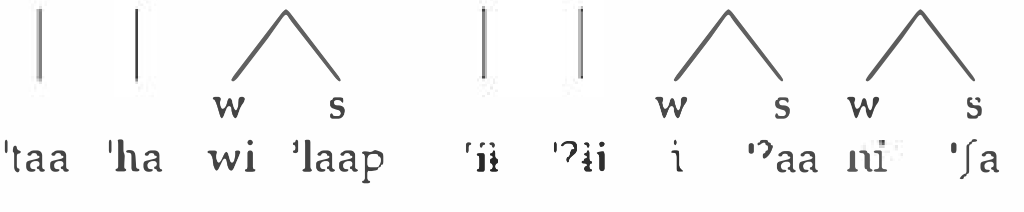


Let us now consider quantity-sensitivity (QS; see CHAPTER 57: QUANTITY-SENSITIVITY). This parameter allows heavy syllables to attract stress. Hayes cites Tübatulabal (Voegelin 1935) as an example of right-to-left right-headed QS bounded feet. The generalization is that stress falls on (a) the final syllable, (b) any long vowel, and (c) any vowel that is two syllables to the left of a stress. Since stresses are unranked, there is no word tree.

(37) 'taa'hawi'laap 'in the summer'
pɨ'tipɨ'tɨɨdi'nat 'he is turning it over repeatedly'
'ɨɨ'ʔɨɨi'ʔaani'ʃa 'he will meat-fast'
'pɔnih'wɨn 'of his own skunk'

Here are two examples of the footings produced by these parameter settings.

(38)



Notice that long vowels count as heavy in Tübatulabal; thus QS refers to the nucleus, not the rhyme.[5]

Hayes' theory of feet also allows for unbounded feet. When these are quantity-insensitive, they simply position stress on the first or last syllable of the word. No actual examples are cited, but we would expect trees like the following for a language with initial stress and QI left-headed unbounded feet:

[5] The data cited by Hayes do not establish unequivocally that codas do not contribute to weight.

We have only exemplified some of the combinations of parameter settings that this theory allows. The claim of the theory is that all settings can be freely combined and that the set of possible stress languages is fully defined by these settings.

The argument for feet per se comes from their role in this parametric system. If the set of possible stress languages is best defined in terms of a theory that adopts the foot as a central descriptive device, then the typology of stress is an argument for the foot.

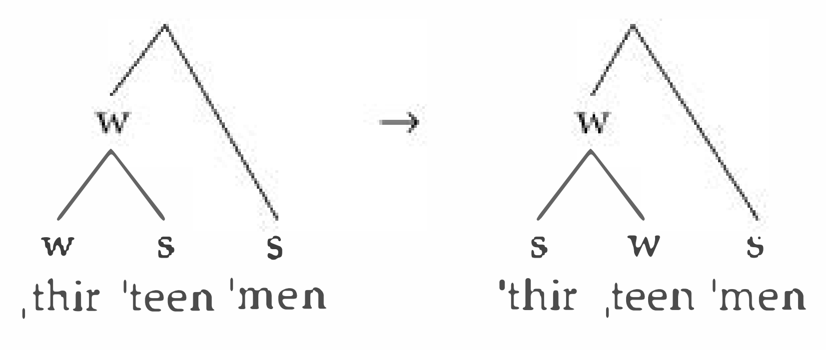
4 Do we need constituency?

One could argue that while the foot is a central computational device in the system Hayes develops, the full predictive power of the foot in that system is not exploited; specifically, while the foot is a *constituent* in metrical trees, its constituency plays no specific role in the system.

Prince (1983) takes this observation to its logical conclusion, proposing an alternative metrical theory without constituency and without feet. To understand this proposal, let us return to Liberman and Prince (1977) and their theory of the *metrical grid.*

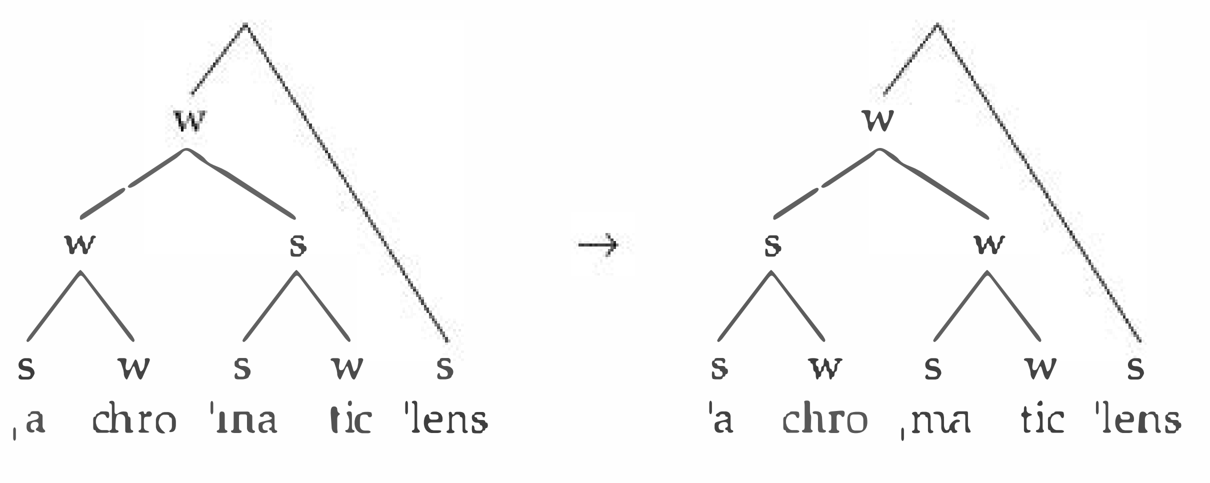
Liberman and Prince propose the metrical grid as a mechanism for identifying the environment for the *rhythm rule*, the phenomenon whereby stress is shifted in certain contexts. Thus, when a word like *ˌthir'teen* is combined with *'men*, we get a shift of stress in the former: *'thirˌteen 'men*. The effect of this shift can be diagrammed as follows:

(46)



Interestingly, the shift also happens with phrases like *'achroˌmatic 'lens*, but not with phrases like *ˌMon'tana 'cowˌboy*. Why this should be the case is not apparent from the metrical trees.

(47)



```
(54)  Feet          Halle and Vergnaud (1987)   Hayes (1995)

      degenerate              x
                             (x)                    (x)
                              σ                      σ

      binary                x
                           (x  x)                  (x  .)
                            σ  σ                    σ  σ
```

The other proposal in response to parallel tree and grid representations of stress was that of Prince (1983). Specifically, Prince proposed a grid-only theory of stress *without* the foot. The basic idea behind the proposal as far as feet are concerned is that binary patterns of iteration are replaced by appeal to the *perfect grid*. This device allows for a binary pattern of stress to be assigned in one of four ways, depending on whether the assignment is from left to right or from right to left and on whether one begins with a stressed syllable or a stressless syllable.

```
(55)  Feet          Left-to-right (LR)     Right-to-left (RL)

      peak first    x   x   x                   x   x   x
                    x x x x x                   x x x x x
                    σ σ σ σ σ ...           ... σ σ σ σ σ

      trough first    x   x                       x   x
                    x x x x x                   x x x x x
                    σ σ σ σ σ ...           ... σ σ σ σ σ
```

Notice how this pattern is achieved with no appeal to binary constituents.

To get the effect of word trees and unbounded foot construction, Prince proposes the *End Rule*. This device assigns a grid mark to the leftmost or rightmost element of the highest level of the grid present. If no stresses have already been assigned, the effect of the End Rule is to assign a stress to the first or last syllable of the word.

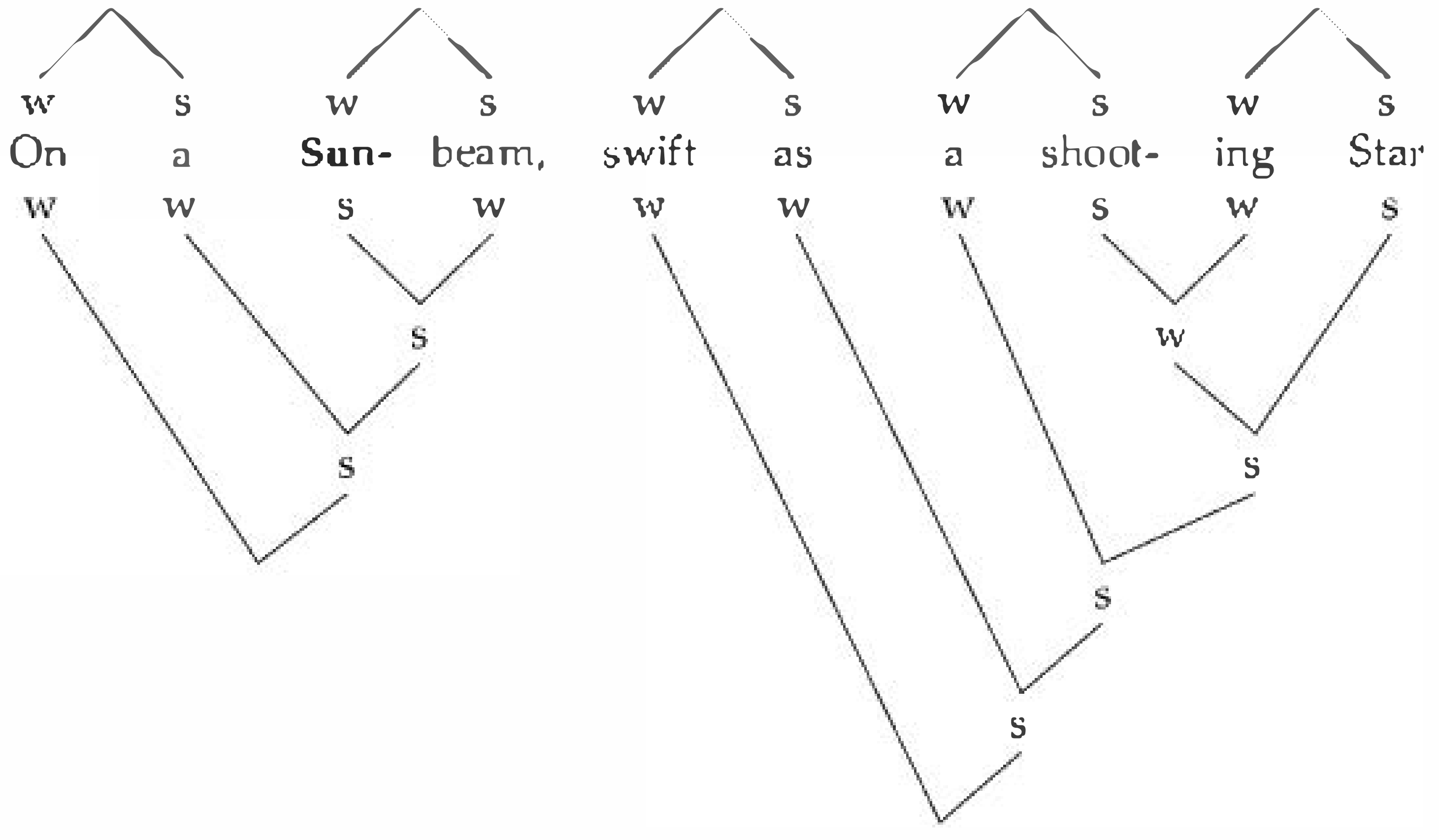
```
(56)  End Rule Left                    x
                        x x x x        x x x x
                        σ σ σ σ   →    σ σ σ σ

      End Rule Right                         x
                        x x x x        x x x x
                        σ σ σ σ   →    σ σ σ σ
```

If stresses are already present, however, then the effect of the End Rule is to promote the leftmost or rightmost stress to primary stress:

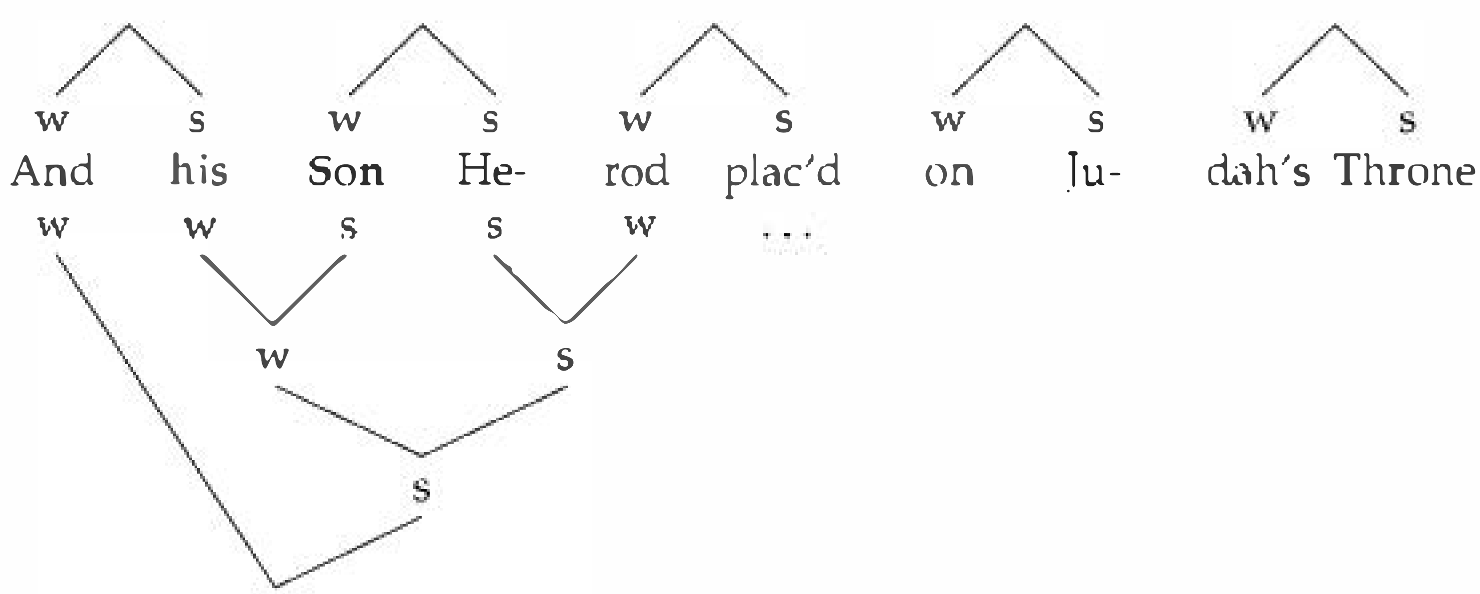
```
(57)  End Rule Left                    x
                        x   x          x   x
                        x x x x        x x x x
                        σ σ σ σ   →    σ σ σ σ
```

(69)



Finally, examples like the following show that the element must be the strongest element of the phrase, and that the bracketing restrictions do not suffice of themselves (Paradise Regained 2.424).

(70)



Hayes (1983) argues that references to foot constituency can be done away with if we define stress peaks over metrical grids and refer to higher-level prosodic constituency. His version of Milton I looks like this:

(71) *Milton I* (grid version)
*Peak / [. . . __]$_{phrase}$

A peak is defined in terms of the grid as a grid column that is higher than at least one of its neighbors.

Let us now look at how this constraint separates the cases we have considered so far. Hayes represents grids in terms of a single symbol "x," rather than numbers. In addition, he represents the line template as a simple single-level grid, rather than with nodes labeled "s" and "w." For the line in (68), we would have this template:

(72) . x . x . x . x . x
Resembling strong youth in his middle age

(79)	*Word*	*Legal*	*Illegal*
	[ˌMinne]['sota]	Minne-fuckin'-sota	Mi-fuckin'-nnesota Minneso-fuckin'-ta
	[ˌTim][ˌbuk]['tu]	Tim-fuckin'-buktu Timbuk-fuckin'-tu	–
	[ˌHali][ˌcar]['nassus]	Hali-fuckin'-carnassus Halicar-fuckin'-nassus	Ha-fuckin'-licarnassus Halicarna-fuckin'-ssus
	[ˌApa][ˌlachi]['cola]	Apa-fuckin'-lachicola Apalachi-fuckin'-cola	A-fuckin'-palachicola Apala-fuckin'-chicola Apalachico-fuckin'-la

Strikingly, there are multiple infixation possibilities just in case we find two medial stressless syllables in a row. This follows directly from the claim that feet in English are binary.

(80)	*Word*	*Legal*	*Illegal*
	[ˌWinne]pe['saukee]	Winne-fuckin'-pesaukee Winnepe-fuckin'-saukee	Wi-fuckin'-nnepesaukee Winnepesau-fuckin'-kee
	[ˌKala]ma['zoo]	Kala-fuckin'-mazoo Kalama-fuckin'-zoo	Ka-fuckin'-lamazoo

The second stressless syllable is affiliated with neither of the adjacent feet allowing the infix to be positioned to either side of it, still satisfying the requirement that there be feet to each side and that the primary stress follows.

There are additional complications to the system (Hammond 1997, 1999). First, the main stress cannot precede the infix. Thus *ˌKalama-fuckin'-'zoo* is decidedly better than *'catama-fuckin'-ˌran*. In addition, if the syllable preceding the infix is stressed, it must be at least bimoraic. Hence, *mun-fuckin'-dane* [ˌmʌnˌfʌkən'den] is better than *ra-fuckin'-ccoon* [ˌræˌfʌkən'kun].

Those complications notwithstanding, the locus of infixation provides additional evidence for foot constituency.

5.5 *Minimum word size*

Lardil (Wilkinson 1988) provides a nice example of a minimum word constraint based on the foot: words in Lardil must have at least two vowels. If they do not, then they are augmented to meet this target with an epenthetic [a]. This provides for alternations in the shape of the stem depending on whether it is suffixed or not; an unsuffixed sub-minimal stem is augmented. Verbs with at least two syllables are inflected as follows:

(81)		'tree'	'dugong'	'beach'	'inside'
	underlying	/ṯuŋal/	/kentapal/	/kela/	/wiṯe/
	uninflected	ṯuŋal	kentapal	kela	wiṯe
	non-future	ṯuŋalin	kentapalin	kelan	wiṯen
	future	ṯuŋaluɽ	kentapaluɽ	kelaɽ	wiṯeɽ

Monosyllabic consonant-final roots with long vowels behave in similar fashion.

(82)

	'ti-tree sp.'	'spear gen.'
underlying	/peeɾ/	/maaɳ/
uninflected	peeɾ	maaɳ
non-future	peeɾin	maaɳin
future	peeɾuɽ	maaɳkuɽ

However, nouns with only a single vowel get augmented when uninflected.

(83)

	'thigh'	'shade'
underlying	/teɾ/	/wik/
uninflected	teɾa	wika
non-future	teɾin	wikin
future	teɾuɽ	wikuɽ

The two-vowel target can be seen as foot based if we treat Lardil as Hayes (1980) treated Southern Paiute: each vowel element is a potential terminal element for footing. Alternatively, if we view vowels as the sole bearers of moras in Lardil, we can view this as a bimoraic target, which was later proposed to be a foot.[15]

5.6 *Language games*

Hammond (1990) discusses a language game in English that provides further evidence for the foot. The game is played by substituting names into the following rhyme.

(84)

Jack, Jack bo back	—, —, bo b —
[ʤæk ʤæk bo bæk]	
banana fana fo fack	banana fana fo f —
[bənænə fæ nə fo fæk]	
me my mo mack	me my mo m —
[mi maj mo mæk]	
Ja–ack	—
[ʤæ–æk]	

The onset of the name undergoes various substitutions not relevant here. The relevant point here is that the name must fit a particular prosodic template: from one to three syllables, where the first syllable is stressed and any subsequent syllables are stressless.

This corresponds to a single left-headed binary foot plus an optional extrametrical syllable. Marking feet with square brackets and the extrametrical syllable with angled brackets, we get a clear difference between names that are acceptable and those that are not.

(85)

Permissible	*Impermissible*
['Jack]	A[n'nette]
['To]<ny>	[ˌIsa]['do]<ra>
['Jenni]<fer>	['Mira][ˌbeil]
['Gwendo]<lyn>	O['livi]<a>

15 Garrett (1999) argues, though, that word minimality is not connected to foot structure.

Burzio, Luigi. 1994. *Principles of English stress*. Cambridge: Cambridge University Press.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
Cole, Jennifer & John Coleman. 1992. No need for cyclicity in generative phonology. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 28. 36–50.
Dixon, R. M. W. 1977a. *A grammar of Yidiɲ*. Cambridge: Cambridge University Press.
Dixon, R. M. W. 1977b. Some phonological rules in Yidiɲ. *Linguistic Inquiry* 8. 1–34.
Echeverría, Max S. & Heles Contreras. 1965. Araucanian phonemics. *International Journal of American Linguistics* 31. 132–135.
Fabb, Nigel & Morris Halle. 2008. *Meter in poetry: A new theory*. Cambridge: Cambridge University Press.
Garrett, Edward, 1999. *Minimal words aren't minimal feet. UCLA Working Papers in Linguistics 1, Papers in Phonology* 2. 68–105 (ROA-1031).
Gordon, Matthew. 2002. A factorial typology of quantity-insensitive stress. *Natural Language and Linguistic Theory* 20. 491–552.
Halle, Morris & Michael Kenstowicz. 1991. The Free Element Condition and cyclic versus noncyclic stress. *Linguistic Inquiry* 22. 457–501.
Halle, Morris & Jean-Roger Vergnaud. 1978. Metrical structures in phonology. Unpublished ms., MIT.
Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress*. Cambridge, MA: MIT Press.
Hammond, Michael. 1982. Foot-domain rules and metrical locality. *Proceedings of the West Coast Conference on Formal Linguistics* 1. 207–218.
Hammond, Michael. 1986. The obligatory-branching parameter in metrical theory. *Natural Language and Linguistic Theory* 4. 185–228.
Hammond, Michael. 1988. *Constraining metrical theory: A modular theory of rhythm and destressing*. New York: Garland.
Hammond, Michael. 1989. Cyclic secondary stresses in English. *Proceedings of the West Coast Conference on Formal Linguistics* 8. 139–153.
Hammond, Michael. 1990. The Name Game and onset simplification. *Phonology* 7. 159–162.
Hammond, Michael. 1997. Vowel quantity and syllabification in English. *Language* 73. 1–17.
Hammond, Michael. 1999. *The phonology of English: A prosodic optimality-theoretic approach*. Oxford: Oxford University Press.
Hayes, Bruce. 1980. A metrical theory of stress rules. Ph.D. dissertation, MIT. Published 1985, New York: Garland.
Hayes, Bruce. 1982. Metrical structure as the organizing principle of Yidiɲ phonology. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part 1, 97–110. Dordrecht: Foris.
Hayes, Bruce. 1983. A grid-based theory of English meter. *Linguistic Inquiry* 14. 357–394.
Hayes, Bruce. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15. 33–74.
Hayes, Bruce. 1987. A revised parametric metrical theory. *Papers from the Annual Meeting of the North East Linguistic Society* 17. 274–289.
Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.
Hyde, Brett. 2002. A restrictive theory of metrical stress. *Phonology* 19. 313–359.
Hyman, Larry M. 1977. On the nature of linguistic stress. In Larry M. Hyman (ed.) *Studies in stress and accent*, 37–82. Los Angeles: Department of Linguistics, University of Southern California.
Idsardi, William J. 1992. The computation of prosody. Ph.D. dissertation, MIT.
Kager, René. 1993. Alternatives to the iambic-trochaic law. *Natural Language and Linguistic Theory* 11. 381–432.
Kahn, Daniel. 1980. *Syllable-based generalizations in English phonology*. New York: Garland.
Kessler, Brett & Rebecca Treiman. 1997. Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language* 37. 295–311.

Key, Harold H. 1961. Phonotactics of Cayuvava. *International Journal of American Linguistics* 27. 143–150.

Key, Harold H. 1967. *Morphology of Cayuvava*. The Hague: Mouton.

Kiparsky, Paul. 1973. "Elsewhere" in phonology. In Stephen R. Anderson & Paul Kiparsky (eds.) *A Festschrift for Morris Halle*, 93–106. New York: Holt, Rinehart & Winston.

Kiparsky, Paul. 1977. The rhythmic structure of English verse. *Linguistic Inquiry* 8. 189–247.

Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10. 421–441.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.

Marantz, Alec. 1982. Re reduplication. *Linguistic Inquiry* 13. 435–482.

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.

McCarthy, John J. 1982. Prosodic structure and expletive infixation. *Language* 58. 574–590.

McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.

McCarthy, John J. & Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University.

Osborn, Henry A. 1966. Warao I: Phonology and morphophonemics. *International Journal of American Linguistics* 32. 108–123.

Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14. 19–100.

Rischel, Jørgen. 1972. Compound stress in Danish without a cycle. *Annual Report of the Institute of Phonetics* 6. 211–228.

Sapir, Edward. 1930. Southern Paiute, a Shoshonean language. *Proceedings of the American Academy of Arts and Sciences* 65. 1–296.

Sebeok, Thomas A. & Frances J. Ingemann. 1961. *An Eastern Cheremis manual: Phonology, grammar, texts and glossary*. Bloomington: Indiana University.

Smolensky, Paul & Géraldine Legendre (eds.) 2006. *The harmonic mind: from neural computation to optimality-theoretic grammar*. Cambridge, MA: MIT Press.

Street, John C. 1963. *Khalkha structure*. Bloomington: Indiana University & The Hague: Mouton.

Treiman, Rebecca & Catalina Danis. 1988. Syllabification of intervocalic consonants. *Journal of Memory and Language* 27. 87–104.

Treiman, Rebecca & Andrea Zukowski. 1990. Toward an understanding of English syllabification. *Journal of Memory and Language* 29. 66–85.

Tryon, Darrell T. 1970. *An introduction to Maranungku (Northern Australia)*. (*Pacific Linguistics* B15.) Canberra: Australian National University.

Voegelin, Charles F. 1935. Tübatulabal grammar. *University of California Publications in American Archaeology and Ethnology* 34. 55–189.

Walker, Rachel. 1997. Mongolian stress, licensing and factorial typology. Unpublished ms., University of California, Santa Cruz (ROA-171).

Wilkinson, Karina. 1988. Prosodic structure and Lardil phonology. *Linguistic Inquiry* 19. 325–334.

41 The Representation of Word Stress

Ben Hermans

1 Introduction

Since the publication of Hayes (1985), the asymmetries between iambs and trochees have been a central theme in the literature on stress. Two types of asymmetries can be distinguished. One has to do with quantity. It has frequently been claimed that iambs do not allow a heavy syllable in the weak position, and require a heavy syllable in the strong position. Kager (1993) deals with asymmetries of this type on the basis of a theory whose central hypothesis is that feet are built over moras, rather than syllables. The recent literature, however, has shown that these "quantitative asymmetries" are not supported empirically. It is simply not true, for instance, that iambs invariably constrain the occurrence of heavy and light syllables in the way just described. A particularly convincing argument to this effect is given in Altshuler (2009), with respect to Osage. (See also CHAPTER 44: THE IAMBIC–TROCHAIC LAW and CHAPTER 57: QUANTITY-SENSITIVITY.)

There is a second class of asymmetries, however, which remains valid. These are "parsing asymmetries," which have to do with the direction of foot construction in a word. Two authors have argued that such parsing asymmetries can only be explained if the representation of word stress is fundamentally changed. Interestingly, however, they disagree as to *how* it should be changed. While Gordon (2002) proposes to *simplify* the representation of word stress by eliminating foot structure completely, Hyde (2001, 2002, 2008) recognizes not only foot structure but, in addition, a new type of structure, the "overlapping foot," thus *complicating* the representation of stress in order to account for the asymmetries.

In this chapter I consider the ongoing debate about the representation of word stress from the perspective of parsing asymmetries. In §2, after presenting some of the most important asymmetries, I briefly sketch Gordon's account, which is as simple as it is radical. In his view, asymmetries can be accounted for if foot structure is abolished. Word stress representations contain only gridmarks, as in Prince (1983) and Selkirk (1984).

In the spirit of Gordon, then, we might say that feet are superfluous if we want to account for the *distribution* of stress in the words of the world's languages. This raises the question of whether feet are necessary at all. Interestingly, if we broaden our scope to include other phenomena as well, the evidence for foot structure becomes overwhelming. In §3 I present some cases from the recent literature which support the existence of feet, thus suggesting that representations with only gridmarks are too impoverished.

If foot structure does exist, then how do we account for the *distribution* of stress in the words of the world's languages? Is it still reasonable to formulate the relevant constraints in terms of the grid only? Or must they be stated in terms of foot construction, with gridmarks playing only a marginal role? In §4 I present an overview of Hyde's work, in which the claim is made that the distribution of stress can best be explained in terms of constraints regulating foot construction. Gridmarks only read off some of the basic properties of a word's foot structure. This is a continuation of the tradition initiated by Liberman and Prince (1977). Other authoritative studies, such as Hayes (1984, 1995), have argued for the same idea, which also led to development of the "bracketed grid" notation (Halle and Vergnaud 1987). There are some important differences between the various "tree-cum-grid" theories, however. In §5 I give a brief overview of one issue where theories seem to differ. This concerns the status of headedness in foot structure. Is a foot inherently headed, even if it is not accompanied by a gridmark? Or is it the case that a foot is inherently headless unless there is a gridmark accompanying it, marking one syllable as the head? In the first approach, some or all of the properties of foot structure can simply be read off the grid. In this view foot structure is primary and the grid secondary. In the second approach, in which the grid is imposed on foot structure, the grid is primary and foot structure secondary.

2 Explaining parsing asymmetries without foot structure

Hyde (2001, 2002) shows that some non-existing systems can easily be derived with generally accepted foot inventories (see CHAPTER 40: THE FOOT). Let us consider three of these cases. The first one is the Australian language Garawa, which can be compared with what Hyde (2002: 329) calls "Anti-Garawa," an unattested system.

(1) *Garawa* *Anti-Garawa*

```
 x   x   x           x   x   x
(1 2)(3 4)(5 6)     (1 2)(3 4)(5 6)
 x      x   x        x   x       x
(1 2) 3 (4 5)(6 7)  (1 2)(3 4) 5 (6 7)
```

In these representations foot structure is indicated by round brackets and headedness is represented by gridmarks. Garawa can be derived with the following rules. One trochee is built at the left edge. Then trochees are built from right to left. Furthermore, degenerate feet are not allowed. In odd-parity words this creates a lapse following the stressed initial syllable. If we change just two ingredients of this system we derive a non-existing pattern. One *iamb* is built at the *right* edge, and then a series of iambs from left to right. We then derive a system in which odd-parity words contain a lapse before the final (stressed) syllable.

The second example is a pair consisting of the Australian language Pintupi, which can easily be derived by the rules of the theory, and Anti-Pintupi, which can be derived just as easily but does not exist.

```
(2) Pintupi              Anti-Pintupi

     x   x   x             x   x   x
    (1 2)(3 4)(5 6)       (1 2)(3 4)(5 6)
     x   x   x                 x   x   x
    (1 2)(3 4)(5 6) 7     1 (2 3)(4 5)(6 7)
```

In Pintupi, trochees are built from left to right. Degenerate feet are not allowed. In odd-parity words this creates a lapse at the right edge of a word. Anti-Pintupi can be derived just as easily, by constructing iambs from right to left, so that the lapse is located at the left edge.

Finally, consider Piro, spoken in Brazil and Peru, and Anti-Piro, an impossible system.

```
(3) Piro                 Anti-Piro

     x   x   x               x   x   x
    (1 2)(3 4)(5 6)       (1 2)(3 4)(5 6)
     x   x     x             x     x   x
    (1 2)(3 4) 5 (6 7)    (1 2) 3 (4 5)(6 7)
```

In Piro, one trochee is built at the right edge. Remaining trochees are built from left to right; degenerate feet are not allowed. In odd-parity words this creates a lapse before the stressed penultimate syllable. Anti-Piro can be derived by changing just two ingredients. One iamb is built at the left edge; remaining iambs are built from right to left. In odd-parity words this creates a lapse after the peninitial syllable.

The three non-existent cases have in common that iambic structure refers to the right edge. In Anti-Carawa one iamb is constructed at the right edge, while in Anti-Pintupi and Anti-Piro foot construction starts at the right edge. One might suppose, then, that iambs cannot refer to a word's right edge. This, however, is not true; languages where iambs are constructed from right to left do exist. One example is Suruwaha, spoken in Brazil (Hyde 2002: 320), which is the mirror image of Araucanian, spoken in Chile and Argentina (Hyde 2002: 320). In this language iambs are constructed from left to right. These systems are illustrated in (4):

```
(4) Araucanian           Suruwaha

       x   x   x           x   x   x
    (1 2)(3 4)(5 6)       (1 2)(3 4)(5 6)
       x   x   x          x   x   x   x
    (1 2)(3 4)(5 6) 7     (1)(2 3)(4 5)(6 7)
```

In Araucanian the final syllable is unparsed in odd-parity words. In Suruwaha the first syllable of odd-parity words is assigned a degenerate foot.

(12) *Anti-Garawa resolved (Suruwaha)*

```
  x   x   x
1 2 3 4 5 6

x   x   x   x
1 2 3 4 5 6 7
```

Insertion of the initial gridmark creates a system with a binary rhythm. This is a possible system, which is exemplified by Suruwaha.

The same reasoning applies to Anti-Piro, whose pattern is given again in (13), without foot structure.

(13) *Anti-Piro*

```
  x   x   x
1 2 3 4 5 6

  x     x   x
1 2 3 4 5 6 7
```

The presence of stress on the final syllable indicates that ALIGNEDGES is high-ranked, as in Suruwaha. To ensure that in even-parity words the final syllable is stressed, ALIGN(x_1,R) must dominate ALIGN(x_1,L). With these rankings, it is impossible to derive the stress pattern of odd-parity words in (13). Again, this pattern has a lapse, and violates ALIGNEDGES. Both violations can be eliminated by inserting a gridmark over the first syllable, and by moving the gridmark of the second syllable to the third, as shown in (14).

(14) *Anti-Piro resolved (Suruwaha)*

```
  x   x   x
1 2 3 4 5 6

x   x   x   x
1 2 3 4 5 6 7
```

Again this pattern is actually attested, and is exemplified by Suruwaha.

I have shown in this section that grid-only frameworks can easily explain the asymmetries occurring in binary rhythmic patterns. They do so in a way that is as simple as it is radical. Foot structure is eliminated, and only stresses are preserved, represented as gridmarks. As far as the *distribution* of stress is concerned, grid-only approaches do rather well. Surely this explains why the grid-only approach is still adopted (see for recent applications Karvonen 2005, 2008; also Gordon 2003: 179, note 4, where he confidently states that "a grid-based theory of stress offers a closer fit to the typology of stress than foot-based metrical theories"). It seems, then, that grid-only approaches are sufficient, and that foot structure is therefore superfluous.

Interestingly, however, if we go beyond the *distribution* of stress, and broaden our scope to other phenomena, we find abundant evidence for feet. In the next section I will present a few arguments in favor of foot structure. Then, in §4, I will investigate what this means for the theory of stress.

tone receives word stress, because it is the head of a trochee. With a grid-only approach it is difficult to explain these facts. It seems as if an unstressed syllable with a low tone behaves differently from an unstressed syllable with a high tone; the former but not the latter seems to avoid a lapse. By its very nature, the grid structure of an unstressed syllable with a low tone is identical to the grid structure of an unstressed syllable with a high tone. This indicates that a grid-only framework cannot account for these facts.[3]

I now turn to the second type of evidence for foot structure: the proper characterization of domains.

3.2 *The characterization of domains*

A stressed syllable and a neighboring unstressed syllable often form a domain within which a phenomenon applies. A grid-only approach is notoriously bad at defining domains of this type, as it can only define a primary stressed position, secondary stressed positions, and unstressed positions. It is impossible to express the fact that some unit creates a bond with another unit.[4]

An example showing this is provided by Guugu Yimidhirr, a language spoken in Australia (Zoll 2004; Elías-Ulloa 2006). In this language long vowels can only occur in the first and/or the second syllable of a word, but nowhere else. Some illustrative examples, from Elías-Ulloa (2006: 231–232), are given in (18):

(18)	a.	ˈguːgu	'language-ASS'
		ˈbuːraˌjaj	'water-LOC'
		ˈdaːbaˌŋahŋaˌla	'ask-RED-IMP'
	b.	maˈgiːl	'branch-ABS'
		maˈjiːŋu	'food-PURP'
		maˈgiːlnajˌgu	'branch-PL-EMPH'
	c.	ˈbuːraːj	'water-ABS'
		ˈdjiːraːlˌgal	'wife-ADESS'
		ˈbuːraːjˌbigu	'water-LOC-EMPH'

In a theory that recognizes feet, it is easy to characterize the domain within which the long vowels can occur; it is the initial foot, which is also the head of the word. This foot can either be a trochee or an iamb, depending on the presence and location of a long vowel. On the other hand, in a grid-only account it is difficult to understand why long vowels are restricted to the first two syllables of the word. This is a consequence of the fact that in this approach the first two syllables cannot be characterized as a domain. The facts illustrated in (18) are therefore

[3] Other phenomena showing that not all syllables behave in the same way in a lapse are high vowel deletion in Old English (Dresher and Lahiri 1991) and vowel balance effects in Scandinavian (Bye 1996) and Old Frisian (Smith 2004; Smith and van Leyden 2007). Unfortunately, due to lack of space I cannot discuss these phenomena here.

[4] The domain-defining character of the foot is the oldest evidence in favor of its existence (Selkirk 1980). It turns out to be very difficult to find cases where it is absolutely impossible to define a domain with grids only. An interesting example is provided by Pearce (2006), who argues that feet in Kera create the domains for tone association. I suspect, however, that an alternative is possible with grid marks only, such that (certain) tones tend to anchor to strong positions on the grid. The role of the foot as a domain delineator is systematically eliminated in Majors (1998). She argues that feet do not play a role in stress-dependent harmony.

problematic for a grid-only framework. Let us now turn to the third type of evidence: syllable weight.

3.3 *Evidence from syllable weight*

Prince (1983) shows that the grid-only framework can explain the relationship between syllable weight and the position of stress. This seems to suggest that the effect which weight has on the position of stress can be described without making use of foot structure. However, once we broaden our view beyond the distribution of stress in the strict sense, it becomes obvious that weight does provide us with convincing evidence for foot structure. Here I present one case: allomorph selection in Shipibo, spoken in Brazil and Peru.[5]

Shipibo has two allomorphs meaning 'again': /ribiː/ and /riːba/. The allomorph /ribiː/ has a long vowel in the second syllable, /riːba/ in the first syllable. Elías-Ulloa (2006) shows that allomorph selection is determined by foot structure. The language has two different feet: moraic trochees and iambs. The former is preferred; iambs are only built if the construction of moraic trochees is not possible. Furthermore, both moraic trochees and iambs must be bisyllabic. Two other high-ranking constraints are relevant here. The constraint Weight-to-Stress is inviolable, so a heavy syllable cannot be left unparsed. There is also one Alignment constraint that is high-ranked: the left edge of a word must be aligned with the left edge of a foot. Consider the following forms (from Elías-Ulloa 2006: 7), where the relevant allomorphs are underlined:

(19) *Allomorph selection in Shipibo*

a.	(pi-'riː)(ba-kɨ)	'eat-again-PAST'
b.	('puta)(riˌbiː)-kɨ	'throw-again-PAST'
c.	('puta)(-ma-ˌriː)(ba-kɨ)	'throw-CAUS-again-PAST'
d.	('puta)(-ˌjama)(-riˌbiː)-kɨ	'throw-NEG-again-PAST'

In (19a), /riːba/ is selected, allowing all syllables of the word to be parsed into feet. If the other allomorph had been selected, the form /pi-ribiː-kɨ/ would have been created. This form cannot be correctly parsed into foot structure. One possible parse is ('pi-ri)(ˌbiː-kɨ), but this representation contains an uneven trochee, which is not allowed in Shipibo. Another realization might be pi-(ri'biː)-ki. But this structure violates the Alignment constraint, which is not possible either.

In (19b), /ribiː/ is selected; this is again explained by foot structure. If the other allomorph had been selected, a form would have been created that could not be properly parsed, viz. /puta-riːba-kɨ/. The parse ('puta)(-ˌriːba)-kɨ is unacceptable, because it contains an uneven trochee. The alternative ('puta)(-ˌriː)(ˌba-kɨ) is also bad, because it contains a monosyllabic iamb, a type of foot that is non-existent in this language. Yet another alternative would be pu(ta-'riː)(ˌba-kɨ), but this representation violates the requirement that a word should begin with the left edge of a foot. This constraint is very highly ranked in Shipibo. The form that is actually realized, (19b), does not suffer from any of these problems. It is parsed

5 Other cases are reduplication in Kosraean (Kennedy 2005) and tonal spread in Capanahua (Hagberg 1993).

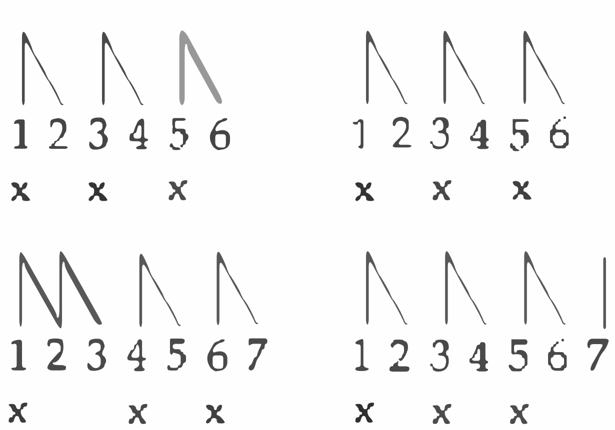
(28) *Excluded by stipulation*

The implication of this stipulation is that if PRWD-L, which requires a trochee at the left edge, is high-ranked, then the other feet cannot be iambic, even though iambs better satisfy right-alignment. The mirror image is also true, of course; if PRWD-R, which requires an iamb at the right edge, is high-ranked, then the other feet cannot be trochaic, even though trochees better satisfy left-alignment.

Maranungku has the ranking PRWD-L >> HDS-R, which derives the representation on the left in (27). This is an instance of maximization, because, in an odd-parity word, as many gridmarks are present as there are feet. Suruwaha has the ranking PRWD-R >> HDS-L, deriving the pattern on the right in (27). Again, in the odd-parity words the number of stresses equals the number of feet. Maranungku has traditionally been analyzed as a language with trochees built from left to right, with a degenerate foot at the right. Suruwaha has been described as its mirror image, with iambs built from right to left and a degenerate foot at the left (cf. (4)).

In principle, every foot receives a gridmark, although, due to overlapping feet, this does not necessarily mean that there are as many gridmarks as there are feet. At the right edge, NON-FINALITY might exclude the presence of a gridmark over the last syllable. Furthermore, at the left edge, the first syllable can be subject to a constraint called INITIALGRIDMARK (Hyde 2002: 320), which requires the presence of a gridmark on the first syllable of a word. If these constraints are high-ranked, this can lead to a situation where a foot is not accompanied by a gridmark. Pintupi has high-ranking NON-FINALITY, while Garawa has high-ranking INITIALGRIDMARK, as illustrated in (29).

(29) *Garawa* *Pintupi*



Garawa is like Nengone (cf. (25)) in the sense that feet are attracted to the left by high-ranking HDS-L. There are also two differences. In Garawa, INITIALGRIDMARK is high-ranked, so the initial syllable must have stress. The constraint against clashing gridmarks is also high-ranked, excluding an immediately following gridmark in the domain of the second foot. With these two constraints dominating GRIDMARKMAPPING, the first foot of the two overlapping feet is stressed, whereas the second foot is not. We thus get a stressless trochee. This is a foot with a head, but without a gridmark. The absence of a gridmark in the second foot creates a lapse immediately after the initial stress.

Pintupi is similar to Maranungku (cf. (27)), but has high-ranking Non-finality. If the final syllable is a foot-head, then it may not receive a gridmark. This creates a lapse at the right edge.

With the system described here it is very easy to explain all the asymmetries mentioned in §2, which are listed again in (30).

(30) *Anti-Garawa* *Anti-Pintupi* *Anti-Piro*

```
  x    x    x          x    x    x          x    x    x
(1 2)(3 4)(5 6)      (1 2)(3 4)(5 6)      (1 2)(3 4)(5 6)
  x    x       x        x    x    x         x       x    x
(1 2)(3 4) 5 (6 7)   1 (2 3)(4 5)(6 7)    (1 2) 3 (4 5)(6 7)
```

In Anti-Garawa, one iamb is built at the right edge, and the other iambs are built from left to right, so that, in odd-parity words, a lapse is created before the final stress. In Hyde's theory it is impossible to derive such a pattern. There are two ways to stress the final syllable of a word: either Hds-R or PrWd-R is high-ranked. With these two systems, it is impossible to create a lapse before the final stress. With high-ranking Hds-R, we derive Araucanian, as already shown in (25). This is a language with a minimization pattern, where in odd-parity words stresses are economically placed, so as to avoid a lapse. With high-ranking PrWd-R we get a system like Suruwaha, whose basic structure is given in (27). This is a maximization pattern, where an extra stress is created so as to avoid a lapse. Anti-Garawa, then, is ruled out, because it cannot be derived.

There is also no place for Anti-Pintupi. The particular property of this imaginary system is that there is no stress on the *initial* syllable. There is only one way to block a stress on a peripheral syllable, Non-finality, which, however, can only block stress on a final syllable. Since there is no equivalent Non-initiality preventing stress from the initial syllable, Anti-Pintupi cannot be derived.

Anti-Piro is also impossible. In this imaginary system, one iamb is built at the left edge, creating a fixed stress on the peninitial syllable. To place a fixed stress at the left periphery two constraints are available. Either Hds-L or PrWd-L must be high-ranked. In the former case we derive Nengone, whose basic configurations are shown in (25). Nengone is a minimization pattern, where the minimal number of stresses is economically placed so as to avoid a lapse. With high-ranking PrWd-L Maranungku is derived, as shown in (27). Maranungku displays a maximization pattern, where an additional stress is inserted so as to avoid a lapse. Thus there is no way to generate a system like Anti-Piro.

Hyde's system is similar to Gordon's in one sense. Neither uses the classical alignment constraints that refer to *feet*. In particular, AllFeet-L/R and PrWd-L/R (McCarthy and Prince 1993) are eliminated.

(31) a. AllFeet-L/R
The left/right edge of every foot is aligned with the left/right edge of some prosodic word.

b. PrWd-L/R
The left/right edge of every prosodic word is aligned with the left/right edge of some foot.

In these constraints, Hyde replaces the argument *foot* by *foot-head*, as shown in (21) and (26), whereas Gordon replaces it by *gridmark*. These two changes are almost identical, since a foot-head normally has exactly the same distribution as a gridmark.

The crucial difference, of course, is the notion of overlapping feet. We might say that, where Gordon eliminates foot structure entirely, Hyde introduces a new type of structure. Of course, Hyde is aware of the fact that this concept must be motivated on independent grounds, a task which he undertakes in Hyde (2008).

Optimality Theory struggles with what Hyde calls the "odd-parity problem." This problem can be divided into two sub-problems: the "even-only problem" and the "odd-heavy problem." The introduction of overlapping feet provides a solution which is not available in standard approaches. For reasons of space I will only discuss the first instance of the odd-parity problem.

The even-only problem is caused by the fact that, in odd-parity words, Faithfulness constraints are in conflict with two other constraints: FootBinarity (the constraint that penalizes degenerate feet) and Parse-σ (the constraint requiring that syllables be dominated by feet). Suppose that the Faithfulness constraints are ranked below FootBinarity and Parse-σ. Under this ranking it is better to insert or delete a syllable than to create a violation of either FootBinarity or Parse-σ, as shown in the following tableau:

(32)

1 2 3 4 5 6 7	Parse-σ	FootBinarity	Max	Dep
a. (1 2)(3 4)(5 6) 7	*!			
b. (1 2)(3 4)(5 6)(7)		*!		
☞ c. (1 2)(3 4)(5 6)(7 8)				*
☞ d. (1 2)(3 4)(5 6)			*	

This tableau shows what happens to a word with an uneven number of syllables. The first candidate, (32a), contains an unparsed syllable, violating Parse-σ. (32b) has a final, monosyllabic foot, violating FootBinarity. In (32c), a vowel is inserted. In this way an extra syllable is created, so that at the right edge a binary foot is built. FootBinarity and Parse-σ are therefore satisfied, although Dep is violated. Finally, in (32d) a vowel is removed, so that both Parse-σ and FootBinarity are again satisfied, although Max is violated.

We can see, then, that in an odd-parity word Parse-σ and FootBinarity can be satisfied if a syllable is inserted or deleted. From this it follows that there should be languages in which all words contain an even number of syllables. These languages would have the ranking Parse-σ, FootBinarity >>Max, Dep. However, no language like this has ever been attested. This illustrates the phenomenon referred to as the even-only problem.

Hyde (2008) shows that the even-only problem does not arise in a theory with overlapping feet. In such a theory, it is possible to satisfy FootBinarity and Parse-σ without violating Faithfulness. In other words, with overlapping feet, no conflict arises between FootBinarity and Parse-σ on the one hand and the two Faithfulness constraints on the other. This is shown in (33):

Nengone has overlapping trochees, with a stressed syllable in the middle. Araucanian has overlapping iambs, also with stress in the middle (cf. (25)). Finally, Garawa has overlapping trochees, with a stress in initial position, as shown in (29).[6]

These representations indicate that headedness is not expressed by gridmarks. In Hyde's framework a foot can be an iamb, even though it does not have a gridmark in its final syllable, as in Araucanian; similarly, a foot can be a trochee, even without a gridmark in its initial syllable, as in Nengone and Garawa.

This is a unique aspect of Hyde's theory. Most recent theories claim that headedness is expressed on the grid. Some theories claim that feet *must have heads*, so that for each foot there must be a gridmark accompanying it. Theories of this type invoke a principle like the Faithfulness Condition, originally formulated in Halle and Vergnaud (1987: 15–16).[7]

Meanwhile, another theory on the relation between foot structure and headedness has been developed. In this theory, headedness is disconnected from foot structure. In these theories there are two foot types; headed feet, which are accompanied by a gridmark, and headless feet, which are not accompanied by a gridmark. The loose connection between headedness and foot structure is expressed by the Separability Hypothesis (Crowhurst and Hewitt 1995b: 39), which states that feet can be headed or headless. Normally, a foot does have a head, but under the pressure of certain constraints, it can happen that a foot is not able to acquire a head. Proponents of the Separability Hypothesis are Hagberg (1993), Crowhurst and Hewitt (1995a, 1995b), Bye (1996), and Crowhurst (1996). Bye occupies a special position among them, because he assumes that feet can sometimes have two heads; when this happens there are two gridmarks in a single foot.[8]

It is clear that Hyde's theory differs from both views just mentioned. On the one hand, it (implicitly) uses the Faithfulness Condition, because all feet are inherently headed. (Heads are not expressed on the grid, but by line structure, as we have already seen; a head has a vertical line, whereas a dependent has a slanted line.) Yet the theory also recognizes the Separability Hypothesis, in the sense that there is a separate mode of representation, the grid, where a gridmark may or may not accompany a foot. Hyde's theory, therefore, does not recognize headless feet, but it does recognize stressed (headed) feet and unstressed (headed) feet. In this sense, Hyde's theory is certainly representationally richer than all other recent theories.

There is some evidence from Vogul, a language of Siberia, that feet can be headed, even if they are not stressed. In Vogul, stress does not seem to be quantity-sensitive (Vaysman 2009). The main stress is realized on the initial syllable, and there are secondary stresses on every other syllable thereafter. If, however, the final syllable is a target for secondary stress, this stress is not realized. Syllable quantity is irrelevant for this distribution. Only syllables with long vowels are heavy: words with a heavy syllable have exactly the same stress patterns as words with only light syllables. These regularities are illustrated in (35).

[6] The mirror image of Garawa (overlapping iambs with stress on the final syllable) does not exist. This system could only arise with a constraint requiring stress on the final syllable, together with high-ranking *Clash. However, in Hyde's system there is no constraint requiring stress on the final syllable. Therefore, two overlapping iambs at the right edge will always have stress in the middle (the minimization pattern).

[7] Halle and Vergnaud's Faithfulness Condition should not be confused with the Faithfulness constraints of OT. The former is a condition on the relation between a foot and its head.

[8] Level stress in Scandinavian dialects necessitates this representation, according to Bye (1996).

words. Overlapping feet can be motivated on independent grounds: overlapping feet solve the odd-parity problem.

In Hyde's theory, the grid is mostly subservient to foot structure. Feet are always inherently headed, with or without a gridmark.

REFERENCES

Altshuler, Daniel. 2009. Quantity-insensitive iambs in Osage. *International Journal of American Linguistics* 75. 365–398.

Bye, Patrik. 1996. Scandinavian "level stress" and the theory of prosodic overlay. *Nordlyd* 24. 23–62.

Crowhurst, Megan J. 1996. An optimal alternative to conflation. *Phonology* 13. 409–424.

Crowhurst, Megan J. & Mark Hewitt. 1995a. Directional footing, degeneracy, and alignment. *Papers from the Annual Meeting of the North East Linguistic Society* 25. 47–61.

Crowhurst, Megan J. & Mark Hewitt. 1995b. Prosodic overlay and headless feet in Yidinʸ. *Phonology* 12. 39–84.

de Lacy, Paul. 2002. The interaction of tone and stress in Optimality Theory. *Phonology* 19. 1–32.

de Lacy, Paul. 2007a. The interaction of tone, sonority, and prosodic structure. In de Lacy (2007b), 281–307.

de Lacy, Paul (ed.) 2007b. *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press.

Dresher, B. Elan & Aditi Lahiri 1991. The Germanic foot: Metrical coherence in Old English. *Linguistic Inquiry* 22. 251–286.

Elías-Ulloa, José A. 2006. Theoretical aspects of Panoan metrical phonology: Disyllabic footing and contextual syllable weight. Ph.D. dissertation, Rutgers University (ROA-804).

Gordon, Matthew. 2002. A factorial typology of quantity-insensitive stress. *Natural Language and Linguistic Theory* 20. 491–552.

Gordon, Matthew. 2003. The phonology of pitch accents in Chickasaw. *Phonology* 20. 173–218.

Hagberg, Lawrence R. 1993. An autosegmental theory of stress. Ph.D. dissertation, University of Arizona.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress*. Cambridge, MA: MIT Press.

Hammond, Michael. 1984. Constraining metrical theory: A modular theory of rhythm and destressing. Ph.D. dissertation, University of California, Los Angeles.

Hayes, Bruce. 1980. A metrical theory of stress rules. Ph.D. dissertation, MIT.

Hayes, Bruce. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15. 33–74.

Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 11. 429–446.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hulst, Harry van der. 1984. *Syllable structure and stress in Dutch*. Dordrecht: Foris.

Hyde, Brett. 2001. Metrical and prosodic structure in Optimality Theory. Ph.D. dissertation, Rutgers University (ROA-476).

Hyde, Brett. 2002. A restrictive theory of metrical stress. *Phonology* 19. 313–359.

Hyde, Brett. 2006. Towards a uniform account of prominence-sensitive stress. In Eric Baković, Junko Itô & John McCarthy (eds.) *Wondering at the natural fecundity of things: Essays in honor of Alan Prince*, 139–183. Santa Cruz: Linguistics Research Center, University of California, Santa Cruz. Available at http://repositories.cdlib.org/lrc/prince/8.

Hyde, Brett. 2008. The odd-parity parsing problem. Unpublished ms., Washington University (ROA-971).

Itô, Junko & Armin Mester. 1992. Weak layering and word binarity. Unpublished ms., University of California, Santa Cruz.

Kager, René. 1993. Alternatives to the iambic-trochaic law. *Natural Language and Linguistic Theory* 11. 381–432.

Kager, René. 1995. The metrical theory of word stress. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 367–402. Cambridge, MA & Oxford: Blackwell.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kager, René. 2007. Feet and metrical stress. In de Lacy (2007b), 195–227.

Karvonen, Daniel. 2005. Word prosody in Finnish. Ph.D. dissertation, University of California, Santa Cruz.

Karvonen, Daniel. 2008. Explaining Nonfinality: Evidence from Finnish. *Proceedings of the West Coast Conference on Formal Linguistics* 26. 306–314.

Kennedy, Robert. 2005. The binarity effect in Kosraean reduplication. *Phonology* 22. 145–168.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.

Majors, Tivoli J. 1998. Stress-dependent harmony: Phonetic origins and phonological analysis. Ph.D. dissertation, University of Texas, Austin.

McCarthy, John J. & Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology 1993*. 79–153.

Pearce, Mary. 2006. The interaction between metrical structure and tone in Kera. *Phonology* 23. 259–286.

Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14. 19–100.

Prince, Alan. 1985. Improving tree theory. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 11. 471–490.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Selkirk, Elisabeth. 1980. Prosodic domains in phonology: Sanskrit revisited. In Mark Aronoff & Mary-Louise Kean (eds.) *Juncture*, 107–129. Saratoga: Anma Libri.

Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Smith, Laura C. 2004. Cross-level interactions in West Germanic phonology and morphology. Ph.D. dissertation, University of Wisconsin, Madison.

Smith, Neil V. & Klaske van Leyden. 2007. The unusual outcome of a level-stress situation: The case of Wursten Frisian. *North-Western European Language Evolution* 52. 31–66.

Vaysman, Olga. 2009. Segmental alternations and metrical theory. Ph.D. dissertation, MIT (ROA-1011).

Zoll, Cheryl. 2004. Positional asymmetries and licensing. In John J. McCarthy (ed.) *Optimality Theory in phonology: A reader*, 365–378. Cambridge, MA & Oxford: Blackwell.

42 Pitch Accent Systems

Harry van der Hulst

1 Introduction

This chapter deals with the typology of *word prosodic systems* and, specifically, discusses the notion of "pitch accent (language)," asking whether there is such a class of pitch accent languages distinct from "stress languages" and "tone languages." Several issues will turn out to be crucial. Firstly, there is the issue of recognizing (or not) a notion of *accent* which could be said to underlie both pitch accent and "stress" (or indeed stress accent), and perhaps other phenomena which are not frequently referred to as accentual (such as phonotactic asymmetries). Secondly, there is the question as to whether we wish to distinguish between pitch as a *non-distinctive* and thus perhaps strictly phonetic property (found in phonetic implementation) and pitch as the exponent of a phonological category (namely *tone*). Thirdly, there is the possibility of having tone, stress, and accent (in various combinations) "side by side" within the same language, which raises the question of how these notions interact in any given language.

The structure of this chapter is as follows. In §2 I will introduce the basic notions and definitions. §3 will briefly discuss examples of languages that have been referred to as pitch accent languages, where accent is apparently realized in terms of *non-distinctive* pitch. In §4, I examine cases in which tone realization or tone distribution has been said to depend on accent (or stress), a class of languages that is also often included in the pitch accent type. §5 and §6 focus on the different ways in which alleged pitch accent languages have been analyzed, with or without using the notion "accent." In §7, I define the notions accent and stress as distinct phonological entities and suggest that stress languages may or may not be accentual. In §8 I offer some conclusions.

2 Accent, tone, and stress: Definitions and usage

2.1 *Accent and stress*

For many languages, researchers have reported word-level "prominence," associated with a specific syllable in the word, which is called "stress" (an English term) or "accent" (the term used in, for example, French or German) (see also CHAPTER 41: THE REPRESENTATION OF WORD STRESS). In literature in English on the subject, both "stress" and "accent" have been used for word-level prominence, which has lead to a good deal of confusion, in particular because there are writers who use the terms for different things. Cutler (1984), for example, regards "stress" as a property of words and "accent" as a property of sentences. There is thus a need for clarity on how these two terms are used.

2.2 *Accent and its cues*

On closer scrutiny, the informal notion of "prominence" can be divided into two distinct phenomena. On the one hand, we have the *location* of the prominent syllable (e.g. penultimate; final if the final syllable has a long vowel, otherwise penultimate; etc.) and on the other hand, there are phonetic (and phonotactic) *cues* that signal the location of the prominent syllable (CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE). In one (fairly old) terminological tradition, the locational aspect of prominence is referred to as *accent*. The characterization of the accent (location) is essentially sequential (or *syntagmatic*); only one syllable within the relevant domain can have this property. This is what Martinet (1960) and Garde (1968) refer to as the *contrastive* or *culminative* function of "accent," a term mainly used by Trubetzkoy (1939). The realizational aspect of prominence is, in a sense, paradigmatic (cf. van Coetsem 1996): there are various (not necessarily incompatible) phonetic and phonotactic means for cueing the accent. Some languages may favor one specific cue (e.g. pitch or duration), but several cues may conspire to manifest the accent. This division of "prominence" correlates with traditional terminological systems such as *musical accent vs. dynamic accent* or (with much the same meaning) *pitch accent* (systems) *vs. stress accent* (systems) (see Fox 2000: ch. 3 for an excellent general review of the notion *accent*; also van Coetsem 1996; van der Hulst 1999b, 2010b). In each case, the modifier of the head noun ("accent") says something about the way in which the accent is "manifested" or "realized." In this chapter I will focus on relationships that involve accent and pitch, whether used distinctively (in terms of contrastive tones) or non-distinctively. However, I will also consider the relationship between accent and stress.

2.3 *Word prosodic types*

While in some languages pitch is a property of words, all languages use pitch features within an *intonational system*, a system that aligns "sentences" with a melody that can be defined in terms of pitch events that mark boundaries of (syntactic or prosodic) units as well as the informational packaging of the utterance with reference to the notion "focus" (Bolinger 1982; Gussenhoven 2004; CHAPTER 32: THE REPRESENTATION OF INTONATION). At the same time there are languages in which pitch is a property of "words." Within this group of languages we commonly find the labels in (1b) and (1c). The label "stress" in (1a) is reserved for languages that need no specification of pitch at the word level, although, like all other languages, they use pitch for intonation purposes.

(1) a. Stress (or stress accent)
    b. Pitch accent
    c. Tone

There is, however, a great deal of controversy concerning the use of the terms tone and pitch accent, and, for that matter, the term stress.[1]

Hyman (2001, 2006, 2009) makes a case for treating systems that we label stress and tone as "prototypes," meaning that languages that belong to one or the other (or both) type(s) display one or more specific defining properties.[2] "Pitch accent," according to Hyman, is not a prototype, but rather a label for a large class of hybrid systems that mix "tone" and "stress" properties in various ways, or systems that are clearly tonal, although displaying various restrictions on the distribution of tones. In effect, Hyman regards the notion accent as unnecessary, whether as a formal mechanism in analysis or as a prosodic type. Other researchers (such as Gussenhoven, e.g. 2004) who also reject the idea of "pitch accent languages" nonetheless recognize the notion of accent as an analytic device. In this chapter these views will be discussed and compared to views that attribute a fundamental role to the notion accent.

2.4 *Definitions and use of tone*

A traditional way of defining the notion tone is in terms of "*distinctive* use of pitch." Thus, if a language uses pitch to distinguish different otherwise identical morphemes, pitch has a phonological or contrastive (distinctive) status. The following frequently quoted definition captures what is perhaps the canonical use of distinctive pitch:

> A tone language may be defined as a language having lexically significant, contrastive, but relative pitch on each syllable (Pike 1948: 3).

If tones are distinctive on all syllables (possibly like properties such as frontness, height, or roundness) we can say that the distribution of tones is *unrestricted*. Most researchers, however, agree that there is no reason to limit the term tonal language to cases in which the distribution of tones is *entirely* unrestricted (see CHAPTER 45: THE REPRESENTATION OF TONE). Presumably, all tonal systems show restrictions resulting from tonal spreading or assimilation (Schuh 1977; Hyman 2007), from using a limited set of tonal melodies which are properties of morphemes rather than of syllables (Leben 1971; Goldsmith 1976b; Halle and Vergnaud 1982), from the avoidance of sequences of identical tones (dissimilatory or OCP effects), or indeed from relations between tone distribution and accent (or "stress") (see §4). Also, it is not uncommon to find that the full range of contrasts is not found in affixes (as opposed to roots or stems) (CHAPTER 104: ROOT–AFFIX ASYMMETRIES). Finally, initial or final syllables may fail to bear tonal contrast (sometimes to leave room for intonational tones or for other, perhaps "perceptual," reasons; CHAPTER 98: SPEECH PERCEPTION AND PHONOLOGY).[3] Since it would be unwise

[1] Typological studies of word prosodic systems are numerous: e.g. Trubetzkoy (1939); Hockett (1955); Garde (1968); Meeussen (1972); Goldsmith (1976a, 1988); Greenberg and Kaschube (1976); Hyman (1977, 1978, 1981, 2006, 2009); Lockwood (1982); Clements and Goldsmith (1984); Beckman (1986); Clark (1987, 1988); Haraguchi (1988); Hollenbach (1988); van der Hulst and Smith (1988); Mock (1988); Wright (1988); Hayes (1995); van der Hulst (1999a, 2010c); de Lacy (2002); Duanmu (2004).

[2] Here Hyman avoids the term "stress accent," presumably because he no longer (cf. Hyman 1977) recognizes the label "pitch accent" as a useful one and thus essentially wants to eliminate the notion accent altogether.

[3] Suárez (1983: 52) observes that in Huichol and Mazahua there is no tone contrast on the last two syllables or the last syllable, respectively. In these languages, inherent lexical tones are removed to free up space for intonational tones.

to maintain the strictness of Pike's definition (according to which perhaps there is not a single tonal language), van der Hulst and Smith (1988) quote the much more liberal definition of Welmers (1973: 2):

> A tone language is a language in which both pitch phonemes and segmental phonemes enter into the composition of at least some morphemes.[4]

Note the use of the term *pitch phoneme* (CHAPTER 11: THE PHONEME), which suggests that Welmers requires that pitch is used contrastively, a crucial point to which I return below. This definition includes languages in which there are tonal contrasts in certain positions, or even in only one position, in some morphemes.

With this broader definition, tonal languages can be ranked on a scale of *tonal density* (Gussenhoven 2004), which indicates how many word positions have how much tonal contrast. In a sense, such a scale indicates the relative *functional load* of tone properties. Stretching Gussenhoven's notion, we could say that relative density arises not only in the syntagmatic dimension (depending on how many positions display tonal restrictions), but also in the paradigmatic dimensions (depending on the number of contrastive options per position):

(2) *Tonal density matrix*

T1	+	+	+	+	+	+	
T2	+	+	+	+	+	+	
T3	+	+	+	+	+	+	
	x	x	x	x	x	x	(tone bearing-units)

Each potential minus would indicate a restriction on the distribution of a distinctive tone. However, no matter how dramatic the restrictions, as long as there is tonal contrast (i.e. distinctive use of pitch), phonological tones *must* be specified in the lexical entries. The smallest tonal system would have two tones, H and L. More extensive systems would add an M tone, or possibly two different M tones (high mid and low mid). In addition, systems can have contour tones (rise, fall, etc.) (CHAPTER 45: THE REPRESENTATION OF TONE).

2.5 Culminativity and obligatoriness

Another frequently cited term in this context is "restricted tone language," introduced by Schadeberg (1973) and Voorhoeve (1973). This term, too, would seem to indicate a scale of restrictiveness, although Voorhoeve introduced it in the context of Bantu languages whose tonal system is so severely restricted (up to one H per word in a H/L system) that he suggested that an accentual analysis might be considered (CHAPTER 114: BANTU TONE). Indeed, adding syntagmatic and paradigmatic restrictions on the distribution of tone together, one could see that a language, despite having a H/L contrast, while allowing at most one H tone per word could easily permit an accentual analysis in which the H "tone" is regarded as the predictable pitch cue of an accent, even in a case where there is no indication of any additional independent cues for this accent.

[4] Strictly speaking this excludes cases in which a language has tonal affixes without having affixes or other morphemes that combine tone and segmental properties.

But what is "accent," and how is it formally represented? Hyman (2009) formulates two necessary properties of what he calls stress, which I will take as a point of departure for establishing what might be seen as characteristics of accent, if these notions are going to be distinguished. One "property" is that each "word" can have stress or accent *at most once* (only one syllable can be stressed or accented) and, additionally, each word must have it *at least once*. These two properties, following Hyman (2006, 2009) can be referred to as:

(3) a. Culminativity (at most one)
 b. Obligatoriness (at least one)

Let us now ask whether the two properties in (3) *must* be regarded as necessary properties of accent. An issue that goes to the heart of what is often seen as problematic for the notion "pitch accent" is that languages which allegedly have a pitch accent system, and thus accent, sometimes have (lexical) words that appear to be unaccented (see the discussion of Tokyo Japanese in §6). This, however, is only problematic if obligatoriness is stipulated to be a *necessary* property of accent. We could investigate a more liberal interpretation of accent, in which unaccented words are permitted in an accentual language. This, of course, has important consequences, because it opens the door to using the presence *vs.* absence of accent as a contrastive option and thus to analyzing alleged tonal languages that have a H–L contrast as fully accentual languages, with H as the exponent of accent and L as the lack of accent.

We might then also question whether culminativity is a necessary requirement for describing accent. If culminativity is not required, even "H/L" languages that allow multiple H "tones" *could* be analyzed as fully accentual. Allowing words to have multiple accents separates the notions stress and accent even more dramatically than just giving up obligatoriness for accent. Still, if accent is not the same thing as stress, there is no a priori reason for believing that any properties of the latter need to be true of the former. I return to these issues in §5.3.

2.6 Representational issues

Answers to the question of whether or not the properties in (3) are definitional of accent have repercussions for, or are implicit in, the manner in which accents are formally represented. In one type of approach the relevant syllables are marked with an "accent mark," as is common in dictionaries and in autosegmental theory (e.g. the "star" of Goldsmith 1976a), or in terms of a segmental feature, as in the phonological theory of Chomsky and Halle (1968). In this "lexicographic" approach there is no commitment to culminativity or obligatoriness.

A different formal approach is to provide the string of syllables with a headed tree structure, as proposed in various versions of metrical theory (Liberman and Prince 1977) and dependency phonology (Anderson and Ewen 1987) (see also CHAPTER 40: THE FOOT). Metrical structures have one designated terminal unit, the head of the word, which counts as the (primary) "stress." This notation (assuming that all syllables must be grouped in one structure) implies culminativity, but not necessarily obligatoriness, because it does not follow from the notation that each word must be provided with a metrical tree.

However, rather than seeing asterisks and trees as competing mechanisms, we should entertain the idea that they are complementary, in that the former represent accents, while the latter represent stress. This point is acknowledged by Anderson and Ewen (1987), who, in addition to headed tree structures, also use asterisks to indicate what we might call "potential heads."[5] I will return to this point in §7.

2.7 *Problems with the notion "pitch accent"*

We have been considering a use of the term accent as an abstract mark of a position that can be cued by various phonetic properties, "stress" being one of them. Beckman (1986) refers to languages that are *not* stress accent languages, as "non-stress accent" languages (thus avoiding the term "pitch accent language"). This, of course, is compatible with the idea that in many non-stress languages pitch is the most salient property of accent. Van der Hulst (1999a, 2010b) points out that if we maintain the term "pitch accent language" we might also expect to find languages that can be labeled as "duration accent" languages (if duration is the only cue). On this view, pitch accent languages are languages in which accent is (mainly) cued by phonetic pitch.

There are, in fact, various factors that make the use of this term problematic. One factor is, obviously, that people may simply define the term differently. For example, as we will see in §4, tonal contrast is often limited to specific syllables in the word; cases of this sort have been analyzed by identifying a notion "accent," with association of tones being dependent on this accent. While in this case, the *presence* of tone can be said to function as a cue of accent, the cue is not a phonetic, but rather a phonological, fact (namely the phonotactic distribution of tones). The fact that the possibility of tonal contrast may signal the accent location is part of a much more general pattern, found in many languages, where accented syllables display contrastive or structural options that are exclusive to a particular syllable (see Downing 2010; van der Hulst 2010b).[6] Pursuing the terminological path that we started out on above, we might refer to such cases in which tonal contrast is limited to the accented syllable as *tone accent* (or *tonal accent*) languages, rather than pitch accent languages. It is apparently the case that accented syllables can be referred to by the phonology as well as by the phonetic implementation system. In fact, accents can be referred to by other grammatical modules as well, e.g. the intonation system. Does this mean that we can refer to English as an "intonation accent" language? Languages cannot be put in a single box as far as cues for accent are concerned.

Tonal accent systems, then, differ from pitch accent systems if we agree that in the latter pitch is not used distinctively. However, some writers (e.g. Downing 2010) use the term "pitch accent" for *any* system in which pitch properties (whether distinctive or not) enter into a relationship with accent or stress. This would include not only what is referred to here as a pitch accent or tone accent language, but also another class of languages which have both tone and accent, in which accent (or "stress") is assigned with reference to tone. Downing's use of the term pitch accent is thus much broader than the one considered above.

[5] Another formal notation (also proposed in Liberman and Prince 1977) is the metrical grid, which does not imply culminativity. See CHAPTER 41: THE REPRESENTATION OF WORD STRESS for extensive discussion.

[6] This relates to the notion of positional faithfulness; cf. Beckman (1998).

3.2.1 *Basque*

The dialects of Basque present a great diversity of word prosodic systems (see Hualde 1999). Gussenhoven (2004: ch. 9) presents an analysis of Northern Bizkaian Basque with reference to the Gernika and Lekeito dialects. Both have accented and unaccented roots, the former being in the minority. There are inflectional and derivational suffixes that are accented or pre-accenting. In Lekeito, if a word has an accent, this accent always ends up on the penultimate syllable. In Gernika, the leftmost (non-final) accent prevails; this is the more common case in Basque dialects. In Lekeito, unaccented words are grouped with an accented word to their left or right, whereas sequences of unaccented words together form a single domain. Each such domain either has an accent (if it contains an accented word) or is unaccented. Unaccented domains receive a default final accent in certain syntactic positions, namely at the end of the sentence or before the finite verb. Each accent, whether lexical or default, is associated with a HL pitch accent. The left edge of the accentual domain is marked by a LH boundary sequence; a high plateau is found between the boundary H and the H of the pitch accent. Systems of this sort seem obvious candidates for accentual analyses, which of course begs the question of whether they *must* be analyzed accentually. One argument that could be made for an accentual approach is that in the various dialects we note a variety of accent locations (ranging from lexical to rule-governed) which is very reminiscent of the distribution of stress in "stress accent languages." The second argument again involves the fact that pitch is non-distinctive in Basque dialects. Note that in Basque, unaccented words are provided with default accent, at least in some cases.

3.2.2 *Japanese*

We also find a broad array of word prosodic systems among the dialects of Japanese (cf. Haraguchi 1977). An overarching property of all systems is the relevance of pitch at the level of the "word," or, as some researchers prefer to put it, the "accentual domain." An interesting overview in the context of autosegmental theory of dialectal differences is offered by Haraguchi (1977, 1988), who divides Japanese dialects into two broad categories: pitch accent systems and unaccented systems. Cross-classifying with this dichotomy, he suggests a "universal" inventory of melodies (H, L, HL, LH, and LHL), from which a system may select at most one or two. In addition to the choice of one or more melodies, the differences among dialects depend on:

(4) a. The location of accent/H: fixed or free.[11]
b. The spreading of H: no spreading, rightward or leftward.

Thus, in Tokyo Japanese, the H tone spreads leftwards (leaving an initial mora low, possibly due to a boundary L tone that comes with the left). We will focus on the pitch aspect of Tokyo Japanese in §6. The system of Tokyo Japanese is such that the constituents of words (stems, affixes) can be accented or unaccented (or, in the case of affixes, pre-accented). When more than one accent is present in the accentual domain (which can be larger than the word and therefore needs careful definition; Gussenhoven (2004) calls it the α-domain), the first (or leftmost

[11] In §6 we will discuss the way accents are distributed in Tokyo Japanese, which is partly lexical and partly rule-based.

accent) predominates, i.e. will attract the high pitch/tone. If no accent is present, the high pitch occurs on the last (rightmost) syllable (and spreads from there). This "First/Last" pattern constitutes a system that is reminiscent of so-called *unbounded stress systems* (Hayes 1995). In fact, Haraguchi (1988) notes that three of the possible unbounded patterns occur in Japanese dialects (see also CHAPTER 120: JAPANESE PITCH ACCENT).

(5) a. *Systems with unaccented words*

First/First[12]	Kumi
First/Last	Tokyo, Osaka
Last/First	—
Last/Last	Hirosaki

b. *Systems without unaccented words*

First	Fukuoka
Last	—

Note that systems without unaccented words have no default clause.

Haraguchi (1977, 1988) also recognizes *unaccented systems*, i.e. systems in which no word is accented. He mentions Sendai, Miyakonojo, and Kagoshima. In such systems the tonal melody is associated either from left to right or from right to left in his analysis:

(6) *Systems with only unaccented words*

First
Last

For these systems, tones are associated to words in terms of association conventions that make no reference to accents, but rather the word edges. These same conventions are invoked for unaccented words in accentual languages (as in (5a)), which implies that in such systems tones are associated partly to accents and partly *directly* (i.e. without "intervening" accents).

In all dialects that use just one melody, the question can again be raised whether this "melody" is a phonological entity or entirely due to phonetic interpretation. Haraguchi (1988) does not consider this issue, but it could be argued, as before, that only dialects that have more than one word melody are truly tonal.

3.3 Bantu languages

Many Bantu languages are commonly described as having both tone and accentual properties, while a few (such as Swahili) have lost tone, and retain only "stress" (CHAPTER 114: BANTU TONE). Bantu word prosodic systems have always been of special interest in the debate regarding the appropriate analysis for languages that have both significant word-level pitch movement and indications that accent plays a role as well; see Schadeberg (1973), Voorhoeve (1973), Goldsmith (1976a, 1988), Hyman (1978, 1981, 1982, 1989), Clements and Goldsmith (1984), Odden (1988), and especially Downing (2010). The accentual analysis of Bantu languages was promoted by Goldsmith (1976b, 1988), although the approach has a long history (see the introduction in Clements and Goldsmith 1984 for a historical perspective).

[12] This can be glossed as: "Associate a tone with the first accented syllable, or, if no accent is present, with the first syllable."

tonal accent system. A question that arises in these cases is whether the accented syllable is cued merely by its attraction of tonal contrast, or, additionally, by other "stress-like" cues. I will consider this issue in §5. I consider here some examples from Suárez (1983), as well as from Yip (2002), in their surveys of Meso/Middle American languages. Isthmus Zapotec has two tones which associate to the accented syllable and from there spread rightwards. "Pre-stress" syllables are low-toned. Suárez also mentions Northern Pame and Yaitepec Chatino as languages that have a tonal contrast only in the syllable that is said to be "stressed" (in both cases the "final" syllable, presumably of the stem). This can be compared to Huautla Mazatec, where every syllable can have contrastive tone. In between, we find cases where the contrast on certain non-accented syllables is limited. In Palantla Chinantec, for example, there is no tonal contrast on post-stress syllables.

Van der Hulst and Smith (1988) cite the case of San Juan Copola Trique, which illustrates how restricted tonal distribution can arise historically (cf. Hollenbach 1988; Yip 2002). In the Otomanguean family in general, we find a continuum of reduction of tonal contrast and, interestingly, an *increase* in tonal contrasts on the accented syllable. A case where accent has only mildly influenced tonal contrast is found in Cajonos Zapotec (Nellis and Hollenbach 1980). Of the four underlying tones H, L, HL, and M, only M is disallowed in unaccented syllables. In this case, then, we do not have a tone accent system, but simply a tone *and* accent system, with accent-driven reduction.

Among the languages in which the distribution of tone is dependent on accent, there is a subclass of cases in which tonal contrast is only found on or near accented syllables, not because tones have been neutralized in other positions, but simply because a tonal contrast historically developed only in this position. In these cases, the accented syllable, in addition to being an attractor for tonal association, has clear stress-like cues. Hence languages of this kind are both stress accent and tonal accent languages, with the proviso that the tone does not always associate directly to the accented syllable, but sometimes on a syllable near it (although this also depends on the details of the analysis). Two well-known cases of this sort are some of the Scandinavian languages and Serbo-Croatian. For discussions of the Scandinavian type I refer to Bruce (1999) and Gussenhoven (2004) (see also CHAPTER 97: TONOGENESIS). For Serbo-Croatian, see e.g. Inkelas and Zec (1988).[13]

We must note that the co-occurrence of stress accent and a lexical pitch contrast enforces a tonal analysis of the latter. If the accent was not manifested in any other way than forming an anchor for lexical pitch, it could be argued that the opposition is one between accented words and unaccented words.

5 The accent debate

5.1 *Accents or no accents*

We have so far discussed two possible interactions between accent and pitch or tone:

[13] In his chapter on central Franconian tones, Gussenhoven (2004: ch. 12) discusses the emergence and representation of a tonal distinction that is very similar to the Scandinavian distinction; see also Gussenhoven and Bruce (1999) and Hermans (1994). We also find a similar contrast (due again to different historical factors) in Scottish Gaelic; see MacAulay (1992: 234–236).

(8) a. Pitch is dependent on accent (pitch accent systems; §3).
b. Tone is dependent on accent (accent-dependent reduction and distribution; §4).

The dividing line between the two types is distinctivity. If pitch is non-distinctive, i.e. if there is no tonal contrast, the system uses pitch to cue accent. But if there is tonal contrast, tones are involved.

The Bantu systems mentioned in the preceding section have been analyzed as involving accent and tone. However, the question of whether the occurrence of tone contrast on one specific syllable requires a notion of accent cannot be taken for granted, even when tonal association seems to be limited to an "accent-like" position. Consider the case in which the alleged accented syllable has no independent property apart from being the locus of tonal contrast. One could then say that there really is no accent at all, and instead assume that the tones, being specified as a property of morphemes, associate to their specific locus *directly*, without first assigning an accent that attracts the tones. In this case we would accept that accent rules and tone association rules fall under the umbrella of a general theory of *positional identification*, and that the principles for positional identification are similar, if not the same, for both accent placement and tone association.

(9) a. *Indirect (accentual) approach*
Step one: Accent goes to position X.
Step two: Tones go to accent.
b. *Direct approach*
Step one: Tones go to position X.

If the direct approach is taken, the category of tonal *accent* systems reduces to tonal systems which are then further differentiated in terms of different principles of association (LR, RL, positional). Below we will see that the direct tonal approach can also be applied in systems that have unpredictable (i.e. lexically specified) loci for accents.

To what degree should tone placement and accent placement be allowed to overlap? If, for example, a tonal contrast occurs on the final syllable if it is closed, and otherwise on the penultimate syllable, should we say that there is a quantity-sensitive accent rule and that tones are attracted to the accent, or should we make the tonal association rules quantity-sensitive? The earlier literature on systems in which tone contrast is limited to specific syllables reflects the view that the theory of accent placement should *not* be duplicated in a theory of tone placement, so that in these cases accent is usually seen as placing a role in tonal association.

On the other hand, Haraguchi (1977, 1988, 1991), as we saw in §3.2, makes a sharp distinction between tones that associate to accents and tones that associate directly to tone-bearing units at edges. The latter case involves only strict directional association in his analysis (from right to left, or from left to right). But, if peripheral tone-bearing units can be "extra-tonal," we can expand the set of cases in which tonal association can be direct. However, we do not expect direct tonal association to be dependent on syllable weight distinctions. Hence, if tones are attracted to positions that reflect weight criteria one would be inclined to associate tones to accent which are assigned in a weight-sensitive fashion.

Given the inevitable overlap between accent placement and direct tonal association, Pulleyblank (1986) launched an attack on the use of accents and suggested replacing accents by tones. This approach, discussed in the next section, has since become dominant.

5.2 *Giving up accents*

The direct tone approach was promoted by Pulleyblank (1986) mainly for various African tonal systems and by Poser (1984) for Tokyo Japanese in particular. The most important of Pulleyblank's arguments against the use of stars (cf. Blevins 1993: 238) are as follows. Firstly, using stars and tone makes the system overly rich, in that we now predict rules referring to stars and to tones, and to both at the same time. Secondly, the inherent culminative nature of stars can also be found in systems that are arguably tonal and non-accentual, i.e. the asymmetry between accent and non-accent finds a counterpart in systems in which H tone contrasts with zero (ending up as default L). Another argument that could be mentioned is that accent (if equated with stress) is a property of syllable, whereas stars sometimes need to be assigned to moras. Finally, as we have already mentioned, the existence of unaccented words in accentual systems, or indeed of words with multiple accents which all surface, can be regarded as problematic.

Pulleyblank applied the direct tone approach to a variety of cases, not only cases in which the position of the tone is predictable, but also those where the former accent location is lexically specified; it was subsequently adopted in much other work (Clark 1988; Hyman 1989). We note again that this move entailed the use of phonological features for non-distinctive, i.e. predictable properties. Even though the location of the alleged tone could be a lexical, unpredictable property, the phonetic nature of the entity (high pitch) would nonetheless be predictable.[14]

The abandonment of stars implies, firstly that the systems discussed in §4, where H tone is restricted (perhaps up to the point of being culminative), but not obligatory, are now analyzed as tonal. However, a further-reaching conclusion is that "straightforward" pitch accent systems (discussed in §3), where high pitch is both obligatory and culminative, are also analyzed as tonal, despite the fact that pitch is not distinctive. This may or may not be considered a (conceptual) problem (cf. Clements 2001, 2009). Another issue is of course that we have rules for tonal association which duplicate the theory of accent which is independently needed for non-tonal accent systems.

Abandoning accent cannot make the Scandinavian (and Serbo-Croatian) case purely tonal, since, as mentioned, in these cases we need the notion of stress (accent), independently of the tonal specifications.

5.3 *In defense of accents*

If accents are rejected for pitch accent and restricted tone languages, the term "accent" can simply be abandoned in favor of the term "stress" (for stress accent languages). Hyman (2007) adopts this position, reducing the typology of word

[14] This might suggest a "compromise" position in which "accents" are regarded as unspecified tonal "root nodes." In an approach which adopts a wider use of accents as possible ingredients of stress accent systems, this idea could not be maintained.

prosodic systems to tone languages and stress languages. In this section I will focus on the use of accents in "tonal" systems and suggest a different approach, one which maximizes the use of accents at the expense not just of non-contrastive "tones," but also of (allegedly) contrastive tones.

The issue here does not revolve around languages that have obligatory and culminative high pitch such as Nubi. Here, the case for accent could be considered uncontroversial, if we believe that culminativity and obligatoriness are necessary for an analysis using accent (which essentially means that accent and stress are the same thing). Rather, let us focus on languages in which H tones violate one or both of these two constraints. I will argue that languages of this sort can also be analyzed as accentual (and thus non-tonal), if obligatoriness and culminativity, while perhaps being typical or even necessary for stress, are not required for accent. These points were anticipated in §2.5.

Let us first consider the type of case in which one syllable per word is either H or L, meaning that H is culminative, but *not* obligatory. In an accent-cum-tone analysis we would postulate an accent, and from there we have several options, depending on how we characterize the tonal contrast (H/L, H/zero, zero/L). But there is also another option. We can also analyze the contrast as accent *vs.* no accent (with accent giving rise to phonetic high pitch and low pitch as default). This means that we can analyze these alleged H/L systems as pitch accent systems as long as we "allow" that accentual languages have a class of unaccented words.

Secondly, even when a "H/L" system allows multiple (non-adjacent) "H tones," this does not necessarily enforce a tonal analysis. If neither criterion proposed by Hyman (2007) for stress applies to accent, there is no reason why a word could not have more than one accent.

Concluding, if we push the use of accents to its limits (at the expense of using tones), this implies allowing unaccented words (violating obligatoriness) and multiple accents (violating culminativity). In this liberal view on accent, only languages that have more than a binary pitch contrast are *necessarily* tonal; in addition, we find languages in which culminativity and obligatoriness of accent is independently required (as in the case discussed in §4).

One could say that "H/L" systems are the real pivotal cases, where, as linguists (or as language learners), we have a choice between an accentual and a tonal analysis. There may be certain diagnostics that will tip the balance to either an accentual or a tonal analysis, and these need to be made explicit. More work is called for in this area.

An accentual approach is favored when the distribution of accent falls squarely within a theory of accent placement that is independently needed for stress accent and other types of accentual languages. This, perhaps, makes an accentual analysis of those languages in which the alleged accents need to be assigned to moras undesirable (cf. the case of Somali; Biber 1981; Hyman 1981; Banti 1988). Another diagnostic pointing to tones is the need to refer to *floating tones*, on the assumption that the notion "floating accent" is suspect. Thirdly, it could be argued that tonal spreading processes might suggest tone, but implementational mechanisms can also be held responsible for pitch extending over several syllables. A fourth potential way to discriminate between accent and /H/ tone would be to look at the details of phonetic implementation. One could conceivably argue that the phonetic pitch target of phonological categories like /H/ is more specifically defined than the pitch target of accents. Lastly, an accentual analysis could account for cases in which we need rules that delete apparent accents in clash, or other rules that refer to accents, irrespective of their pitch or "tonal" correlates.

McCawley (1978) suggested that in some cases one might want to say that a system is accentual first, and then becomes tonal in the course of the derivation. The question is, however, whether the tonal end of the derivation is still part of the phonology or part of the phonetic implementation.

In this section I have suggested that accentual systems should be "allowed" to have unaccented words or multiple accented words, or even both. This seems to imply that obligatoriness and culminativity are not necessary properties of accent and that the case in which accents are both obligatory and culminative is just one of four possibilities; see §7.

6 The case of Tokyo Japanese

Tokyo Japanese is a language that is often mentioned as a prime example of a pitch accent system, but differs from both Nubi and Somali, while apparently sharing properties with each. Every word is said to have high pitch, but, at the same time, some words are accented and others are non-accented. Let us first consider the basic facts; references to various types of analyses are offered below.

In Tokyo Japanese, nouns have a specific pitch contour, which in some but not all cases involves LHL. In those words that have the full LHL pattern, the L occurs on the initial mora. This mora is followed by a high plateau, which may drop to low at some point. After the drop, remaining syllables are low. In some words the initial L is missing, and in other words the final L. Thus we find the following patterns, taking trisyllabic nouns to illustrate the possibilities:

(10)	a.	HLL	b.	LHL	c.	LHHL	d.	LHHH
		σσσ		σσσ		σσσ(-σ)		σσσ(-σ)
		inoti		kokoro		atama		sakana
		'life'		'heart'		'head'		'fish'

This system can be and has been analyzed in many different ways; here we will specifically focus on accounting for the difference between (10c) and (10d). For (10a)–(10c) we have three options; depending on which one is chosen, various approaches can be suggested for class (10d):

(11)	(10a)–(10c)	(10d)		
		(i)	(ii)	(iii)
a.	Accent → /H/ → [H]	default accent	/H/ to last σ	implementation
b.	/H/		/H/ to last σ	implementation
c.	Accent → [H]			implementation

In (11a), the accent-cum-tone analysis, the (10d) case would be lexically unaccented. Since such words surface with an apparent H tone throughout (except for the initial mora), one could consider assigning a default final accent (11.i), which then triggers an H tone. This analysis encounters a problem, however. Words that have no lexical accent must be identifiable as such in the phonetic interpretation because there is a phonetic difference between (10c) and (10d). Roughly speaking, (10c) is LHH and (10d) is LHM, with the stem-final "H" in the latter not quite as high as the other Hs in both examples. The two types of words also have different effects on following words (or "accentual phrases") inside the Intermediate Phrase: (10c) causes downstep, (10d) does not; cf. Haraguchi (1988); Pierrehumbert and Beckman (1988); Gussenhoven (2004). Alternative (ii), which would use the H tone assignment rule in (12), resolves this issue, because it can be argued that a H tone on an accented syllable and a H tone on an unaccented syllable are interpreted differently (cf. (5a)):

(12) Assign /H/ to the first accent or, if there is no accent, to the final syllable.[15]

The difference between (10c) and (10d) could also be described if the pitch properties of the latter class are entirely accounted for in terms of phonetic implementation (11.iii), since this system could respond to the presence *vs.* absence of an accent.

In the second (tone-only) approach, (10d) should be accounted for by method (iii), since method (ii), available in principle, would wrongly conflate (10c) and (10d), as there is no longer an accent to differentiate between them. Finally, in the third method, (11c) (accent only), both classes *must* be differentiated in the phonetic implementation: accent is interpreted as high pitch, while lack of accent is interpreted differently, although also in terms of elevated pitch.

I have briefly discussed three different approaches to a system such as that of Tokyo Japanese nouns, namely those mentioned in (11). All three approaches have been defended in the literature in one form or another. The tone accent approach (although often *called* "pitch accent approach"), (11a), comes closest to the analysis offered in McCawley (1968). Lexically, the language is accentual, but in the course of the derivation (presumably at the word level) tone is added, and from that point on the language is tonal. This approach was adopted as part of the autosegmental analysis of languages like Japanese and other monomelodic systems (cf. Goldsmith 1976b; Haraguchi 1977, 1988). The tone-only approach, (11b), has been advocated by Meeussen (1972), Poser (1984), Pulleyblank (1986), Clark (1987), and Pierrehumbert and Beckman (1988). Lockwood (1983) is a clear representative of (11c), the pitch accent analysis.

To what extent do these linguists recognize the possibilities in (11), other than the one that they propose for Japanese, as valid for other languages? Clark (1988) rejects (11a) as a theoretical option, but claims that (11c) represents an independent possibility, alongside (11b). She makes a distinction between restricted tonal systems, i.e. (11b), and metrical pitch accent systems, i.e. (11c). The difference between the two types is claimed to be that only metrical pitch accent systems have the characteristics that we also find in non-tonal accent languages with respect to accent location (e.g. influence of syllable weight) and other phonetic cues that occur as the manifestation of accent. In her restricted tonal languages, the alleged accent is simply a tone at every level of representation (Clark 1988: 52). An argument for analyzing Tokyo Japanese as tonal would be the fact that we have words like (10d), distinct from (10c). In a tonal analysis, this difference is expected, since words do not have to have a tone. But in an accentual analysis,

15 Here I have added "first" to the rule because, if a word ends up having more than one accent, it is always the leftmost accent that attracts the H tone.

prosodic boundaries. We can, if accents have high pitch, first assign [H] to accent and then do the actual implementation. The same applies to boundaries; we can assign a [H] to the left boundary of a certain type of prosodic phrase. Strictly speaking, we only do this in order to make the implementation rules refer to only one type of entity (namely tonal entities, whether phonetic or phonological) instead of to three different types of entities (tones, accents, and boundaries). In any event, it would seem that the pitch profiles of Tokyo Japanese words do not require reference to word-level tones.

7 Accent and stress

A discussion of pitch accent systems forms part of the broader discussion of word prosodic systems. However, having made reference in the preceding sections to a view that recognizes both accent and stress as independent notions, this section will briefly discuss their properties and interaction.

We might entertain the idea that the alleged accents in Tokyo Japanese are simply "marks" which are to be compared to syllable weight. If this comparison holds, we might refer to the accents as "diacritic weight marks" and in that case there is no reason for every word to have one such mark, just as languages that have a contrast between CV (light) and CVX (heavy) syllables typically do not necessarily demand that each word has a "heavy syllable." Nor, for that matter, do we expect words to have only one "accent," since words also can have more than one heavy syllable. This interpretation of "accents" explains the occurrence of unaccented words and multiple accented words in specific systems.

A problem with this approach is that weight diacritics have characteristics that are more reminiscent of "stress" than of heavy syllables, notably predictability. This can be illustrated by taking a closer look at the accentual system of Tokyo Japanese. See CHAPTER 120: JAPANESE PITCH ACCENT, where it is shown that the Tokyo Japanese accent rule is very similar to the Latin-style English accent rule.

We now have a new problem. If the Tokyo Japanese accents are like weight, (a) why is their distribution predictable by rule, and (b) why is the rule so similar to the typical "stress" placement rules? And why are there accentual systems in which accent is culminative and/or obligatory? To resolve these issues, van der Hulst (2009, 2010c) proposes to account for accent and "rhythm," which traditional metrical theories conflate in one representation, in two different modules. The accentual module accounts for the location of so-called primary accent or primary "stress" in systems where this location shows influence of lexical factors (exceptions, morphological classes, etc.), while the rhythmic module associates words with metrical structures. This separation of tasks allows a simpler version of the metrical system, which, as van der Hulst shows, cannot handle all varieties of primary accent locations in bounded systems and is simply not designed to deal with accent locations in unbounded systems.

The theory of accent that has been suggested is "liberal," in that accent is required neither to be culminative nor obligatory. While this allows four different kinds of pitch accentual systems, it might be argued that we now predict four kinds of *any* sort of accent system, whatever the cues for accent. Focusing on the specific case of stress accent languages, Hyman (2009) argues that in such systems "stress" is always culminative and obligatory.

of accent) if, at the same time, cases in which accent is cued by non-distinctive duration or vowel quality are *not* analyzed as involving lexical specification of length or of non-distinctive vowel features?

ACKNOWLEDGMENTS

I would like to thank Carlos Gussenhoven, Marc van Oostendorp, Keren Rice, and two anonymous reviewers for valuable comments on earlier drafts of this chapter.

REFERENCES

Anderson, John M. & Colin J. Ewen. 1987. *Principles of dependency phonology*. Cambridge: Cambridge University Press.

Banti, Giorgio. 1988. Two Cushitic systems: Somali and Oromo nouns. In van der Hulst & Smith (1988), 11–49.

Beckman, Jill N. 1998. Positional faithfulness. Ph.D. dissertation, University of Massachusetts, Amherst.

Beckman, Mary E. 1986. *Stress and non-stress accent*. Dordrecht: Foris.

Biber, Douglas. 1981. The lexical representation of contour tones. *International Journal of American Linguistics* 47. 271–282.

Blevins, Juliette. 1993. A tonal analysis of Lithuanian nominal accent. *Language* 69. 237–273.

Bolinger, Dwight L. 1982. Intonation and its parts. *Language* 58. 505–532.

Bruce, Gösta. 1977. Swedish word accents in sentence perspective. *Lund University Department of Linguistics Working Papers* 12. 219–228.

Bruce, Gösta. 1999. Word tone in Scandinavian languages. In van der Hulst (1999b), 605–633.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clark, Mary. 1987. Japanese as a tone language. In Takashi Imai & Mamoru Saito (eds.) *Issues in Japanese linguistics*, 53–106. Dordrecht: Foris.

Clark, Mary. 1988. An accentual analysis of the Zulu noun. In van der Hulst & Smith (1988), 51–79.

Clements, G. N. 2001. Representational economy in constraint-based phonology. In T. A. Hall (ed.) *Distinctive feature theory*, 71–146. Berlin & New York: Mouton de Gruyter.

Clements, G. N. 2009. The role of features in speech sound inventories. In Raimy & Cairns (2009), 19–68.

Clements, G. N. & John A. Goldsmith (eds.) 1984. *Autosegmental studies in Bantu tone*. Dordrecht: Foris.

Coetsem, Frans van. 1996. *Towards a typology of lexical accent: Stress accent and pitch accent in a renewed perspective*. Heidelberg: Carl Winter.

Cutler, Anne. 1984. Stress and accent in language understanding and production. In Dafydd Gibbon & Helmut Richter (eds.) *Intonation, accent and rhythm*, 77–90. Berlin & New York: Mouton de Gruyter.

de Lacy, Paul. 2002. The interaction of tone and stress in Optimality Theory. *Phonology* 19. 1–32.

Downing, Laura J. 2010. Accent in African languages. In van der Hulst *et al.* (2010b), 381–428.

Duanmu, San. 2000. *The phonology of Standard Chinese*. Oxford: Oxford University Press.

Duanmu, San. 2004. Tone and non-tone languages: An alternative to language typology and parameters. *Language and Linguistics* 5. 891–924.

Fox, Anthony. 2000. *Prosodic features and prosodic structure: The phonology of suprasegmentals.* Oxford: Oxford University Press.

Fromkin, Victoria A. (ed.) 1978. *Tone: A linguistic survey.* New York: Academic Press.

Fukui, Rei. 2003. Pitch accent systems in Korean. In Shigeki Kaji (ed.) *Cross-linguistic studies of tonal phenomena: Historical development, phonetics of tone, and descriptive studies*, 275–286. Tokyo: ICLAA.

Garde, Paul. 1968. *L'Accent.* Paris: Presses Universitaires de France.

Goldsmith, John A. 1976a. Autosegmental phonology. Ph.D. dissertation, MIT.

Goldsmith, John A. 1976b. An overview of autosegmental phonology. *Linguistic Analysis* 2. 23–68.

Goldsmith, John A. 1981. English as a tone language. In Didier L. Goyvaerts (ed.) *Phonology in the 1980s*, 287–308. Ghent: E. Story-Scientia.

Goldsmith, John A. 1988. Prosodic trends in the Bantu languages. In van der Hulst & Smith (1988), 81–93.

Gordon, Matthew. Forthcoming. Disentangling stress and pitch accent: toward a typology of different prosodic levels. In Harry van der Hulst (ed.) *Word accent: Theoretical and typological issues*. Berlin & New York: Mouton de Gruyter.

Greenberg, Joseph H. & Dorothea Kaschube. 1976. Word prosodic systems: A preliminary report. *Working Papers on Language Universals* 20. 1–18.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation.* Cambridge: Cambridge University Press.

Gussenhoven, Carlos & Gösta Bruce. 1999. Word prosody and intonation. In van der Hulst (1999b), 233–271.

Halle, Morris & Jean-Roger Vergnaud. 1982. On the framework of autosegmental phonology. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, part II, 65–82. Dordrecht: Foris.

Haraguchi, Shōsuke. 1977. *The tone pattern of Japanese: An autosegmental theory of tonology.* Tokyo: Kaitakusha.

Haraguchi, Shōsuke. 1988. Pitch accent and intonation in Japanese. In van der Hulst & Smith (1988), 123–150.

Haraguchi, Shōsuke. 1991. *A theory of stress and accent.* Dordrecht: Foris Publications.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Hermans, Ben. 1994. The composite nature of accent, with case studies of the Limburgian and Serbo-Croatian pitch accent. Ph.D. dissertation, Free University, Amsterdam.

Hockett, Charles F. 1955. *A manual of phonology.* Baltimore: Waverly Press.

Hollenbach, Barbara. 1988. The asymmetrical distribution of tone in Copala Trique. In van der Hulst & Smith (1988), 167–182.

Hualde, José Ignacio. 1999. Basque accentuation. In van der Hulst (1999b), 947–994.

Hulst, Harry van der. 1999a. Word accent. In van der Hulst (1999b), 3–115.

Hulst, Harry van der (ed.) 1999b. *Word prosodic systems of the languages of Europe.* Berlin & New York: Mouton de Gruyter.

Hulst, Harry van der. 2009. Brackets and grid marks or theories of primary accent and rhythm. In Raimy & Cairns (2009), 225–245.

Hulst, Harry van der. 2010a. Word accent systems in the languages of Europe. In van der Hulst *et al.* (2010b), 429–508.

Hulst, Harry van der. 2010b. Word accent: Terms, typologies and theories. In van der Hulst *et al.* (2010b), 3–54.

Hulst, Harry van der. 2010c. Word accentual systems. Unpublished ms., University of Connecticut.

Hulst, Harry van der & Norval Smith. 1988. *Autosegmental studies on pitch accent.* Dordrecht: Foris.

Hulst, Harry van der, Keren Rice & W. Leo Wetzels. 2010a. Word accent systems in the languages of Middle America. In van der Hulst *et al.* (2010b), 249–312.

Hulst, Harry van der, Rob Goedemans & Ellen van Zanten (eds.) 2010b. *A survey of word accentual systems in the language of the world*. Berlin & New York: Mouton de Gruyter.

Hyman, Larry M. 1977. On the nature of linguistic stress. In Larry M. Hyman (ed.) *Studies in stress and accent*, 37–82. Los Angeles: Department of Linguistics, University of Southern California.

Hyman, Larry M. 1978. Tone and/or accent. In Donna Jo Napoli (ed.) *Elements of tone, stress and intonation*, 1–20. Washington: Georgetown University Press.

Hyman, Larry M. 1981. Tonal accent in Somali. *Studies in African Linguistics* 12. 169–203.

Hyman, Larry M. 1982. Globality and the accentual analysis of Luganda tone. *Journal of Linguistic Research* 2–3. 1–40.

Hyman, Larry M. 1989. Accent in Bantu: An appraisal. *Studies in the Linguistic Sciences* 19. 115–134.

Hyman, Larry M. 2001. Tone systems. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.) *Language typology and language universals: an international handbook*, vol. 2, 1367–1380. Berlin & New York: Mouton de Gruyter.

Hyman, Larry M. 2006. Word-prosodic typology. *Phonology* 23. 225–257.

Hyman, Larry M. 2007. Universals of tone rules: 30 years later. In Tomas Riad & Carlos Gussenhoven (eds.) *Tones and tunes*, vol. 1: *Typological studies in word and sentence prosody*, 1–34. Berlin & New York: Mouton de Gruyter

Hyman, Larry M. 2009. How (not) to do phonological typology: The case of pitch-accent. *Language Sciences* 31. 213–238.

Inkelas, Sharon & Draga Zec. 1988. Serbo-Croatian pitch accent: The interaction of tone, stress, and intonation. *Language* 64. 227–248.

Kodzasov, Sandro. 1999. Caucasian: Daghestanian languages. In van der Hulst (1999b), 995–1020.

Ladd, D. Robert. 2008. *Intonational phonology*. 2nd edn. Cambridge: Cambridge University Press.

Leben, William R. 1971. Suprasegmental and segmental representation of tone. *Studies in African Linguistics*. Supplement 2. 183–200.

Liberman, Mark. 1975. The intonational system of English. Ph.D. dissertation, MIT. Published 1979, New York: Garland.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.

Lockwood, D. G. 1982. Parameters for a typology of word accent. *Lacus Forum* 9. 231–240.

Lockwood, D. G. 1983. Tone in a non-substantive theory of language. *Lacus Forum* 10. 131–140.

MacAulay, Donald (ed.) 1992. *The Celtic languages*. Cambridge: Cambridge University Press.

Martinet, André. 1960. *Eléments de linguistique générale*. Paris: Librairie Armand Colin.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague & Paris: Mouton.

McCawley, James D. 1978. What is a tone language? In Fromkin (1978), 113–131.

Meeussen, A. E. 1972. Japanese accentuation as a restricted tone system. *Papers in Japanese Linguistics* 1. 267–270.

Mock, Carol. 1988. Pitch accent and stress in Isthmus Zapotec In van der Hulst & Smith (1988), 197–223.

Nellis, Donald G. & Barbara Hollenbach. 1980. Fortis versus lenis in Cajonos Zapotec phonology. *International Journal of American Linguistics* 46. 92–105.

Odden, David. 1988. Predictable tone systems in Bantu. In van der Hulst & Smith (1988), 225–251.

Odé, Cecilia & Vincent van Heuven (eds.) 1994. *Phonetic studies of Indonesian prosody*. University of Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië.

Pierrehumbert, Janet B. 1980. The phonology and phonetics of English intonation. Ph.D. dissertation, MIT.

Pierrehumbert, Janet B. & Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.

Pike, Kenneth L. 1948. *Tone languages.* Ann Arbor: University of Michigan Press.
Poser, William J. 1984. The phonetics and phonology of tone and intonation in Japanese. Ph.D. dissertation, MIT.
Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14. 19–100.
Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology.* Dordrecht: Reidel.
Raimy, Eric & Charles Cairns (eds.) 2009. *Contemporary views on architecture and representations in phonology.* Cambridge, MA: MIT Press.
Rice, Keren. 2010. Accent in the native languages of North America. In van der Hulst *et al.* (2010b), 155–248.
Schadeberg, Thilo C. 1973. Kinga: A restricted tone system. *Studies in African Linguistics* 4. 23–47.
Schiering, René & Harry van der Hulst. 2010. Word accent systems in the languages of Asia. In van der Hulst *et al.* (2010b), 509–614.
Schuh, Russell C. 1977. Tone rules. In Fromkin (1978), 221–256.
Selkirk, Elisabeth. 1996. The prosodic structure of function words. In James L. Morgan & Katherine Demuth (eds.) *Signal to syntax: Bootstrapping from speech to grammar in early acquisition,* 187–213. Mahwah, NJ: Lawrence Erlbaum.
Suárez, Jorge A. 1983. *The Mesoamerican Indian languages.* Cambridge: Cambridge University Press.
Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie.* Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology.* Berkeley & Los Angeles: University of California Press.
Voorhoeve, Jan. 1973. Safwa as a restricted tone system. *Studies in African Linguistics* 4. 1–21.
Weidert, Alfons. 1981. *Tonologie: Ergebnisse, analysen, vermutungen.* Tübingen: Niemeyer.
Welmers, William E. 1973. *African language structures.* Berkeley: University of California Press.
Wetzels, W. Leo & Sergio Meira. 2010. A survey of South American stress systems. In van der Hulst *et al.* (2010b), 313–380.
Williamson, Kay. 1988. Tone and accent in Ịjọ. In van der Hulst & Smith (1988), 253–278.
Wright, Martha. 1983. A metrical approach to tone sandhi in Chinese dialects. Ph.D. dissertation, University of Massachusetts, Amherst.
Wright, Martha. 1988. A level-based model for pitch accent languages. In van der Hulst & Smith (1988), 295–316.
Yip, Moira. 1980. The tonal phonology of Chinese. Ph.D. dissertation, MIT.
Yip, Moira. 2002. *Tone.* Cambridge: Cambridge University Press.
Zanten, Ellen van & Philomena Dol. 2010. Word stress and pitch accent in Papuan languages. In van der Hulst *et al.* (2010b), 113–154.
Zanten, Ellen van, Ruben Stoel & Bert Remijsen. 2010. Stress types in Austronesian languages. In van der Hulst *et al.* (2010b), 87–112.

43 Extrametricality and Non-finality

Brett Hyde

1 Introduction

Extrametricality and *non-finality* are often treated as two terms that refer to the same principle of final stress avoidance, the former being implemented in a rule-based framework and the latter being implemented in a constraint-based framework. The labels and native frameworks of the two approaches, however, do not constitute the full extent of their differences. As Prince and Smolensky (1993) explain, extrametricality focuses on the parsability of prosodic constituents, while non-finality focuses on the position of stress peaks. Extrametricality is concerned primarily with dominance relations, and non-finality is concerned primarily with prominence relations.

Liberman and Prince (1977) introduced extrametricality in their foundational work on metrical stress theory to capture the apparent exclusion of certain English suffixes from the domain of stress rules.[1] Recognizing the potentially wide range of applications, Hayes (1980) proposed the general formulation for extrametricality rules in (1), where the initial or final constituent of a particular domain is designated as extrametrical.

(1)

$$C \rightarrow [+\text{extrametrical}] \; / \; \left\{ \begin{array}{l} \_\_\_\,]_D \\ {}_D[\, \_\_\_ \end{array} \right\}$$

The result of extrametricality is essentially invisibility to the application of subsequent rules. When a constituent is designated as extrametrical, it is excluded from the domain of rules that might incorporate it into higher levels of prosodic structure. An extrametrical segment cannot be associated with a mora; for example, an extrametrical syllable cannot be footed, and an extrametrical foot cannot be included in a prosodic word.

[1] Liberman and Prince introduce the notion of extrametricality to account for the apparent invisibility to stress rules of final *-y* in English: "From our point of view, *-y* functions as a kind of 'extrametrical' syllable; it simply does not take part in the metrical calculation" (Liberman and Prince 1977: 293). Later in the same paragraph: "*-y* is effectually *hors de combat* in the basic determination of metrical structure."

As part of his general approach, Hayes proposed four restrictions on extrametricality. The first, *constituency*, ensures that only constituents – segments, syllables, feet, affixes, and so on – can be designated as extrametrical. *Peripherality* restricts extrametrical constituents to the edges of a domain, while *edge markedness* prefers that they occur at the right edge. Finally, *non-exhaustivity* ensures that extrametricality cannot exhaust the domain of a rule, preventing it from applying altogether.

Prince and Smolensky (1993) incorporated similar restrictions into non-finality when they presented it as a replacement for extrametricality as part of their initial work on Optimality Theory. As (2) indicates, non-finality only applies at the edge of a domain (peripherality), and it only applies at the right edge in particular (edge markedness). The stress peaks that must avoid the right edge are prosodic categories (constituency) that are the heads of larger categories.

(2) *Head-based non-finality*

No head *Cat1* of a *Cat2* occurs in final position in *Cat3* (where *Cat1*, *Cat2*, and *Cat3* are prosodic categories).

The effect of non-finality constraints is to prevent prominent categories – the heads that represent stress peaks – from occurring at the right edge of a domain. Non-finality might prevent the head moras of syllables from occurring at the right edges of feet, for example, or head syllables of feet from occurring at the right edge of a prosodic word.

Although it is usually a simple matter to distinguish non-finality from extrametricality, some approaches do exhibit characteristics of both. This is especially true of approaches that target relationships between final constituents and entries on the metrical grid, the classical device for representing stress patterns (see CHAPTER 41: THE REPRESENTATION OF WORD STRESS). As (3) indicates, the non-finality constraints of Hyde (2003) prohibit stress peaks – grid entries – in final position, but they specify a particular final constituent that stress must avoid.

(3) *Grid-based non-finality*

No *Cat1*-level grid entry occurs over the final *Cat2* of *Cat3* (where *Cat1*, *Cat2*, and *Cat3* are prosodic categories).

Under the grid-based approach, a non-finality constraint might prevent foot-level grid entries (secondary stress) from occurring over the final mora of a foot, for example, or prosodic word-level entries (primary stress) from occurring over the final foot of a prosodic word.

The grid-based non-finality approach is like the head-based approach, then, in that it focuses on stress peaks, but it is similar to an extrametricality approach in that it excludes a particular final element from associating with some structure (in this case, certain levels of the metrical grid). A similar mixture of characteristics can be found in approaches that are typically considered extrametricality approaches. Since the grid-based account of Prince (1983) lacked feet, for example, the effect of extrametricality was to prevent syllables from mapping to the metrical grid – in other words, from associating with a stress peak – rather than to prevent them from being footed.

from left to right. With the final syllable extrametrical, the last two syllables in an even-parity form cannot form an iambic foot, so the expected final stress fails to appear. The difference between the two languages would be that Aguaruna tolerates degenerate feet – and can parse the penult as a degenerate foot after iambic footing is no longer possible – but Choctaw does not. Since Aguaruna can parse the penultimate syllable as a degenerate foot, as (6) illustrates, the expected final stress shifts to the penult.

(6) *Aguaruna extrametricality analysis*

		extrametricality		parsing
i.ʧi.na.ka	→	i.ʧi.na⟨ka⟩	→	(i.ˈʧi)(ˌna)⟨ka⟩

Since Choctaw cannot parse the penult as a degenerate foot, however, as (7) illustrates, the expected final stress is absent altogether.

(7) *Choctaw extrametricality analysis*

		extrametricality		parsing
ʧi.pi.sa.li	→	ʧi.pi.sa⟨li⟩	→	(ʧi.ˈpi)sa⟨li⟩

For additional, and more detailed, extrametricality analyses of final stress avoidance, see Halle and Vergnaud (1987) and Hayes (1995).

A non-finality approach produces the same patterns, although a bit more directly, simply by prohibiting stress at the right edge of the word. Head-based non-finality, where heads represent stress, prohibits the head syllable of a foot from occurring in final position. Grid-based non-finality, where grid entries represent stress, prohibits a foot-level gridmark from occurring over the final syllable. In either case, prohibiting final stress effectively prohibits a final iambic foot.

The difference between Aguaruna and Choctaw is in the options that they employ to avoid a final iamb. As (8) illustrates, Aguaruna employs a final trochaic foot, shifting the expected final stress to the penult. Notice that the non-finality analysis does not necessarily require underparsing like the extrametricality analysis. (In (8) and examples throughout, the expression "X >> Y" indicates that X is more harmonic than Y or that some constraint, in this case non-finality, prefers X to Y.)

(8) *Aguaruna non-finality analysis*

	non-finality	
(i.ˈʧi)(ˌna.ka)	>>	(i.ˈʧi)(na.ˌka)

In contrast, as (9) illustrates, Choctaw prefers to leave its final two syllables unparsed in order to avoid a final iamb.[3]

(9) *Choctaw non-finality analysis*

	non-finality	
(ʧi.ˈpi)sa.li	>>	(ʧi.ˈpi)(sa.ˈli)

[3] An alternative to leaving the final two syllables unparsed is to parse them into a stressless foot: (ʧi.ˈpi)(sa.li). See Hyde (2002) for discussion.

See Hyde (2002, 2003) for more detailed grid-based non-finality analyses of final stress avoidance, and McCarthy and Prince (1993) and Kenstowicz (1995) for more detailed head-based non-finality analyses.

In this section, then, we have seen that extrametricality and non-finality provide equally effective analyses for the avoidance of stress on final syllables. Both approaches account for cases where a binary stress pattern is perturbed at the right edge, whether the final stress arrives early or is absent altogether.

3 Final feet

Another important use of extrametricality and non-finality has been to prevent primary stress from occurring over a word-final foot. In the clearest examples of the phenomenon, the primary stress is the penultimate stress, the presence of a secondary stress further to the right being the clearest indication that there is a final foot that primary stress might have occupied.

Consider Banawá (Buller *et al.* 1993; Everett 1996, 1997) and Paumari (Everett 2003). In Banawá, consonant-initial words have a trochaic pattern, and vowel-initial words have an iambic pattern. In both the trochaic pattern and the iambic pattern, however, the primary stress is the penultimate stress. The secondary stress that follows indicates that there is a final foot that primary stress might have occupied if it had been drawn as far to the right as possible.

(10) *Primary stress in Banawá*

a.	aˈbariˌko	'moon'
b.	ˌmetuˈwasiˌma	'find them'
c.	ˌtinaˈrifaˌbune	'you are going to work'

The primary stress is also the penultimate stress in the consistently iambic Paumari, indicating the presence of a final foot that primary stress has avoided.

(11) *Primary stress in Paumari*

a.	kaˈbahaˌki	'to get rained on'
b.	ˌahaˈkabaˌra	'dew'
c.	aˌtʰanaˈrariˌki	'sticky consistency'
d.	biˌkanaˌtʰaraˈraviˌni	'to cave in, to fall apart quickly'

It is a relatively simple matter to produce the Banawá and Paumari patterns with either extrametricality or non-finality. In the extrametricality approach, a word-final foot is designated as extrametrical, excluding it from the prosodic word. As (12) illustrates, when a right-headed prosodic word is constructed, it positions the primary stress over the penultimate foot, rather than the final.

(12) *extrametricality* *word layer*

```
                                                  (        x    )
( x   )( x   )( x    ) → ( x   )( x   )⟨( x    )⟩ → ( x   )( x   )⟨(x     )⟩
 ti na  ri fa  bu ne      ti na  ri fa   bu ne       ti na  ri fa   bu ne
```

The non-finality approach produces a similar result, although it does not require that final feet be excluded from the prosodic word. Head-based non-finality avoids primary stress on final feet by prohibiting head feet from occurring in final position. Grid-based non-finality prohibits a prosodic word-level gridmark from occupying the final foot.

(13) *non-finality*

```
(       x          )    (             x    )
( x   )( x   )( x    )   ( x   )( x   )( x    )
 ti na  ri fa  bu ne >>  ti na  ri fa  bu ne
```

In either case, the primary stress and the foot associated with it are pushed back from the right edge. As a result the primary stress is the penultimate stress, and the associated foot the penultimate foot.

It should be noted at this point that many of the examples cited in the literature on primary stress avoiding final feet are not as compelling as those discussed above. Hayes (1995) presents several languages as potential examples of foot extrametricality: Bedouin Arabic (Blanc 1970), Cayuga (Chafe 1977; Foster 1982; Michelson 1988), Delaware (Goddard 1979, 1982), Eastern Ojibwa (Piggott 1980, 1983), and Palestinian Arabic (Kenstowicz and Abdul-Karim 1980; Kenstowicz 1983). McCarthy (2003) points out, however, that these are not especially clear cases, because no secondary stress has been reported in a position associated with the supposed extrametrical foot. While McCarthy's point overreaches a bit – Piggott (1983) reports post-tonic secondary stresses in Ojibwa, and patterns of reduction and non-reduction suggest post-tonic feet in Delaware – it is true of many of the traditional examples. As Banawá and Paumari demonstrate, however, the avoidance of stress on final feet is still one of the important functions performed by extrametricality and non-finality.

4 Final moras

The avoidance of final moras can make stress sensitive to the weight of syllables generally or to the weight of domain-final syllables (see CHAPTER 57: QUANTITY-SENSITIVITY). As we shall see below, non-finality offers a relatively straightforward analysis in such cases. §4.1 demonstrates how avoidance of syllable-final moras promotes general weight-sensitivity, §4.2 how avoidance of prosodic word-final moras promotes sensitivity to the weight of prosodic word-final syllables, and §4.3 how avoidance of foot-final moras promotes rhythmic lengthening. §4.4 examines the difficulties confronting extrametricality analyses.

4.1 General weight-sensitivity

A fairly common type of weight-sensitivity is the type where stress avoids light syllables. It has been addressed, for example, using the Obligatory Branching Parameter (Hayes 1980) of classical metrical theory and the PEAK-PROMINENCE (Prince and Smolensky 1993) and STRESS-TO-WEIGHT (Hammond and Dupoux 1996; Lorentz 1996) constraints of Optimality Theory. As Hyde (2006, 2007b) demonstrates, Similarly, the Aguacatec pattern would emerge when non-finality in the syllable takes precedence over rightward orientation.

Moraic non-finality constraints applied to the syllable domain, then, have the same effect as Obligatory Branching, PEAK-PROMINENCE and STRESS-TO-WEIGHT. One point that favors the non-finality approach over the others is that non-finality constraints are motivated by their usefulness in a much wider range of contexts – the avoidance of stress on final syllables (§2) and feet (§3), for example – many of which have nothing to do with syllable weight.

4.2 *Weight-sensitivity in word-final syllables*

In this section, we examine the situation where stress is sensitive to the weight of prosodic word-final syllables only. To make stress sensitive to the weight of prosodic word-final syllables, all that is necessary is to require that prosodic word-final moras be stressless. When word-final moras cannot be stressed, stress can occupy a heavy final syllable, but it cannot occupy a light final syllable.

One situation where stress is sensitive to the weight of prosodic word-final syllables only arises in syllabic trochee systems. Consider the case of Wergaia (Hercus 1986), where heavy syllables are syllables with long vowels (typically limited to initial position), syllables with diphthongs (limited to initial or final position), and closed syllables. As (17) illustrates, Wergaia stress is largely weight-insensitive. It falls automatically on every odd-numbered syllable counting from the left, except the final syllable. Stress falls on final syllables only if they are odd-numbered and heavy, as in (17f) and (17g). It avoids final syllables if they are light, as in (17d) and (17e).

(17) *Avoidance of final light syllables in Wergaia*

a.	ˈLL	ˈwuru	'mouth'
b.	ˈHL	ˈŋaːri	'oak tree'
c.	ˈLH	ˈŋarau	'wild turkey'
d.	ˈHLL	ˈmaːbila	'to tell lies'
e.	ˈLHL	ˈdaguŋga	'to punch someone'
f.	ˈLLˌH	ˈbunaˌɖug	'broad-leaved mallee'
g.	ˈLLˌH	ˈwaɲaˌgai	'catfish'
h.	ˈLLˌLL	ˈbunaˌmala	'fine-leaved mallee'
i.	ˈLLˌLH	ˈwuregˌwuraŋ	'speaking together, gabbing'

In Hyde's (2007b) grid-based non-finality approach, the Wergaia pattern emerges when it is more important that foot-level gridmarks avoid prosodic word-final moras than it is that feet contain a stressed syllable. When the final syllable of an odd-parity form is light, stress cannot occur on the final syllable without occurring on the final mora, so the final foot emerges without a stress, as in (18).

(18) *Moraic non-finality preferences in the prosodic word*

```
( x      )( )     ( x      )(x)
  μμ   μ  μ  >>    μμ   μ  μ
 maa . bi . la    maa . bi . la
```

When the final syllable is heavy, however, stress can occupy the final syllable without occupying the final mora, so the final foot emerges with a stress, as in (19).

(19)
```
( x      )( x )
  μ   μ   μμ
 wa . ɲa . gai
```

The same result can be produced in a more standard structural framework when non-finality in the prosodic word takes precedence over the constraints that require syllables to be parsed into feet. Odd-parity forms with a light final syllable would emerge with the final syllable unparsed and stressless. Odd-parity forms with a heavy final syllable would emerge with the final syllable parsed and stressed. The result can also be produced with head-based non-finality by prohibiting the head mora of a foot from being final in the prosodic word.[5]

4.3 *Rhythmic lengthening*

Non-finality can be used not only as a simple detector of syllable weight – the use focused on in §4.1 and §4.2 – but also as a trigger to increment syllable weight. When stress would fall on an underlyingly light syllable, non-finality can force the syllable to become heavy on the surface. Rhythmic lengthening is an example of this effect. It results from avoidance of stress on foot-final moras or syllable-final moras.

There are two types of rhythmic lengthening: iambic lengthening and trochaic lengthening. The former adds a mora to the stressed syllable of an iamb; the latter adds a mora to the stressed syllable of a trochee. The iambic type appears to occur more frequently than the trochaic type (Hayes 1985, 1987, 1995; Kager 1993; CHAPTER 44: THE IAMBIC–TROCHAIC LAW), but both are well attested. Iambic lengthening can be found in Carib (Hoff 1968), for example. As (20) illustrates, Carib lengthens even-numbered syllables counting from the left, but not the final syllable, producing a fairly typical iambic pattern.

(20) *Iambic lengthening in Carib*

a.	tonoro	→ tono:ro	'large bird'
b.	kurijara	→ kuri:jara	'canoe'
c.	woturoporo	→ wotu:ropo:ro	'cause to ask'
d.	woturopotake	→ wotu:ropo:take	'I shall ask'

Trochaic lengthening can be found in Chimalapa Zoque (Knudson 1975), a dual stress language based on trochaic footing. In Chimalapa Zoque, stress occurs on the initial syllable and the penult, with the stress on the penult being primary. As (21) illustrates, every stressed syllable must be heavy on the surface. When an underlyingly light syllable is stressed, the syllable is made heavy by lengthening its vowel.

[5] An alternative approach is to rely on a foot minimality requirement to distinguish between light and heavy final syllables. This is essentially the approach adopted by Hayes (1995). As Hyde (2007b) points out, however, such an approach produces the same type of weight-sensitivity in non-final syllables, as well, where it is, unfortunately, unattested.

(25) a. *Final* CV

i. *Preceding consonant stray*

```
( x        )
  μ     μ   ⟨μ⟩
[CV] [CV] CV
```

ii. *Preceding consonant as coda*

```
( x          )
  μ      μ   ⟨μ⟩
[CV] [CVC] V
```

b. *Final* CVC

```
( x         )(  x)
  μ      μ      μ ⟨μ⟩
[CV] [CV] [CV] C
```

Hayes does not actually consider this second option and, therefore, does not reject it explicitly. Its rejection is implied, however, by his assumption that extrametricality never prevents syllabification. Unfortunately, he only justifies the assumption as it relates to extrametrical consonants, but there do seem to be good reasons for applying it to extrametrical moras as well.

The analysis presents some fairly obvious problems in CV-final words. When the final mora and its associated vowel remain unsyllabified, there are essentially two options for dealing with any consonants that would otherwise be part of the final vowel's onset. First, they might be left stray, as in (25a.i), in which case they would be subject to Stray Erasure and deleted (see §6.3). The result would be a language where final CVC and CVV sequences are preserved and their vowels stressed but final CV sequences have their consonants deleted and their vowels left stressless. To my knowledge, such an outcome is unattested. Second, preceding consonants might be incorporated into the preceding syllable as a coda, as in (25a.ii). The result would be a language where final CVV and CVC sequences always have their consonants syllabified as onsets and their vowels stressed but final CV sequences always have their consonants syllabified as codas and their vowels left stressless. To my knowledge, this outcome is also unattested.[6]

There is an additional, primarily theoretical, reason for rejecting extrasyllabic moras, not only in the particular situation under consideration, but in all situations. Moras are unique in that the primary motivation for including them in the prosodic hierarchy in the first place is to provide an effective representation of syllable weight, a function that cannot be performed outside the syllable.

Neither of the options that might be used to achieve the desired results under an extrametricality approach, then, appears to be viable.

5 Final consonants

Evolved from proposals by Mohanan (1979) and Hayes (1980), traditional consonant extrametricality rules prevent word-final consonants from having moraic status and, therefore, from contributing to the weight of final syllables. The result is that final syllables that end in a consonant are lighter than we would otherwise

[6] Similar arguments can be made against proposals that involve extrasyllabic moras acting as a sort of prosodic licenser for otherwise stray segments in order to protect them from deletion through Stray Erasure (Downing 1993; Everett 1996). Although such licensing has only been employed at the left edge, there would seem to be nothing to prevent it applying at the right edge, as well, leading to the results illustrated in (25).

expect. As (26) illustrates, final CV and CVV are unaffected. Final CVC, CVVC, and CVCC, however, are all lighter than they would be otherwise. Final CVC, normally bimoraic, emerges as monomoraic and counts as light. Final CVVC and CVCC, normally trimoraic, emerge as bimoraic and count as heavy.

(26) *Weight contrasts under consonant extrametricality*

a. *Light syllables*

i.	μ CV	ii.	μ CV⟨C⟩

b. *Heavy syllables*

i.	μμ CVV	ii.	μμ CVV⟨C⟩	iii.	μμ CVC⟨C⟩

Among the languages that have been argued to exhibit consonant extrametricality are English (Hayes 1982), various dialects of Arabic (McCarthy 1979; Hayes 1995), Ancient Greek (Steriade 1988), Spanish (Harris 1983), and Estonian (Hint 1973; Prince 1980). Examples from Estonian are provided in (27).

(27) *Final syllables in Estonian*

a. ˈkavaˌlatt 'cunning'
b. ˈpaheˌmait 'worse (PART PL)'
c. ˈpimestav 'blinding'
d. ˈpimesˌtavale 'blinding (ILL SG)'

Like Wergaia, Estonian automatically stresses every odd-numbered syllable except the final syllable. Final syllables are stressed only if they are heavy, as in (27a) and (27b). When a final syllable is light, as in (27c) and (27d), it is unstressed. Since final CVV, CVVC, and CVCC are always stressed, they must pattern together in counting as heavy. Since final CV and CVC are always stressless, they must pattern together in counting as light. This is exactly the division predicted by consonant extrametricality.

Since moras are not stress peaks, non-finality cannot directly prohibit moras from associating with final consonants. Non-finality can only affect a final consonant's moraic status by referring to a stress peak that coincides with moras. The success of a non-finality approach, however, depends crucially on the representation of stress peaks. Under head-based non-finality, no stress peak coincides with moras generally. A mora coincides with a stress peak only if it is a head mora, and banning head moras from final position does not ban all moras.[7] In contrast,

[7] As an anonymous reviewer points out, the claim that head-based non-finality cannot prevent final consonants from being moraic depends on the assumption that moras – unlike the higher prosodic categories – do not have heads. If moras have head segments, as argued by de Lacy (1997), then final consonants might be prevented from being moraic by prohibiting head segments from being final in the prosodic word. There are several arguments against this approach, however, one of which is that segments are not constituents of moras in the usual sense. It is often the case that multiple moras are associated with single segments. In such cases, not only would each mora have exactly the same single constituent, but it would also have exactly the same head. Neither situation is tolerated at higher prosodic levels, even in fairly permissive theories, like Hyde (2002), that allow prosodic categories to share constituents.

assuming that moras map to the base-level of the grid, there are stress peaks that coincide with moras generally under grid-based non-finality. By prohibiting base-level gridmarks from occurring over prosodic-word final consonants, non-finality can prevent final consonants from associating with moras.

To illustrate, when it is more important for final consonants to avoid associating with base-level gridmarks than it is for coda consonants to be moraic, final consonants will give up their moraic status to avoid associating with base-level gridmarks. Final CVC syllables emerge as monomoraic and light, and final CVVC and CVCC syllables emerge as bimoraic and heavy, resulting in the same weight distinctions among final syllables as those created by consonant extrametricality.

(28) *Consonantal non-finality preferences*

```
a.   x            x x
     μ     >>     μ μ
    CVC           CVC

b.   x x          x x x
     μ μ   >>     μ μ μ
    CVVC          CVVC

c.   x x          x x x
     μ μ   >>     μ μ μ
    CVCC          CVCC
```

Given its parsability focus, then, the extrametricality analysis is the most straightforward for cases like Estonian. Since it makes final consonants invisible to the process of mora assignment, consonant extrametricality produces the desired weight distinctions in a fairly direct fashion. While it is also possible to provide a non-finality analysis, it is only possible to do so with a grid-based approach.

For additional discussion of this and other issues concerning final consonants, see CHAPTER 36: FINAL CONSONANTS.

6 The classic arguments

We turn now to some of the classic arguments marshaled in support of extrametricality and briefly consider whether or not they also provide support for non-finality. Below we consider three of extrametricality's traditional uses: establishing trisyllabic stress windows, helping to capture generalizations about the stress patterns of different lexical classes, and helping to provide a general account of the deletion of unsyllabifiable segments.

6.1 *Eliminating ternary foot templates*

In many languages, a word's final three syllables form a domain that is crucial in creating the appropriate stress pattern. The most direct option for creating such a domain – establishing it with a trisyllabic foot – has the disadvantage of making it necessary to expand the foot inventory beyond the well-motivated binary templates to include less well-motivated ternary templates. As Hayes (1980) demonstrates, a less direct extrametricality approach allows us to maintain the smaller inventory. It allows the theory to create trisyllabic windows using a binary foot followed by an unparsed syllable.

Consider the stress pattern of Latin. In Latin words of at least three syllables, stress falls on either the antepenult or penult, depending on the weight of the latter. If the penult is heavy, it is stressed; otherwise, the antepenult is stressed.

(29) *Trisyllabic stress window in Latin*

a.	L'HH	a'miːkus	'friend (NOM SG MASC)'
b.	LH'HH	moneː'baːmus	'warn (1PL IMPERF INDIC ACT)'
c.	L'LLH	ko'mitium	'the election site in the forum (NOM SG)'
d.	L'HLH	do'mestikus	'domestic (NOM SG MASC)'

Without extrametricality, the Latin pattern requires the quantity-sensitive ternary template (σ́ L σ) to establish the appropriate trisyllabic domain at the right edge. When the penult is light, the template is used to construct a ternary foot at the right edge of the word, resulting in antepenultimate stress. When the penult is heavy, however, the template allows only a binary foot, resulting in penultimate stress.

(30) *Ternary foot analysis*

			parsing
a.	do.mes.ti.kus	→	do('mes.ti.kus)
b.	mo.neː.baː.mus	→	mo.neː('baː.mus)

Extrametricality makes the ternary template unnecessary, allowing the trisyllabic domain to be formed with an unparsed syllable and a maximally disyllabic foot. The unparsed syllable is the result of syllable extrametricality. The maximally disyllabic foot is produced with the quantity-sensitive template (σ́ L). If the penult is light, as in (31a), the template allows for a disyllabic foot at the right edge. In combination with the extrametrical syllable, the result is stress on the antepenult. If the penult is heavy, however, as in (31b), the template only allows for a monosyllabic foot, resulting in stress on the penult.

(31) *Extrametricality analysis*

			extrametricality		*parsing*
a.	do.mes.ti.kus	→	do.mes.ti⟨kus⟩	→	do('mes.ti)⟨kus⟩
b.	mo.neː.baː.mus	→	mo.neː.baː⟨mus⟩	→	mo.neː('baː)⟨mus⟩

As Prince and Smolensky (1993) demonstrate, a head-based non-finality approach can also construct trisyllabic domains from a binary foot and an unparsed syllable. When it is more important for the head foot to avoid final position than it is for the head foot to occur as far to the right as possible, the desired pattern emerges.

(32) *Head-based non-finality analysis*

		non-finality	
a.	do('mes.ti)kus	>>	do.mes('ti.kus)
b.	mo.neː('baː)mus	>>	mo.neː.baː('mus)

With the head foot pushed back from the right edge by non-finality, a disyllabic foot positions stress on the antepenult when the penult is light, and a monosyllabic foot positions it on the penult when the penult is heavy.

In the case of Latin, extrametricality and head-based non-finality have a very similar effect. They both result in the final syllable being left unparsed. The similarity arises because the stress peak that must avoid final position in the non-finality analysis happens to be a foot, the head foot of the prosodic word. If the head foot must be the rightmost foot but cannot be final, then the final syllable must remain unfooted, the very situation demanded when a final syllable is made extrametrical.

Two points should be kept in mind, however. The first is that grid-based non-finality is unable to produce this same result. Since stress peaks do not double as prosodic constituents, grid-based non-finality cannot require that final syllables remain unfooted.[8] Second, as Hyde (2008) points out, even head-based non-finality does not offer a general approach to trisyllabic stress windows. It is unable to produce the stress window of Macedonian (Comrie 1976), for example. An alignment-based analysis actually provides a more successful general approach.

For a discussion of ternary stress intervals more generally, not just those limited to word edges, see CHAPTER 52: TERNARY RHYTHM.

6.2 *Similarities between lexical classes*

In many languages, one class of lexical items exhibits one stress pattern, while a different class exhibits a slightly different pattern. In many cases, the difference can be reduced to an extrametricality effect that one class exhibits and the other does not. Once the extrametricality effect is recognized, the similarities between the patterns become apparent, and it is possible to address both with a more unified approach. English (Hayes 1982), Spanish (Harris 1983), and Yawelmani (Archangeli 1984) are among the languages where extrametricality has played an important role in this context. English is used to illustrate below.

At first glance, English verbs and nouns seem to have very different stress patterns. In verbs, the position of stress depends on the shape of the final syllable. If the ultima is CVV, CVVC, or CVCC, the ultima is stressed. If the ultima is CV or CVC, the penult is stressed.

[8] As an anonymous reviewer points out, whether or not grid-based non-finality can prevent the final syllable from being footed depends on the particular structures that are assumed to be the constituents of feet. If feet are actually built on base-level gridmarks, rather than syllables, preventing the final syllable from mapping to a base-level gridmark would also prevent it from being footed. The grid-based non-finality approach presented here, however, assumes that metrical structure and prosodic structure are independent, so that feet are built on syllables. Under this approach, the failure of a final syllable to map to the grid would not prevent it from being footed.

(33) *English verbs*

o'bey	de'velop
a'tone	as'tonish
tor'ment	

In nouns, the position of stress depends on the weight of the penult. If the penult is heavy, stress appears on the penult. Otherwise, it appears on the antepenult.[9]

(34) *English nouns*

a'genda	A'merica
e'litist	'discipline
Ari'zona	'labyrinth

As Hayes (1982) demonstrates, the difference between verbs and nouns is that they show the effects of two different types of extrametricality. The verb pattern is influenced by consonant extrametricality, the evidence being the characteristic weight distinctions among final syllables (see §5). The noun pattern is influenced by syllable extrametricality, the evidence being the presence of a trisyllabic stress window (see §6.1).

Once we allow for the two different types of extrametricality, the correct patterns emerge for both verbs and nouns when we use the quantity-sensitive binary template (σ́ L) to construct a foot at the right edge. In verbs, the (σ́ L) template positions stress on the penult when the ultima emerges as light, once the effects of consonant extrametricality have been taken into account. It positions stress on the ultima when the ultima emerges as heavy.

(35) *English verbs and consonant extrametricality*

```
a.                              ( x        )
      μ    μ    μ       →     μ    μ    μ
      de . ve . lo⟨p⟩         de . ve . lo⟨p⟩

b.                               ( x    )
      μμ    μμ          →     μμ    μμ
      tor . men⟨t⟩            tor . men⟨t⟩
```

Once syllable extrametricality excludes final syllables from the foot layer in nouns, the same (σ́ L) template positions stress on the antepenult when the penult is light. It positions stress on the penult when the penult is heavy.

(36) *English nouns and syllable extrametricality*

```
a.                        →     ( x     )
      A . me . ri ⟨ca⟩          A . me . ri ⟨ca⟩

b.                        →     ( x )
      a . gen ⟨da⟩              a . gen ⟨da⟩
```

[9] This generalization applies to English nouns with a stressless final syllable. Nouns with final stress must be treated differently.

Extrametricality, then, allows us to extract the aspects of the English verb and noun patterns that differ, in order to capture the similarities in a single general stress rule. The analysis consists of two independently motivated extrametricality rules, the source of the differences, and a single, general stress rule, the source of the similarities. If extrametricality were unavailable, we would be forced to incorporate its effects directly into separate stress rules for verbs and nouns, making both that much more complicated and obscuring the similarities between the patterns.

It is not a straightforward matter to reproduce the extrametricality analysis in this case with a non-finality analysis. As mentioned in §6.1, head-based non-finality can produce the type of stress window found in Latin and in English nouns, but grid-based non-finality cannot. As mentioned in §5, however, grid-based non-finality can reproduce the consonant extrametricality effect seen in English verbs, but head-based non-finality cannot. Although non-finality could, in principle, help to capture similarities between the stress patterns of different lexical classes, then, its success depends very much on the facts of the particular case.

6.3 Licensing segments

Itô (1986) puts extrametricality to a use that is quite different from those discussed thus far. In the types of effects discussed above, extrametricality makes a domain-final constituent invisible to rules that create prosodic structure. Itô, however, uses extrametricality to make domain-final segments invisible to Stray Erasure (Harris 1983), a rule that deletes unsyllabified segments. The result is a theory of syllabification that relies on general, rather than idiosyncratic, deletion rules.

As a simple illustration, consider deletions that occur as part of the syllabification process in Diola Fogny (Sapir 1965). Diola prefers not to syllabify obstruents as codas. As seen in (37a)–(37c), a medial obstruent that would otherwise be syllabified as a coda ends up being deleted instead. The preference to avoid obstruent codas seems to be thwarted at the right edge of the word, however, as seen in (37d). Final obstruents are not deleted, even though they cannot be syllabified as anything other than a coda.

(37) *Obstruent deletion in Diola Fogny*

a.	letkuɟaw	→	lekuɟaw	'they won't go'
b.	uɟukɟa	→	uɟuɟa	'if you see'
c.	kobkoben	→	kokoben	'yearn, long for'
d.	kuɲilak	→	kuɲilak	'the children'

Extrametricality accounts for the different treatment of final and medial obstruents. In the lexical phonology, Diola's coda condition prevents obstruents from being syllabified if they would syllabify as codas. Stray Erasure then deletes any segment that remains unsyllabified and has not been designated as extrametrical. Since medial consonants cannot be designated as extrametrical – due to the Peripherality restriction – medial obstruents that fail to syllabify are always deleted, as in (38a). Since final consonants can be designated as extrametrical, however, final obstruents are invisible to Stray Erasure and escape deletion, as in (38b), even syllables often have a longer rhyme might make them seem more compatible with final lengthening. The fact that intensity declines in the rhyme under phonetic final lengthening but increases under stress, however, suggests that this is really not the case. The increased intensity in the rhyme and the strengthening of the onset makes stress more compatible with initial lengthening.

Based on a parallel phenomenon in music (Gabrielsson 1987, 1993), Hyde (2009) suggests that different types of tempo changes at prosodic boundaries might account for the different characteristics of initial and final lengthening. Initial lengthening is the result of a strong attack and acceleration to medial tempo, while final lengthening is the result of a deceleration from medial tempo. An initial acceleration results in strengthening of initial segments and increased intensity in initial syllables, characteristics consistent with stress. A final deceleration results in declining intensity in final rhymes, a characteristic consistent with stresslessness.

7.2 *Stress typologies*

The second line of evidence against initial extrametricality and "non-initiality" is that they result in a decline in the accuracy of typological predictions (Hyde 2002; Altshuler 2009). Consider, for example, the iambic patterns of Aguaruna and Choctaw, discussed in §2. They emerge when rightward binary alternation of unstressed and stressed syllables is perturbed at the right edge of even-parity forms in order to avoid final stress. Aguaruna avoids final stress by shifting it one syllable to the left, and Choctaw avoids it simply by not assigning it.

(39) a. *Unattested*
σσ́σ́σσ́σ
σσ́σσ́σσ́σ

b. *Aguaruna*
σσ́σσ́σ́σ
σσ́σσ́σσ́σ

c. *Unattested*
σσσ́σσ́σ
σσ́σσ́σσ́σ

d. *Choctaw*
σσ́σσ́σσ
σσ́σσ́σσ́σ

Although the trochaic mirror images of these patterns are both unattested, they would be predicted to occur if leftward binary alternation of unstressed and stressed syllables could be perturbed at the left edge in order to avoid initial stress.

Among the attested binary patterns in general, final stress avoidance is often a reason to perturb binary alternation, but initial stress avoidance is not. Including a principle of initial stress avoidance in the grammar, then, would only result in the prediction of unattested patterns.

The only requirement for initial syllables that produces attested patterns is a requirement that initial syllables be stressed. For example, in the trochaic Passamaquoddy (LeSourd 1993) and Garawa (Furby 1974) patterns, an initial stress requirement perturbs leftward binary alternation at the left edge.

(40) a. *Passamaquoddy*
σ́σσ́σσ́σ
σ́σ́σσ́σσ́σ

b. *Unattested*
σσ́σσ́σσ́
σσ́σσ́σσ́σ́

c. *Garawa*
σ́σσ́σσ́σ
σ́σσσ́σσ́σ

d. *Unattested*
σσ́σσ́σσ́
σσ́σσ́σσσ́

Not coincidentally, given the repulsion of stress by final syllables, the iambic mirror images of these patterns are both unattested.

Both final stresslessness and initial stress, then – the two aspects of the asymmetry suggested by the phonetic and rhythmic considerations discussed above – are confirmed by the typology of binary stress patterns.

7.3 *Potential counterexamples*

While the vast majority of extrametricality and non-finality effects have been found at the right edges of prosodic domains, a few languages have been argued to exhibit extrametricality effects at the left edge. In most such cases, however, alternative analyses are readily available.

Halle and Vergnaud (1987), for example, attribute the unstressability of initial vowels in Western Aranda to initial segment extrametricality. Subsequent research, however, has resulted in a number of alternative analyses of Western Aranda and similar languages, analyses that do not require initial extrametricality or non-initiality. Typically, they require the left edge of an appropriate prosodic structure to align with a consonant, preventing initial vowels from being included in that structure and, therefore, from being stressed. In Goedemans (1996), for example, the left edges of feet must align with a consonant. This prevents the initial vowel from being footed and, therefore, from being stressed. In Hyde (2007a), it is the left edges of head syllables that must align with a consonant; in Downing (1998), it is the left edges of prosodic words. Smith's (2002) approach simply requires stressed syllables to have onsets.

As a second example, in the stress patterns of Winnebago (Miner 1979; Hale and White Eagle 1980) and Kashaya (Oswalt 1961, 1988; Buckley 1992) the primary stress in a form is the leftmost stress, but it typically does not appear until the third syllable. Since this ternary interval is characteristic of both even- and odd-parity forms, the most straightforward analysis is to establish a trisyllabic stress window at the left edge of the word. An initial extrametricality approach could establish the stress window by making the initial syllable extrametrical and then constructing a maximally disyllabic foot just to the right of the initial syllable. This is not necessarily strong evidence for initial extrametricality, however. As mentioned in §6.1, there are alternative approaches to trisyllabic stress windows in the literature, some of them addressing a greater variety of windows than is possible with extrametricality.

8 Summary

Extrametricality and non-finality have much in common. Both deal with peripheral positions in a domain, both deal primarily with the right edge of the domain, and both often result in final stresslessness. An important difference, however, is that extrametricality focuses on constituent parsability, while non-finality focuses on the position of stress peaks. Extrametricality rules typically prevent some domain-final constituent from being parsed into higher prosodic structure; non-finality constraints typically prevent a stress peak from occurring in some domain-final position. While they have been used to address many of the same phenomena, the difference in focus ensures that they do not address all types with equal success.

In §2 and §3, we saw that extrametricality and non-finality provide equally effective analyses of situations where stress avoids larger final constituents like syllables and feet. In situations where stress is avoided on final syllables, an expected final stress either arrives early or is absent altogether. An extrametricality analysis achieves the desired effect by excluding the final syllable from the foot layer, a non-finality analysis by prohibiting head syllables in final position or by prohibiting foot-level gridmarks over final syllables. In situations where primary stress avoids final feet, the primary stress emerges as the penultimate stress. An extrametricality analysis excludes the final foot from the prosodic word; non-finality either prohibits head feet in final position or prohibits prosodic-word level gridmarks over final feet.

In contrast, extrametricality and non-finality do not perform equally well in accounting for phenomena involving smaller final constituents. In §4, we saw how the avoidance of stress on word-final moras makes stress sensitive to the weight of word-final syllables, how the avoidance of stress on foot-final moras results in iambic lengthening, and how the avoidance of stress on syllable-final moras results in general weight-sensitivity, iambic lengthening, and trochaic lengthening. In these cases, a non-finality analysis is much more straightforward than an extrametricality analysis. With its stress peaks focus, non-finality can prohibit stress on domain-final moras directly. With its parsability focus, however, extrametricality can only prohibit stress on domain-final moras by excluding them from some higher prosodic structure, a requirement that seems impossible to implement without either violating syllable integrity or requiring moras to remain unsyllabified.

In §5, we saw how the failure of final consonants to contribute to syllable weight affects the stressability of final syllables. Extrametricality achieves the desired result directly by making final consonants invisible to mora assignment. A grid-based non-finality approach achieves the desired result indirectly by prohibiting mora-level gridmarks – and, thus, the moras associated with them – from occurring over final consonants. A head-based non-finality approach, however, appears to be unable to capture the effect at all.

In §6, we examined some of the classic arguments for extrametricality, focusing on trisyllabic stress windows and segmental licensing, and we considered the possibility of non-finality approaches. While head-based non-finality offers analyses for some types of trisyllabic windows, grid-based non-finality does not. Recent alternative proposals for a general approach to stress windows, however, make non-finality's limitations in this area less problematic. With respect to segmental licensing, it is not clear that a non-finality approach is even possible.

Finally, §7 outlined the evidence for the edge asymmetry in extrametricality and non-finality formulations. First, the types of effects analyzable in terms of extrametricality or non-finality occur almost exclusively at right edges. Second, phonetic and rhythmic considerations motivate stresslessness in final positions, but they actually motivate stress in initial position. Third, the inclusion of initial extrametricality or non-initiality in the grammar negatively impacts the accuracy of typological predictions.

REFERENCES

Abbott, Stan. 1985. A tentative multilevel multiunit phonological analysis of the Murik language. *Papers in New Guinea Linguistics* 22. 339–373.

Altshuler, Daniel. 2009. Quantity-insensitive iambs in Osage. *International Journal of American Linguistics* 75. 365–398.

Archangeli, Diana. 1984. Extrametricality in Yawelmani. *The Linguistic Review* 4. 101–120.

Beckman, Mary E. 1986. *Stress and non-stress accent*. Dordrecht: Foris.

Blanc, Haim. 1970. The Arabic dialect of the Negev Bedouins. *Proceedings of the Israeli Academy of Science and Humanities* 4. 112–150.

Buckley, Eugene. 1992. Theoretical aspects of Kashaya phonology and morphology. Ph.D. dissertation, University of California, Berkeley.

Buller, Barbara, Ernest Buller & Daniel L. Everett. 1993. Stress placement, syllable structure, and minimality in Banawá. *International Journal of American Linguistics* 59. 280–293.

Chafe, Wallace L. 1977. Accent and related phenomena in the Five Nations Iroquois languages. In Larry M. Hyman (ed.) *Studies in stress and accent*, 169–181. Los Angeles: Department of Linguistics, University of Southern California.

Comrie, Bernard. 1976. Irregular stress in Polish and Macedonian. *International Review of Slavic Linguistics* 1. 227–240.

de Lacy, Paul. 1997. Prosodic categorisation. M.A. thesis, University of Auckland. (ROA-236.)

Derbyshire, Desmond C. 1985. *Hixkaryana and linguistic typology*. Dallas: Summer Institute of Linguistics & University of Texas at Arlington.

Downing, Laura J. 1993. Unsyllabified vowels in Aranda. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 29. 171–185.

Downing, Laura J. 1998. On the prosodic misalignment of onsetless syllables. *Natural Language and Linguistic Theory* 16. 1–52.

Everett, Daniel L. 1996. Prosodic levels and constraints in Banawá and Suruwaha. Unpublished ms., University of Pittsburgh (ROA-121).

Everett, Daniel L. 1997. Syllable integrity. *Proceedings of the West Coast Conference on Formal Linguistics* 16. 177–190.

Everett, Daniel L. 2003. Iambic feet in Paumari and the theory of foot structure. *Linguistic Discovery* 2(1). 22–44.

Foster, Michael. 1982. Alternating weak and strong syllables in Cayuga words. *International Journal of American Linguistics* 48. 59–72.

Furby, Christine. 1974. *Garawa phonology*. (*Pacific Linguistics* A37.) Canberra: Australian National University.

Gabrielsson, Alf. 1987. Once again: The theme from Mozart's "Piano sonata in A major" (K. 331). A comparison of five performances. In Alf Gabrielsson (ed.) *Action and perception in rhythm and music*, 81–103. Stockholm: Royal Swedish Academy of Music.

Gabrielsson, Alf. 1993. The complexities of rhythm. In Thomas J. Tighe & W. Jay Dowling (eds.) *Psychology and music: Understanding of melody and rhythm*, 94–120. Hillsdale, NJ: Lawrence Erlbaum.

Goddard, Ives. 1979. *Delaware verbal morphology: A descriptive and comparative study*. New York: Garland.

Goddard, Ives. 1982. The historical phonology of Munsee. *International Journal of American Linguistics* 48. 16–48.

Goedemans, Rob. 1996. An optimality account of onset-sensitive stress in quantity-insensitive languages. *The Linguistic Review* 13. 33–47.

Gordon, Matthew. 2002. A phonetically driven account of syllable weight. *Language* 78. 51–80.

Hale, Kenneth & Josie White Eagle. 1980. A preliminary metrical account of Winnebago accent. *International Journal of American Linguistics* 46. 117–132.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress*. Cambridge, MA: MIT Press.

Hammond, Michael & Emmanuel Dupoux. 1996. Psychophonology. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: Models and methods*, vol. 1, 274–297. Salford: ESRI.

Harris, James W. 1983. *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: MIT Press.

Hayes, Bruce. 1980. A metrical theory of stress rules. Ph.D. dissertation, MIT. Published 1985. New York: Garland.

Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13. 227–276.

Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 11. 429–446.

Hayes, Bruce. 1987. A revised parametric metrical theory. *Papers from the Annual Meeting, North East Linguistics Society* 17. 274–289.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Hercus, Luise A. 1986. *Victorian languages: A late survey.* (*Pacific Linguistics* B77.) Canberra: Australian National University.

Hint, Mati. 1973. *Eesti Keele Sõnafonoloogia I.* Talinn, Estonia: Eeste NSV Teaduste Akadeemia.

Hoff, Berend J. 1968. *The Carib language: Phonology, morphonology, morphology, texts and word index.* The Hague: Martinus Nijhoff.

Hung, Henrietta. 1994. The rhythmic and prosodic organization of edge constituents. Ph.D. dissertation, Brandeis University (ROA-24).

Hyde, Brett. 2002. A restrictive theory of metrical stress. *Phonology* 19. 313–359.

Hyde, Brett. 2003. Nonfinality. Unpublished ms., Washington University in St Louis (ROA-633).

Hyde, Brett. 2006. Towards a uniform account of prominence-sensitive stress. In Eric Baković, Junko Itô & John McCarthy (eds.) *Wondering at the natural fecundity of things: Essays in honor of Alan Prince*, 139–183. Santa Cruz: Linguistics Research Center, University of California, Santa Cruz. http://repositories.cdlib.org/lrc/prince/8.

Hyde, Brett. 2007a. Issues in Banawá prosody: Onset sensitivity, minimal words, and syllable integrity. *Linguistic Inquiry* 38. 239–285.

Hyde, Brett. 2007b. Non-finality and weight-sensitivity. *Phonology* 24. 287–334.

Hyde, Brett. 2008. Alignment continued: distance-sensitivity, order-sensitivity, and the Midpoint Pathology. Unpublished ms., Washington University in St Louis (ROA-998).

Hyde, Brett. 2009. The rhythmic foundations of Initial Gridmark and Nonfinality. *Papers from the Annual Meeting, North East Linguistics Society* 38(1). 397–410.

Itô, Junko. 1986. Syllable theory in prosodic phonology. Ph.D. dissertation, University of Massachusetts, Amherst.

Kager, René. 1993. Alternatives to the iambic–trochaic law. *Natural Language and Linguistic Theory* 11. 381–432.

Kager, René. 1995. Review article of Hayes (1995). *Phonology* 12. 437–464.

Keating, Patricia, Taehong Cho, Cécile Fougeron & Chai-Shune Hsu. 2003. Domain-initial articulatory strengthening in four languages. In John Local, Richard Ogden & Rosalind Temple (eds.) *Phonetic interpretation: Papers in laboratory phonology VI*, 145–163. Cambridge: Cambridge University Press.

Kenstowicz, Michael. 1983. Parametric variation and accent in the Arabic dialects. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 19. 205–213.

Kenstowicz, Michael. 1995. Cyclic *vs.* non-cyclic constraint evaluation. *Phonology* 12. 397–436.

Kenstowicz, Michael & Kamal Abdul-Karim. 1980. Cyclic stress in Levantine Arabic. *Studies in the Linguistic Sciences* 10(2). 55–76.

Knudson, Lyle M. 1975. A natural phonology and morphophonemics of Chimalapa Zoque. *Papers in Linguistics* 8. 283–346.

LeSourd, Philip S. 1993. *Accent and syllable structure in Passamaquoddy.* New York: Garland.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.

Lieberman, Philip. 1960. Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America* 32. 451–454.

Lorentz, Ove. 1996. Length and correspondence in Scandinavian. *Nordlyd* 24. 111–128.

Lunden, Anya. 2007. Weight, final lengthening and stress: A phonetic and phonological case study of Norwegian. Ph.D. dissertation, University of California, Santa Cruz (ROA-833).

McArthur, Henry & Lucille McArthur. 1956. Aguacatec (Mayan) phonemes within the stress group. *International Journal of American Linguistics* 22. 72–76.

McCarthy, John J. 1979. On stress and syllabification. *Linguistic Inquiry* 10. 443–465.

McCarthy, John J. 2003. OT constraints are categorical. *Phonology* 20. 75–138.

McCarthy, John J. & Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. Unpublished ms., University of Massachusetts, Amherst & Rutgers University.

Michelson, Karin. 1988. *A comparative study of Lake-Iroquoian accent*. Dordrecht: Kluwer.

Miner, Kenneth L. 1979. Dorsey's Law in Winnebago-Chiwere and Winnebago accent. *International Journal of American Linguistics* 45. 25–33.

Mohanan, K. P. 1979. Word stress in Hindi, Malayalam, and Sindhi. Paper presented at MIT, as reported by Hayes (1980).

Nicklas, Thurston Dale. 1972. The elements of Choctaw. Ph.D. dissertation, University of Michigan, Ann Arbor.

Nicklas, Thurston Dale. 1975. Choctaw morphophonemics. In James M. Crawford (ed.) *Studies in Southeastern Indian Languages*, 237–250. Athens: University of Georgia Press.

Oller, D. K. 1973. The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* 54. 1235–1247.

Oswalt, Robert L. 1961. A Kashaya grammar (Southwestern Pomo). Ph.D. dissertation, University of California, Berkeley.

Oswalt, Robert L. 1988. The floating accent of Kashaya. In William Shipley (ed.) *In honor of Mary Haas*, 611–622. Berlin: Mouton de Gruyter.

Payne, David. 1990. Accent in Aguaruna. In Doris L. Payne (ed.) *Amazonian linguistics: Studies in lowland South American languages*, 161–184. Austin: University of Texas Press.

Payne, David, Judith Payne & Jorge Sanchez Santos. 1982. *Morfología, fonología, y fonética del asheninca del Apurucayali*. Pucallpa, Peru: Instituto Lingüístico de Verano.

Piggott, Glyne L. 1980. *Aspects of Odawa morphophonemics*. New York: Garland.

Piggott, Glyne L. 1983. Extrametricality and Ojibwa stress. *McGill Working Papers in Linguistics* 1. 80–117.

Prince, Alan. 1976. "Applying" stress. Unpublished ms., University of Massachusetts, Amherst.

Prince, Alan. 1980. A metrical theory for Estonian quantity. *Linguistic Inquiry* 11. 511–562.

Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14. 19–100.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Sapir, Edward. 1930. Southern Paiute, a Shoshonean language. *Proceedings of the American Academy of Arts and Sciences* 65. 1–296.

Sapir, Edward. 1949. The psychological reality of phonemes. In David G. Mandelbaum (ed.) *Selected writings of Edward Sapir in language, culture, and personality*, 46–60. Berkeley: University of California Press.

Sapir, J. David. 1965. *A grammar of Diola-Fogny*. Cambridge: Cambridge University Press.

Smith, Jennifer. 2002. Phonological augmentation in prominent positions. Ph.D. dissertation, University of Massachusetts, Amherst.

Steriade, Donca. 1988. Greek accent: A case for preserving structure. *Linguistic Inquiry* 19. 271–314.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91. 1707–1717.

44 The Iambic–Trochaic Law

Brett Hyde

1 Introduction

In the development of metrical stress theory, several influential approaches (Hayes 1985, 1987, 1995; McCarthy and Prince 1986; Prince 1990) have employed the Iambic–Trochaic Law (ITL) to provide extralinguistic grounding for an account of the differences between iambic and trochaic stress systems (see also CHAPTER 39: STRESS: PHONOTACTIC AND PHONETIC EVIDENCE; CHAPTER 40: THE FOOT; CHAPTER 41: THE REPRESENTATION OF WORD STRESS). The ITL, given in (1), is a statement about the naturalness of two types of rhythmic groupings in two different contexts. According to the ITL, sequences of elements that contrast in intensity most naturally divide into groups with trochaic prominence, and sequences of elements that contrast in duration most naturally divide into groups with iambic prominence.

(1) *The Iambic–Trochaic Law* (Hayes 1985, 1987)

a. Elements contrasting in intensity naturally form groupings with initial prominence.
b. Elements contrasting in duration naturally form groupings with final prominence.

For approaches to metrical stress based on the ITL, this difference in naturalness is responsible for the duration-related differences found in iambic and trochaic stress patterns.

The ITL is based on a long tradition of experimental investigation into the perception of rhythmic grouping (Bolton 1894; Woodrow 1909). In the typical experiment, participants are asked to group a sequence of artificially created alternating sounds. The sounds alternate either in intensity, as in (2a), or in duration, as in (2b).

(2) a. . . . o O o O o O o O o O o O o O o O o O o . . .
b. . . . – — – — – — – — – — – — – — – — – — – . . .

The outcome, under certain circumstances, is that participants tend to divide intensity alternations into groups where the more intense element appears first, as in (3a), and they tend to divide duration alternations into groups where the longer element appears second, as in (3b).[1] The ITL is essentially a statement of these results.

(3) a. *Intensity contrasts: Left-prominent groupings*
. . . [O o][O o][O o][O o][O o][O o][O o][O o][O o] . . .

b. *Duration contrasts: Right-prominent groupings*
. . . [– —][– —][– —][– —][– —][– —][– —][– —] . . .

Though the ITL is an extralinguistic principle, it seems to be reflected in the stress patterns of numerous languages, suggesting, at least initially, that it plays an important role in shaping them. For example, many trochaic languages are like Cahuilla (Seiler 1965, 1967, 1977; Seiler and Hioki 1979). They exclude heavy syllables from disyllabic feet, ensuring that durational contrasts never arise in a foot with trochaic prominence. In Cahuilla, heavy syllables are CVV and CVʔ.

(4) *Exclusion of H from disyllabic feet in Cahuilla*

('LL)(ˌL)	('taxmu)(ˌʔat)	'song'
('LL)(ˌLL)	('taka)(ˌliʧem)	'one-eyed ones'
('H)(ˌL)	('paʔ)(ˌli)	'the water (OBJ)'
('H)(ˌLL)	('qaː)(ˌniʧem)	'palo verde (PL)'
('L)(ˌH)(ˌL)	('su)(ˌkaʔ)(ˌti)	'the deer (OBJ)'
('L)(ˌH)(ˌLL)	(ˌnesun) ('ka)(ˌviː)(ˌʧi-wen)	'I was surprised'

Many iambic languages are like Hixkaryana (Derbyshire 1985). They lengthen the vowel of stressed syllables, if necessary, to ensure that feet with iambic prominence always contain durational contrasts. In Hixkaryana, heavy syllables are CVV and CVC.

(5) *Iambic lengthening in Hixkaryana*

a.	(L'L)('H)L	→	(L'H)('H)L	
	(khæ'næ)('nɯh)nɔ	→	(khæ'næː)('nɯh)nɔ	'I taught you'
b.	(L'L)(L'H)L	→	(L'H)(L'H)L	
	(mɯ'hæ)(næ'nɯh)nɔ	→	(mɯ'hæː)(næ'nɯh)nɔ	'you taught him'
c.	('H)(L'L)L	→	('H)(L'H)L	
	('ɔw)(tɔ'hɔ)næ	→	('ɔw)(tɔ'hɔː)næ	'to the village'
d.	('H)(L'L)LL	→	('H)(L'H)LL	
	('tɔh)(ku'rʲi)hɔnæ	→	('tɔh)(ku'rʲiː)hɔnæ	'to Tohkurye'

This chapter reviews the strengths and weaknesses of ITL approaches to metrical stress, and examines some of the most promising alternatives. We shall see that the ITL does not actually offer an adequate foundation for an account

[1] The effect emerges in the range of one half to five beats per second. (The syllable rate of "ordinary conversational speech" is typically toward the upper limits of this range; Bell 1977.) Hayes (1995) states that the right-prominent effect illustrated in (3b) requires a durational contrast where the longer elements are 1.5 to 2 times as long as the shorter elements, noting that Woodrow (1909) found that smaller durational contrasts actually result in left-prominent groupings.

of stress systems in general, but it may provide an adequate foundation for an account of quantity-sensitive stress systems in particular (see CHAPTER 57: QUANTITY-SENSITIVITY). This is not to say that it provides the best foundation. There is a clear sense in which the superficial and descriptive ITL is itself an observation in need of an explanation, much like the stress patterns found in natural language. Part of the appeal of the most promising alternatives is that they have the potential to account not only for the stress patterns of natural language, but also for the ITL itself.

Before reviewing the various proposals, we should note the results of more recent investigations into the perception of rhythmic grouping. In some cases, more recent studies have confirmed the grouping preferences found in the earlier studies on which the ITL was based. In other cases, they have challenged their universality. The studies of Rice (1992), Vos (1977), and Hay and Diehl (2007), for example, found grouping preferences among English, French, and Dutch speakers similar to those found in the earlier studies of Bolton (1894) and Woodrow (1909).[2] The studies of Kusumoto and Moreton (1997) and Iversen *et al.* (2008), however, found significant differences between speakers of English and Japanese. Iversen *et al.*, for example, found that English speakers had a fairly strong preference (68 percent) for dividing sequences of amplitude contrasts into trochaic (loud–soft) groups, but Japanese speakers had a much stronger preference (91 percent) for trochaic grouping. English speakers showed a very strong preference (89 percent) to divide duration contrasts into iambic (short–long) groups, but Japanese speakers showed no preference. While the challenge to universality may be troubling to those particularly concerned with extralinguistic grounding, and it certainly presents an interesting problem in this connection, it does not necessarily tell us anything about the ITL's ability to predict differences between iambic and trochaic stress patterns in language. Having noted the problem with respect to extralinguistic grounding, then, I will not address the issue further.

2 Interpretations of the ITL

The most recent ITL accounts (McCarthy and Prince 1986; Hayes 1987, 1995; Prince 1990) reflect two distinct interpretations. The stronger of the two, given in (6), takes the ITL to be concerned with the actual presence or absence of durational contrasts within rhythmic groupings.

(6) *Strong interpretation of the ITL*

a. If a foot contains a durational contrast, it is iambic.
b. If a foot lacks a durational contrast, it is trochaic.

The weaker interpretation, given in (7), takes the ITL to be concerned with sensitivity to the positions of the heavy syllables that might help to create durational contrasts.

[2] Rice's study also found a preference for iambic grouping when elements contrasted in pitch, a result not found in previous studies.

(7) *Weak interpretation of the ITL* (Hayes 1985)

a. If parsing is sensitive to the position of heavy syllables, it is iambic.
b. If parsing is insensitive to the position of heavy syllables, it is trochaic.

Even at the point at which the ITL was introduced to metrical stress theory, it was clear that the strong interpretation in (6) was unsustainable, at least when applied to stress systems generally. Under the strong interpretation, iambic footing and the presence of durational contrasts are intimately connected: only iambs contain durational contrasts; durational contrasts arise only in iambic feet; and only iambic systems employ rules that create durational contrasts. Similarly, trochaic footing and the absence of durational contrasts are intimately connected: only trochees lack durational contrasts; durational contrasts are absent only in trochaic feet; and only trochaic systems employ rules that destroy durational contrasts.

Even a cursory look at the general typology of attested stress patterns reveals that the strong interpretation misses the mark by a wide margin. As mentioned above, many iambic languages are like Hixkaryana, lengthening the vowel of stressed syllables, if necessary, to ensure that surface iambs contain durational contrasts. Many other iambic languages are like Araucanian (Echeverria and Contreras 1965), however. They tolerate surface iambs that have no durational contrasts.

(8) *Even iambs in Araucanian*

(L'L)	(wu'le)	'tomorrow'
(L'L)L	(ti'pan)to	'year'
(L'L)(L,L)	(e'lu)(mu,ju)	'give us'
(L'L)(L,L)L	(e'lu)(a,e)new	'he will give me'

A similar situation obtains with trochaic languages. As mentioned above, several are like Cahuilla in prohibiting foot-internal durational contrasts. Several others, however, are like Chimalapa Zoque (Knudson 1975). They tolerate foot-internal durational contrasts and even have rules that create them. In Chimalapa Zoque, heavy syllables are CVV and CVC.

(9) *Trochaic lengthening in Chimalapa Zoque*

a.	('LH)	→	('HH)	
	('kosaʔ)	→	('koːsaʔ)	'scold (IMP)'
b.	(,L)('LL)	→	(,H)('HL)	
	(,hu)('kuti)	→	(,huː)('kuːti)	'fire'
c.	(,LL)L('LL)	→	(,HL)L('HL)	
	(,witi) hu('kuti)	→	(,wiːti) hu('kuːti)	'big fire'
d.	(,LH)H('HL)	→	(,HH)H('HL)	
	(,wituʔ)paj('niksɨ)	→	(,wiːtuʔ)paj('niksi)	'he is coming and going'

There appears to be no close connection between iambs and the presence of durational contrasts, then, or between trochees and the absence of durational contrasts, at least in the general case.

Given the shortcomings of the strong interpretation, Hayes (1985) introduced the ITL to metrical stress theory under the weak interpretation in (7). Under the weak interpretation, the crucial connections are between iambic footing and quantity-sensitivity and trochaic footing and quantity-insensitivity. While iambic feet and trochaic feet might both contain durational contrasts, parsing is iambic if and only if it is sensitive to the positions of heavy syllables. Parsing is trochaic if and only if it is insensitive to the positions of heavy syllables.

There are three problems with the weak interpretation. The first is conceptual. The ITL is plainly a generalization about the appropriateness of durational contrasts within two different types of feet. Since its requirements concerning durational contrasts affect both types, the ITL does not countenance situations where either type is quantity-insensitive (where either type simply ignores the differences in syllable weight that help to create durational contrasts). In viewing the primary concern of the ITL to be the appropriateness of quantity-sensitivity for different types of feet, the weak interpretation seems really to be a misinterpretation.

The second problem is a loss of empirical coverage. Since it only addresses quantity-sensitivity, the weak interpretation tells us nothing about the status of lengthening and shortening rules addressed by the strong interpretation.

The final problem is that the weak interpretation is false. A significant number of trochaic systems are quantity-sensitive, falsifying (7a), and a significant number of iambic systems are quantity-insensitive, falsifying (7b). In (4), for example, we saw that heavy syllables consistently perturb the basic stress pattern of the trochaic Cahuilla, indicating that it is quantity-sensitive. In (10), we see that heavy syllables consistently fail to perturb the basic pattern of the iambic Paumari (Everett 2003), indicating that it is quantity-insensitive. In the basic pattern, stress appears on every odd-numbered syllable from the right. CVV syllables are heavy.

(10) *Quantity-insensitive iambs in Paumari*

(ˌL)(LˈL)	(ˌma)(siˈko)	'moon'
(LˌL)(LˈL)	(kaˌʤo)(wiˈri)	'island'
(ˌH)(HˈL)	(ˌkai)(haiˈhi)	'type of medicine'
(HˌL)(LˈL)	(waiˌʧa)(naˈwa)	'little ones'

Additional quantity-sensitive trochaic languages include Palestinian Arabic (Brame 1973, 1974; Kenstowicz and Abdul-Karim 1980; Kenstowicz 1983) and Fijian (Schütz 1978, 1985; Dixon 1988). Additional quantity-insensitive iambic languages include Araucanian, Osage (Altshuler 2009), Suruwaha (Everett 1996), and Weri (Boxwell and Boxwell 1966).

As we shall see in §3, parts of both interpretations, (6a) of the strong interpretation and (7b) of the weak interpretation, are brought together to form the basis for two subsequent ITL accounts, those of Hayes (1987, 1995) and of McCarthy and Prince (1986). This marriage between the halves of two very different interpretations often makes the connection between the ITL and the phenomena that these approaches attempt to account for less than clear. This is part of the reason, perhaps, that some have concluded that there is actually little of the ITL left in ITL-based approaches (see van der Hulst 1999, for example). A third ITL approach, that of Prince (1990), employs only the strong interpretation, but seeks to avoid the problems discussed above by employing it only in the context of quantity-sensitive systems and only as a relative "preference" rather than an absolute "law."

Though I will point out the aspects of the more recent ITL accounts that derive from the weak interpretation, it should be clear at this point that the weak

The reason is simply that the foot that is constructed immediately after the heavy syllable is parsed, rather than the foot that is constructed to parse the heavy syllable itself, determines how the basic alternation resumes. Whether heavy syllables are included in disyllabic feet or parsed as monosyllabic feet, the basic alternations of both iambic and trochaic systems would resume with an unstressed syllable (the underlined syllable in the examples below).

(12) *Parsing directionality matches headedness*

a. *Left-to-right iambic*

```
i. Iamb                     ii. Monosyllable
     (   x)(   x)                (x)(   x)
   ...L  H  L  L...            ...H  L  L...
   ---------------->           ---------------->
```

b. *Right-to-left trochaic*

```
i. Trochee                  ii. Monosyllable
     (x   )(x   )                (x   )(x)
   ...L  L  H  L...            ...L  L  H...
   <----------------           <----------------
```

In left-to-right iambic systems, the heavy syllable must occur at the right edge of a foot whether the foot is an iamb, as in (12a.i), or a monosyllable, as in (12a.ii). Since the next foot constructed would be iambic in either case, the alternation resumes with an unstressed syllable. In right-to-left trochaic systems, the heavy syllable would be parsed at the left edge of a foot whether the foot is a trochee, as in (12b.i), or a monosyllable, as in (12b.ii). Since the next foot constructed would be trochaic in either case, the alternation again resumes with an unstressed syllable.

A difference in the resumption of basic alternations emerges only in situations where parsing directionality and the headedness of the foot do not match. In right-to-left iambic languages like Tübatulabal (Voegelin 1935), as (13) illustrates, heavy syllables are always preceded by stressless syllables. In Tübatulabal, heavy syllables are CVV(C).

(13) *Resumption with a stressless syllable in Tübatulabal*

'LL'L	'ʧiŋi'jal	'the red thistle'
L'LL'H	ti'ŋija'laap	'on the red thistle'
L'LL'L	wi'taŋha'tal	'the Tejon Indians'
'LL'LL'HL'L	'witaŋ'hata'laaba'tsu	'away from the Tejon Indians'
'H'LL'L	'taa'hawi'la	'the summer'
'H'LL'H	'taa'hawi'laap	'in the summer'

In left-to-right trochaic languages like Cahuilla, however, as illustrated in (4), heavy syllables are always followed by stressed syllables. The difference between right-to-left iambic systems and left-to-right trochaic systems, then, is that the former resume their basic alternations with stressless syllables while the latter resume them with stressed syllables.

The reason that a difference emerges when headedness and parsing directionality do not match is that the resumption of basic alternations depends directly on how the heavy syllable itself is footed.

existence of several quantity-insensitive iambic languages. While Hayes argues that the quantity-insensitivity of such systems is only apparent, as they do not actually contain heavy syllables to perturb the basic pattern, the argument is plausible only in the cases of Araucanian and Weri. It is not plausible in the cases of Osage, Paumari, and Suruwaha, each of which has long vowels, diphthongs, or both.

Given that both iambic and trochaic systems can be quantity-sensitive, clause (6a) of the strong interpretation, "if a foot contains a durational contrast, it is iambic," motivates a disparity in precisely how the two types can be quantity-sensitive. As (15b) indicates, Hayes's account requires trochaic systems to deal with heavy syllables differently than iambic systems. Iambs allow heavy syllables in strong position in disyllabic feet, where they can create durational contrasts. They exclude them only from weak position. Trochees, however, exclude heavy syllables from disyllabic feet entirely. Trochaic systems must parse heavy syllables into monosyllabic feet, where no durational contrast is possible.

The disparity in how the two types of feet can be quantity-sensitive predicts the difference, discussed above, in how right-to-left iambic languages and left-to-right trochaic languages resume basic stress alternations after encountering a heavy syllable. The fact that iambic systems parse heavy syllables into disyllabic feet in Hayes's account correctly predicts that right-to-left iambic languages will resume their basic alternation with a stressless syllable, as in (14a.i). The fact that trochaic systems must parse heavy syllables into monosyllabic feet correctly predicts that left-to-right trochaic languages will resume their basic alternation with a stressed syllable, as in (14b.ii).

McCarthy and Prince (1986) arrive at a foot inventory similar to Hayes's, but they arrive at it through a slightly different route and in service of a different purpose. They posit one type of quantity-insensitive foot: the balanced [σ σ] template, and two types of quantity-sensitive feet: the balanced [μ μ] template and the unbalanced [σ_μ $\sigma_{\mu\mu}$] template.

(16) a. *Quantity-insensitive*
   Balanced [σ σ]

b. *Quantity-sensitive*
   i. Balanced [μ μ]
   ii. Unbalanced [σ_μ $\sigma_{\mu\mu}$]

The ITL contributes to McCarthy and Prince's account in two ways. First, clause (6a) of the strong interpretation, "if a foot contains a durational contrast, it is iambic," motivates the iambic configuration of the quantitatively unbalanced foot. To guarantee that quantitatively iambic feet are also iambic with respect to stress, they posit the Quantity/Prominence Homology principle. It ensures that the heavier syllable in feet with a quantity contrast – in effect, the heavy syllable in a [σ_μ $\sigma_{\mu\mu}$] foot, given the limited possibilities in (16) – bears the stress.

(17) *Quantity/Prominence Homology*

For a, b ∈ F, if a > b quantitatively, then a > b stresswise.

Prominence in unbalanced feet is determined by the Trochaic Default principle, which ensures that [σ σ] and [μ μ] feet both stress their initial syllable.

(22)	*satisfy:*	IQ/TQ, Binarity		Binarity only		neither
a.	Iambic	[L H]	>	[L L], [H]	>	[L]
b.	Trochaic	[L L], [H]	>	[H L]	>	[L]

The asymmetry in this case arises in how the iambic and trochaic hierarchies order balanced and unbalanced feet. Iambs prefer unbalanced [L H] to balanced [L L] and [H], but trochees prefer balanced [L L] and [H] to unbalanced [H L].

The main difference between the inventory of quantity-sensitive feet in Prince's and Hayes's accounts is the possibility of unbalanced [H L] trochees. As Prince notes, however, with their lesser status, there are limited situations in which unbalanced trochees might arise. First, consider the result of Harmonic Parsing in a left-to-right trochaic system. Since feet are constructed serially, and parsing the next syllable in line is the overriding concern, Harmonic Parsing would parse the H of an HL sequence as monosyllabic foot, just like Hayes's moraic trochees. If the following L could not combine with another L to form a disyllabic foot, giving [H][LL], it might be parsed as a monosyllable, giving [H][L], or left unparsed, giving [H]L, depending on whether or not degenerate feet are tolerated. These are the same options available under Hayes's moraic trochees. The results are rather different in right-to-left systems. Harmonic Parsing would always parse an HL sequence into an [HL] foot, but Hayes's moraic trochees would yield either [H][L] or [H]L, depending on whether or not degenerate feet are tolerated. The latter option results in the same stress pattern, but the former does not. I am not aware, however, of a right-to-left trochaic language that would allow us to distinguish between the two approaches.

3.2 *A symmetric foot inventory*

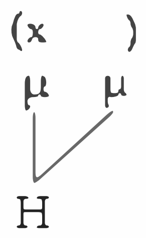
As discussed above, the evidence for a difference between iambic and trochaic quantity-sensitivity comes from systems where parsing directionality is opposite the headedness of the foot. After left-to-right trochaic parsing encounters a heavy syllable, binary alternation resumes with a stressed syllable. In contrast, after right-to-left iambic parsing encounters a heavy syllable, binary alternation resumes with a stressless syllable. Where ITL approaches posit an asymmetric foot inventory to account for this difference, Kager (1993) proposes a symmetric foot inventory, arguing that the difference is best explained in terms of the metrical principles of clash and lapse avoidance.

Kager distinguishes between parsing feet and the surface feet that can be formed later through adjunction of unparsed syllables. The inventory of parsing feet is symmetric. The quantity-insensitive syllabic trochee corresponds to a mirror-image syllabic iamb. The quantity-sensitive moraic trochee corresponds to a mirror-image moraic iamb. Iambic quantity-sensitivity and trochaic quantity-sensitivity are identical, then, in that both exclude heavy syllables from disyllabic feet.

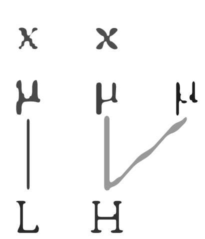
(23) *Parsing feet*

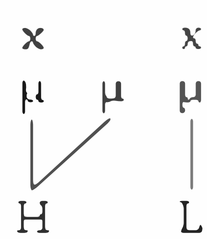
	trochaic	*iambic*
	(x)	(x)
Syllabic (quantity-insensitive)	σ σ	σ σ
	(x)	(x)
Moraic (quantity-sensitive)	μ μ	μ μ

Crucial to Kager's account is the claim that heavy syllables contain an internal prominence contrast corresponding to a decline in sonority between the first and second mora. According to Kager, the internal contrast is the characteristic of heavy syllables responsible for their attraction of stress (Prince's 1990 Weight-to-Stress principle). The decline in sonority ensures that stress occurs over the first mora and that the second mora is stressless, as in (24).

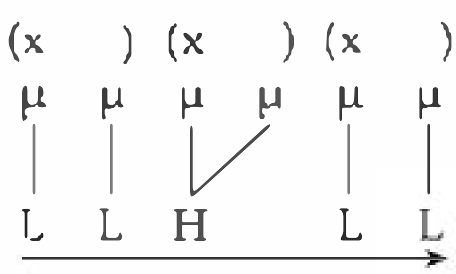
(24)


A heavy syllable's strong–weak contour translates into different results with respect to clash at the mora level. When a stressed heavy syllable immediately follows another stressed syllable, as in (25a), the result is a clash. When the order is reversed, as in (25b), there is no clash.

(25) a. *Clash*


b. *No clash*


Assuming that clash is never tolerated at the point the basic alternation is resumed, the internal prominence contrast accounts for the different modes of resumption after a heavy syllable.[4] In left-to-right trochaic systems, a trochaic foot can immediately follow the heavy syllable without creating clash, so the pattern resumes with a stressed syllable.

(26) *Left-to-right trochees: No clash*


In right-to-left iambic systems, however, an iambic foot cannot immediately precede the heavy syllable without creating clash. This being the case, the parsing algorithm must skip a syllable before constructing an iambic foot, and the pattern resumes with an unstressed syllable.

[4] More precisely, Kager assumes that the construction of a foot cannot introduce clash within the parsing window. The parsing window consists of the syllables being parsed in the current iteration plus the string of syllables encountered by the parsing algorithm in previous iterations. It does not include syllables that the algorithm has not yet encountered.

pattern. Rhythmic lengthening appears to be based solely on the alternation of strong and weak positions, affecting only the former. In contrast, rhythmic shortening can affect both strong and weak positions and seems in many cases to be motivated, at least partially, by a preference for exhaustive parsing.

Rhythmic lengthening increases the duration of stressed syllables either through vowel lengthening (CHAPTER 20: THE REPRESENTATION OF VOWEL LENGTH) or gemination of an adjacent consonant (CHAPTER 37: GEMINATES), the former method being more common than the latter. It can be found in both iambic and trochaic languages. In (5), for example, we saw that the stressed vowels of underlyingly light syllables lengthen in the iambic Hixkaryana, making them heavy on the surface. Other iambic lengthening languages include Carib, Choctaw, and several varieties of Yupik (Woodbury 1981, 1987; Jacobson 1984, 1985; Krauss 1985a; Leer 1985; among others). In (9), we saw that the stressed vowels of underlyingly light syllables lengthen in the trochaic Chimalapa Zoque. Other trochaic lengthening languages include Chamorro (Topping and Dungca 1973; Chung 1983), Icelandic (Árnason 1980, 1985), Mohawk (Michelson 1988), and Selayarese (Mithun and Basri 1986).

An interesting difference between iambic and trochaic lengthening is that lengthening occurs in iambic systems only when they are quantity-sensitive, and in trochaic systems only when they are quantity-insensitive.[5] When it is seen as shaping the possibilities of stress patterns generally, then, the existence of trochaic lengthening in quantity-insensitive systems clearly undermines the ITL. If we restrict the ITL's scope to quantity-sensitive systems, however, the distribution of lengthening gives it considerable support. The presence of lengthening in iambic languages, where durational contrasts are encouraged, is consistent with the ITL, as is the absence of lengthening in trochaic languages, where durational contrasts are prohibited.

Another important generalization concerning rhythmic lengthening is the correlation between what I will refer to as *regular lengthening* and certain types of minimal words. *Regular lengthening* is the exceptionless lengthening in non-minimal forms characteristic of many lengthening languages: vowels lengthen in underlyingly light syllables whenever they receive the appropriate degree of stress.[6] As (32) indicates, languages with regular lengthening allow only three types of minimal word: H, LL, and HL.

(32) *Minimal words associated with regular lengthening*

a. *Monosyllabic*		b. *Disyllabic*	
L	*unattested*	LL	Choctaw verbs (iambic)
H	Chimalapa Zoque (trochaic)	HL	Carib (iambic)
	Choctaw nouns (iambic)		Hixkaryana (iambic)
	Icelandic (trochaic)		Selayarese (trochaic)
	Yupik varieties (iambic)		
		LH	*unattested*

[5] The clearest cases of quantity-insensitive iambs – Osage, Paumarí, and Suruwaha – do not exhibit lengthening. The less clear cases – Weri and Araucanian – also do not exhibit lengthening.

[6] Lengthening is not "regular" when it is prohibited in various positions in non-minimal forms, especially in final position. Syllables with primary stress in Italian, for example, lengthen if they are penultimate but not if they are antepenultimate or final. Languages like Unami and Munsee Delaware (Goddard 1979), which make stressed syllables heavy through consonant gemination, also fall outside the generalization.

Iambic lengthening languages and trochaic lengthening languages can both insist on H or HL minimal words, but only iambic lengthening languages can insist on an LL minimal word. There appear to be no regular lengthening languages with L minimal words, and none with LH minimal words.

If we exclude alternations better described as vowel reduction (see below), rhythmic shortening is a marginal phenomenon. It occurs in only a few trochaic systems, and, as Mellander (2003) points out, each of these few is quantity-sensitive. Trochaic shortening can affect either a stressed syllable or an unstressed syllable. In Fijian, for example, stressed syllables shorten, converting HL sequences into LL sequences.

(33) *Trochaic shortening in Fijian*

ŋbuː-rŋu	→	ˈmbu-rŋu	'my grandmother'
θaː-j-a	→	ˈta-j-a	'chop-TRANS-3 SG OBJ'
nreː-ta	→	ˈnre-ta	'pull-TRANS'

In Pre-Classical Latin (Allen 1973; Mester 1994), stressless syllables shorten, converting LH sequences into LL sequences.[7]

(34) *Trochaic shortening in Latin*

egoː	→	ˈego	'I'
maleː	→	ˈmale	'bad'
amiːkitiam	→	ˌamiˈkitiam	'friendship'

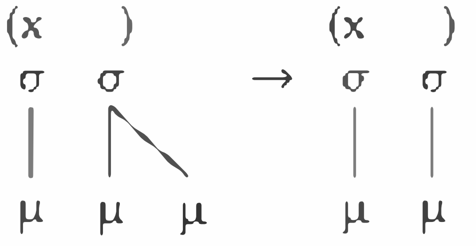
Rhythmic shortening, though marginally attested, is consistent with the ITL. Among quantity-sensitive languages, it occurs only in trochaic systems, destroying the durational contrasts that the ITL prohibits. It does not occur in iambic systems, where it would destroy durational contrasts that the ITL requires. Quantity-insensitive languages, of either type, apparently do not exhibit rhythmic shortening.

Before we proceed, it should be noted at this point that there are at least two languages with shortening phenomena that are potential counterexamples to the generalizations presented in the preceding paragraph: Central Slovak (Dvonč 1955; Bethin 1998; Mellander 2003) and Gidabal (Geytenbeek and Geytenbeek 1971; Rice 1992; Mellander 2003). The stress patterns of both languages are fairly complex, however, and their analyses are not at all straightforward. It is not clear whether they are examples of shortening in trochaic feet (resulting in unbalanced HL trochees), shortening in iambic feet, or both, or whether they are simply examples of shortening in non-head syllables generally. Since it is unclear exactly how such examples are relevant, I have set them aside here.

I have also set aside phenomena involving vowel strengthening in stressed syllables and vowel reduction and deletion in unstressed syllables. Some of the alternatives to an ITL approach draw to a significant extent on such phenomena as evidence that the difference between iambs and trochees with respect to rhythmic lengthening and shortening is not as great as previously thought (Revithiadou and van de Vijver 1997; van de Vijver 1998; Revithiadou 2004). Strengthening, reduction, and deletion phenomena are fairly common in both iambic and trochaic systems.

[7] The Latin-type shortening is often referred to as *iambic shortening*, because it affects the second syllable in a two-syllable sequence rather than the first.

b. *Latin-type*



The effect of the Latin-type shortening rule in (36b) is to convert an ill-formed (LH) trochee to a well-formed (LL) trochee. An ill-formed (LH) trochee might be created inadvertently, for example, when an extrametrical H syllable is adjoined to a degenerate (L) foot, and (36b) repairs the defect.

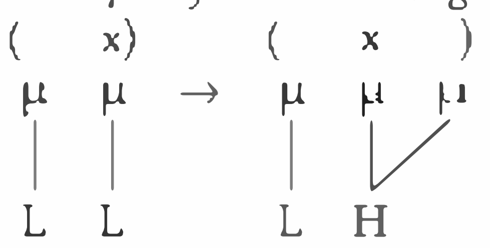
Though the ITL account captures the distribution of rhythmic lengthening and shortening in quantity-sensitive systems, it falls short in two ways. First, it fails to allow for trochaic lengthening in quantity-insensitive systems (a phenomenon whose existence Hayes denies). Second, it does not account for the correlation between regular lengthening and the group of minimal words in (32). Based on the ITL, lengthening rules might be employed to create durational contrasts in iambic feet (or, possibly, to destroy them in trochaic feet), but there is nothing in the law entailing that lengthening languages should prefer H, HL, and LL minimal words above L and LH minimal words. If the ITL is actually the motivation for lengthening, then the correlation of regular lengthening with these particular minimal words is a mystery.

4.2 *Lapse avoidance and non-finality*

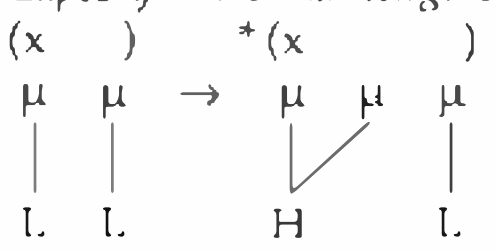
Kager's (1993) approach to the asymmetries in rhythmic lengthening and shortening is based on the same principles that governed his approach to quantity-sensitivity (see §3.2). In conjunction with a prohibition against foot-internal lapse, the internal prominence contrast in heavy syllables restricts the occurrence of lengthening. Kager views lengthening of stressed syllables in general as phonetically motivated, but the restriction against foot-internal lapse ensures that such lengthening is more common in iambs than in trochees. As (37) illustrates, the grammar tolerates lengthening that creates (LH) iambs, because they contain no foot-internal lapse, but it does not tolerate lengthening that creates (HL) trochees, because they do contain a foot-internal lapse.

(37) *Lengthening asymmetry through lapse avoidance*

a. *No lapse after iambic lengthening*



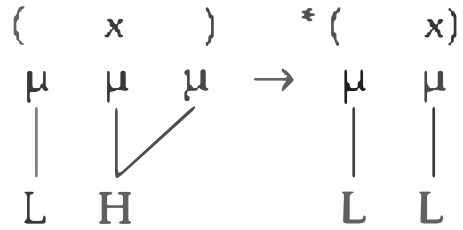
b. *Lapse after trochaic lengthening*



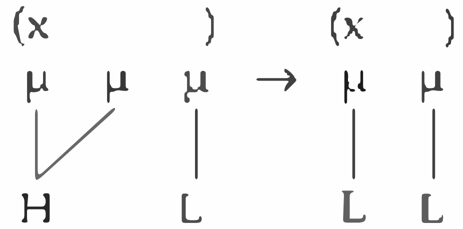
The prohibition against foot-internal lapse also accounts for the shortening asymmetry, but in this case it acts as a trigger. In Kager's view, the purpose of trochaic shortening is to eliminate foot-internal lapses like those found in (HL) trochees. Since there is no foot-internal lapse in (LH) iambs, the motivation for shortening never arises in iambic systems.

(38) *Lapse avoidance predicts shortening asymmetry*

a. *No lapse in iambs to trigger shortening*



b. *Lapse in trochees triggers shortening*



Although foot-internal lapse avoidance effectively addresses the lesser frequency of lengthening in trochaic systems, it does not address the actual phonological triggers for lengthening. Hyde's (2007) non-finality approach addresses the lesser frequency of lengthening in trochaic systems, but it also provides phonological triggers for rhythmic lengthening and addresses its correlation with certain types of minimal words (CHAPTER 43: EXTRAMETRICALITY AND NON-FINALITY).

Under the non-finality approach, rhythmic lengthening is a special case of the type of weight-sensitivity where stress avoids light syllables. To avoid stressing a light syllable, the syllable is lengthened to make it heavy. Non-finality produces this type of weight-sensitivity by prohibiting stress on domain-final moras. Following Kager (1995), Hyde applies non-finality to the foot domain to promote iambic lengthening. Going a step further, he also applies non-finality to the syllable domain. This gives the approach a second mechanism for promoting iambic lengthening but it also gives it a mechanism for promoting trochaic lengthening.

(39) a. NON-FINALITY(Ft)
No stress occurs over the final mora of a foot.

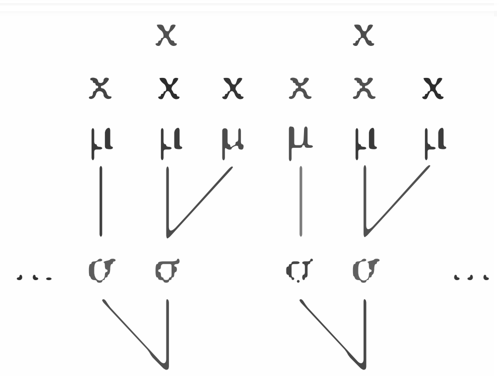
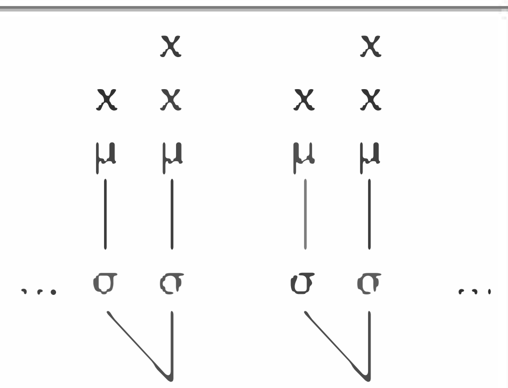
b. NON-FINALITY(σ)
No stress occurs over the final mora of a syllable.

NON-FINALITY(Ft) effectively prohibits stress on light foot-final syllables. Since it bans foot-level gridmarks from foot-final moras, foot-final syllables must be at least bimoraic to support stress. NON-FINALITY(σ) effectively prohibits stress on light syllables generally. Since it bans foot-level gridmarks from syllable-final moras, syllables generally must be at least bimoraic to support stress.

To produce lengthening, one of the non-finality constraints must dominate DEP-μ, the faithfulness constraint that prevents mora insertion. Under such rankings, when stress would otherwise occupy a light syllable, a mora can be added to make the syllable heavy on the surface. The two non-finality constraints do not, however, have equal ability to promote lengthening in every type of foot. Since NON-FINALITY(Ft) prohibits stress over light foot-final syllables in particular, it can lengthen the stressed syllable of an iamb but not the stressed syllable of a trochee. In contrast, since NON-FINALITY(σ) prohibits stress over light syllables in general, it can lengthen the stressed syllables of both.

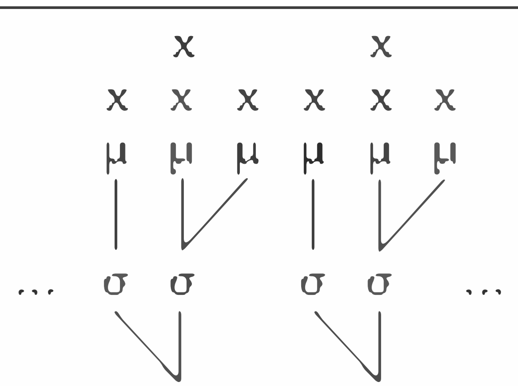
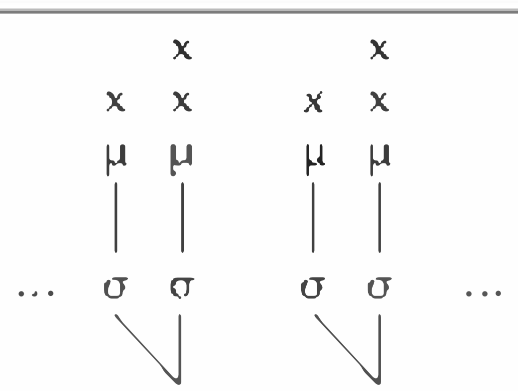
Consider first the situation where the stressed syllable occurs in an iamb. When NON-FINALITY(Ft) dominates DEP-μ, a second mora is added to underlyingly light syllables to avoid stress on foot-final moras.

(40)

LLLL		NON-FIN(Ft)	DEP-μ
☞ a.			**
b.		*!*	

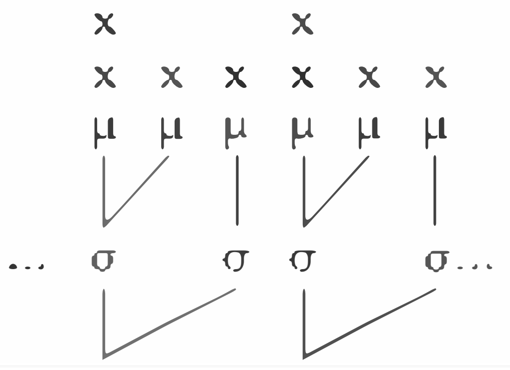
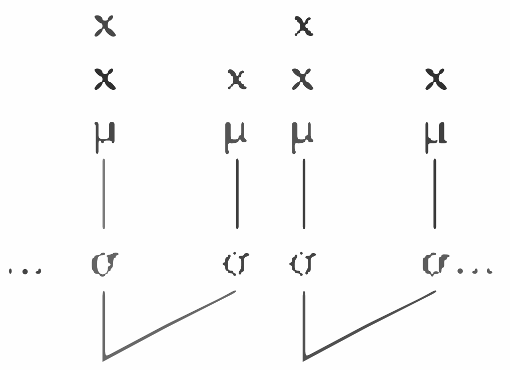
The result is similar when NON-FINALITY(σ) dominates DEP-μ: a second mora is added to the underlyingly light syllables to avoid stress on syllable-final moras.

(41)

LLLL		NON-FIN(σ)	DEP-μ
☞ a.			**
b.		*!*	

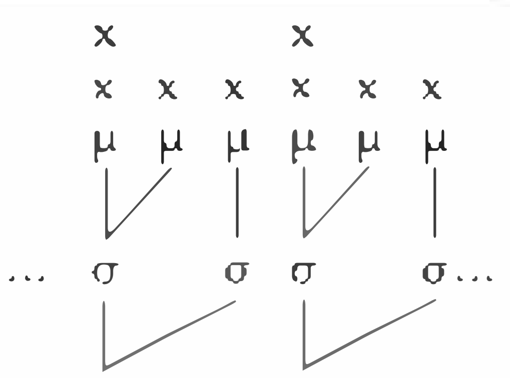
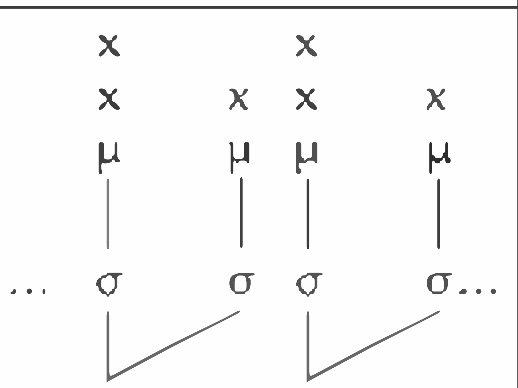
Now consider the situation where the stressed syllable occurs in a trochee. When NON-FINALITY(Ft) dominates DEP-μ, as in (42), there is no lengthening. Because stress does not occupy the foot-final syllables in either candidate, there is no danger that it will occupy the foot-final moras, and NON-FINALITY(Ft) cannot distinguish between them. The lower ranked DEP-μ settles on the faithful (42b) candidate.

(42)

LLLL		NON-FIN(Ft)	DEP-μ
a.			*!*
☞ b.			

When NON-FINALITY(σ) dominates DEP-μ, however, as in (43), the lengthening candidate emerges as the winner. The stressed syllables become heavy, to allow stress to avoid syllable-final moras.

(43)

LLLL		NON-FIN(σ)	DEP-μ
☞ a.			**
b.		*!*	

One advantage of the non-finality approach is that it has a built-in explanation for the lesser frequency of lengthening among trochaic systems. Non-finality in the syllable and non-finality in the foot both produce iambic lengthening, but only non-finality in the syllable produces trochaic lengthening. Every ranking that produces trochaic lengthening, then, also produces iambic lengthening, but some rankings that produce iambic lengthening do not produce trochaic lengthening. Since the percentage of possible rankings that produce iambic lengthening is greater than the percentage of possible rankings that produce trochaic lengthening, we would expect lengthening to occur with greater frequency in iambic systems than it does in trochaic systems, all else being equal.

A second advantage of the non-finality approach is that it helps to account for the particular group of minimal words associated with regular lengthening. As discussed above, languages that automatically lengthen appropriately stressed vowels only allow three types of minimal word: H, LL, and HL. They never allow L or LH minimal words. Using the same non-finality constraints to produce rhythmic lengthening and the minimal word restrictions predicts this situation.

L minimal words are absent, because the lengthening constraints themselves both establish H minimal words. NON-FINALITY(σ) has the same effect in monosyllabic feet that it has in disyllabic feet, and NON-FINALITY(Ft) has the same effect that it has in iambs. They both force the stressed syllable to lengthen. As (44) indicates, if either of the lengthening constraints ranks highly enough to produce lengthening in the disyllabic feet of longer forms, then it also ranks highly enough to produce lengthening in the monosyllabic feet of monosyllabic forms.

(44) a. NON-FINALITY(σ) >> DEP-μ
Iambic or trochaic lengthening + H minimal word
b. NON-FINALITY(Ft) >> DEP-μ
Iambic lengthening + H minimal word

Two desirable predictions result from this situation: regular lengthening is always accompanied by a minimal word that is at least bimoraic, and iambic lengthening languages and trochaic lengthening languages can both have H minimal words.

Although the lengthening constraints cannot produce disyllabic minimal words on their own, they do help to determine which type of disyllable emerges. Assuming that disyllabic minimal words have a trochaic strong–weak stress contour, we can explain the two-syllable requirement with an additional non-finality constraint, NON-FINALITY(ω), which bans stress from the final syllable of a prosodic word.[8] Once the strong–weak contour is established, lengthening constraints determine the weight of the initial syllable. NON-FINALITY(σ), which produces lengthening in both iambic feet and trochaic feet, requires that the initial syllable be heavy. NON-FINALITY(Ft), which produces lengthening only in iambic feet, tolerates a light initial syllable.

(45) a. NON-FINALITY(ω), NON-FINALITY(σ) >> DEP-μ
Iambic or trochaic lengthening + HL minimal word
b. NON-FINALITY(ω), NON-FINALITY(Ft) >> DEP-μ
Iambic lengthening + LL minimal word

[8] Plausible cases of iambic minimal words appear to be extremely rare.

This correctly predicts that either iambic or trochaic lengthening can be accompanied by an HL minimal word, but only iambic lengthening can be accompanied by an LL minimal word.

Van de Vijver (1998) and Revithiadou (2004) propose an approach to rhythmic lengthening that is similar in some respects to the non-finality approach. Although it does not rely on the non-finality formulation, it does posit two lengthening mechanisms. One lengthens stressed syllables generally, which produces both iambic lengthening and trochaic lengthening, and the other lengthens foot-final syllables in particular, which produces only iambic lengthening.

(46) *Lengthening constraints* (van de Vijver 1998)

a. STRESSED SYLLABLE LENGTH
   A stressed syllable is long and an unstressed syllable is short.
b. FOOTFINAL
   Foot-final elements are lengthened.

Since there are two sources for iambic lengthening and only one for trochaic lengthening, Revithiadou's and van de Vijver's proposals, like the non-finality approach, provide an account of the different frequencies with which the two types of lengthening occur. The advantage of the non-finality approach is that it incorporates the lengthening mechanisms into the much more general non-finality formulation, a formulation independently motivated by its ability to account for a surprisingly broad range of phenomena at different prosodic levels. (See CHAPTER 43: EXTRAMETRICALITY AND NON-FINALITY.)

5 Summary

The most interesting interpretation of the Iambic–Trochaic Law is a strong interpretation that focuses on the presence or absence of durational contrasts in disyllabic feet. Since the general typology of attested stress systems offers very little support for the strong interpretation, Hayes (1985) introduced the ITL to metrical theory under a weaker interpretation that focused on quantity-sensitivity. This also turned out to be inadequate, however, as it was soon recognized that both iambic languages and trochaic languages could be either quantity-sensitive or quantity-insensitive. Two subsequent accounts – McCarthy and Prince (1986) and Hayes (1987, 1995) – pursued a hybrid approach, combining aspects of the weak interpretation and the strong interpretation. Another, Prince (1990), returned to a strong interpretation of the ITL, but applied it, in effect, only to quantity-sensitive systems and as a preference rather than an absolute requirement.

Since the ITL is inherently quantity-sensitive, it seems more natural to employ it as a foundation for an account of quantity-sensitive systems in particular than as the foundation for an account of stress systems generally. There is, in fact, considerable support for the ITL among quantity-sensitive systems. Iambic quantity-sensitivity differs from trochaic quantity-sensitivity, as indicated by the different ways in which alternating patterns resume after encountering a heavy syllable, and the asymmetric foot inventory of ITL accounts very effectively for this difference. Standard iambs exclude heavy syllables from weak position

Derbyshire, Desmond C. 1985. *Hixkaryana and linguistic typology*. Dallas: Summer Institute of Linguistics & University of Texas at Arlington.

Dixon, R. M. W. 1988. *A grammar of Boumaa Fijian*. Chicago: University of Chicago Press.

Dvonč, Lad'islav. 1955. *Rytmický zákon v spisovnej slovenčine*. Bratislava: Vydavatel'stvo Slovenskej Akadémie Vied.

Echeverría, Max S. & Heles Contreras. 1965. Araucanian phonemics. *International Journal of American Linguistics* 31. 132–135.

Everett, Daniel L. 1996. Prosodic levels and constraints in Banawá and Suruwaha. Unpublished ms., University of Pittsburgh (ROA-121).

Everett, Daniel L. 2003. Iambic feet in Paumari and the theory of foot structure. *Linguistic Discovery* 2(1). 22–44.

Geytenbeek, Brian B. & Helen Geytenbeek. 1971. *Gidabal grammar and dictionary*. Canberra: Australian Institute of Aboriginal Studies.

Goddard, Ives. 1979. *Delaware verbal morphology: A descriptive and comparative study*. New York: Garland.

Hay, Jessica S. F. & Randy L. Diehl. 2007. Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception and Psychophysics* 69. 113–122.

Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 11. 429–446.

Hayes, Bruce. 1987. A revised parametric metrical theory. *Papers from the Annual Meeting of the North East Linguistic Society* 17. 274–289.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hoff, Berend J. 1968. *The Carib language: Phonology, morphonology, morphology, texts and word index*. The Hague: Martinus Nijhoff.

Hulst, Harry van der. 1999. Issues in foot typology. In S. J. Hannahs & Mike Davenport (eds.) *Issues in phonological structure: Papers from an international workshop*, 95–127. Amsterdam & Philadelphia: John Benjamins.

Hyde, Brett. 2007. Non-finality and weight-sensitivity. *Phonology* 24. 287–334.

Iversen, John R., Aniruddh D. Patel & Kengo Ohgushi. 2008. Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America* 124. 2263–2271.

Jacobson, Steven A. 1984. The stress conspiracy and stress-repelling bases in the Central Yup'ik and Siberian Yupik Eskimo languages. *International Journal of American Linguistics* 50. 312–324.

Jacobson, Steven A. 1985. Siberian Yupik and Central Yupik prosody. In Krauss (1985b), 25–45.

Kager, René. 1989. *A metrical theory of stress and destressing in English and Dutch*. Dordrecht: Foris.

Kager, René. 1993. Alternatives to the iambic–trochaic law. *Natural Language and Linguistic Theory* 11. 381–432.

Kager, René. 1995. Review article of Hayes (1995). *Phonology* 12. 437–464.

Kenstowicz, Michael. 1983. Parametric variation and accent in the Arabic dialects. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 19. 205–213.

Kenstowicz, Michael & Kamal Abdul-Karim. 1980. Cyclic stress in Levantine Arabic. *Studies in the Linguistic Sciences* 10(2). 55–76.

Knudson, Lyle M. 1975. A natural phonology and morphophonemics of Chimalapa Zoque. *Papers in Linguistics* 8. 283–346.

Krauss, Michael. 1985a. Supplementary notes on Central Siberian Yupik prosody. In Krauss (1985b), 47–50.

Krauss, Michael (ed.) 1985b. *Yupik Eskimo prosodic systems: Descriptive and comparative studies*. Fairbanks: Alaska Native Language Center.

Kusumoto, Kiyomi & Elliott Moreton. 1997. Native language determines parsing of nonlinguistic rhythmic stimuli. *Journal of the Acoustical Society of America* 102. 3204.

Leer, Jeff. 1985. Toward a metrical interpretation of Yupik prosody. In Krauss (1985b), 159–172.

Lombardi, Linda & John J. McCarthy. 1991. Prosodic circumscription in Choctaw morphology. *Phonology* 8. 37–71.

McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.

Mellander, Evan W. 2003. (HL)-creating processes in a theory of foot structure. *The Linguistic Review* 20. 243–280.

Mester, Armin. 1994. The quantitative trochee in Latin. *Natural Language and Linguistic Theory* 12. 1–61.

Michelson, Karin. 1988. *A comparative study of Lake-Iroquoian accent*. Dordrecht: Kluwer.

Mithun, Marianne & Hasan Basri. 1986. The phonology of Selayarese. *Oceanic Linguistics* 25. 210–254.

Nicklas, Thurston Dale. 1972. The elements of Choctaw. Ph.D. dissertation, University of Michigan, Ann Arbor.

Nicklas, Thurston Dale. 1975. Choctaw morphophonemics. In James M. Crawford (ed.) *Studies in Southeastern Indian languages*, 237–250. Athens: University of Georgia Press.

Prince, Alan. 1990. Quantitative consequences of rhythmic organization. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 26(2). 355–398.

Revithiadou, Anthi. 2004. The Iambic/Trochaic Law revisited: Lengthening and shortening in trochaic systems. *Leiden Papers in Linguistics* 1. 37–62.

Revithiadou, Anthi & Ruben van de Vijver. 1997. Durational contrasts and the iambic/trochaic law. *Proceedings of Western Conference on Linguistics* 9. 229–242.

Rice, Curt. 1992. Binarity and ternarity in metrical theory: Parametric extensions. Ph.D. dissertation, University of Texas, Austin.

Schütz, Albert J. 1978. English loanwords in Fijian. In Albert J. Schütz (ed.) *Fijian language studies: Borrowing and pidginization*, 1–50. Suva: Fiji Museum.

Schütz, Albert J. 1985. *The Fijian language*. Honolulu: University of Hawaii Press.

Seiler, Hansjakob. 1965. Accent and morphophonemics in Cahuilla and in Uto-Aztecan. *International Journal of American Linguistics* 31. 50–59.

Seiler, Hansjakob. 1967. Structure and reconstruction in some Uto-Aztecan languages. *International Journal of American Linguistics* 33. 135–147.

Seiler, Hansjakob. 1977. *Cahuilla grammar*. Banning, CA: Malki Museum Press.

Seiler, Hansjakob & Kojiro Hioki. 1979. *Cahuilla dictionary*. Banning, CA: Malki Museum Press.

Topping, Donald M. & Bernadita C. Dungca. 1973. *Chamorro reference grammar*. Honolulu: University Press of Hawaii.

Vijver, Ruben van de. 1998. The iambic issue: Iambs as a result of constraint interaction. Ph.D. dissertation, University of Leiden.

Voegelin, Charles F. 1935. Tübatulabal grammar. *University of California Publications in American Archaeology and Ethnology* 34. 55–189.

Vos, P. 1977. Temporal duration factors in the perception of auditory rhythmic patterns. *Scientific Aesthetics* 1. 183–199.

Woodbury, Anthony. 1981. Study of the Chevak dialect of Central Alaskan Yupik. Ph.D. dissertation, University of California, Berkeley.

Woodbury, Anthony. 1987. Meaningful phonological processes: A consideration of Central Alaskan Yupik Eskimo prosody. *Language* 63. 685–740.

Woodrow, Herbert. 1909. A quantitative study of rhythm. *Archives of Psychology (New York)* 14. 1–66.

45 The Representation of Tone

Larry M. Hyman

1 Introduction

No issue has had a greater impact on phonological representations than the study of tone. Although receiving only passing attention in both pre- and early generative phonology, tone quickly moved away from its marginal status to occupy center stage in the development of non-linear phonology. While both level and contour tones had been traditionally transcribed with either accents or numerals, as in (1) and (2), the assumption in early generative phonology, e.g. Wang (1967), was that tones consisted of features that could be added at the bottom of a segmental feature matrix, as in (3).

(1)

	Awa		*Falam*		*Obokuitai*	
High (H)	ná	'breast'	páa	'mushroom'	kuik^{1}	'rock'
Low (L)	nà	'house'	kèe	'leg'	kuik^{2}	'insect (sp.)'
HL (falling)	nâ	'taro'	sâa	'animal'	kuik^{12}	'lizard (sp.)'
LH (rising)	pǎ	'fish'	zǔu	'bear'		
	(Loving 1966: 25)		(personal notes)		(Jenison and Jenison 1991: 85)	

(2)

Jingpho			*Ayutla Mixtec*		
H	mu^{55}	'word'	H–H	ʃi^{1}ɲuʔ^{1}	'pineapple'
Mid (M)	mu^{33}	'delicious'	H–L	ʃi^{1}niʔ^{3}	'hat'
L	mu^{31}	'to see'	M–L	ʃi^{2}niʔ^{3}	'head'
HL	mu^{51}	'mother'	L–L	ti^{3}kuʔ^{3}	'louse'
(Qingxia and Diehl 2003: 401)			(Pankratz and Pike 1967: 291)		

(3) a. H *tone* /á/

$$\begin{bmatrix} +\text{syll} \\ -\text{cons} \\ +\text{back} \\ +\text{low} \\ +\textsc{high} \end{bmatrix}$$

b. HL *falling tone* /â/

$$\begin{bmatrix} +\text{syll} \\ -\text{cons} \\ +\text{back} \\ +\text{low} \\ +\textsc{falling} \end{bmatrix}$$

A major representational problem was how to account for the properties of contour tones. Although falling and rising tones could be expressed less formally by combining accents (â, ǎ) or numerals (31, 13, etc.), features such as [±FALLING] and [±RISING] fail to capture what are known as "edge effects": a high to low falling tone acts like a H tone with respect to what precedes, but as a L tone with respect to what follows. Similarly, a low to high rising tone acts like a L tone with respect to what precedes, but as a H tone with respect to what follows. Representations such as in (4), which were occasionally entertained, would simply be incoherent in a framework in which a segment consists of a single vertical matrix of features:

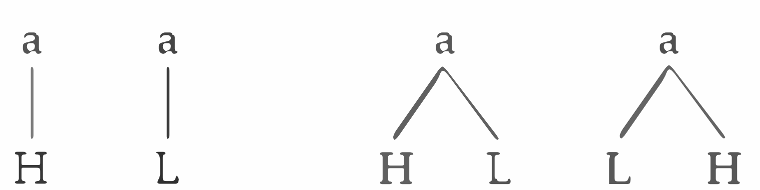
(4) a. *Falling tone /â/* b. *Rising tone /ǎ/*

$$\begin{bmatrix} +\text{syll} \\ -\text{cons} \\ +\text{back} \\ +\text{low} \\ [+\text{HIGH}][+\text{LOW}] \end{bmatrix} \qquad \begin{bmatrix} +\text{syll} \\ -\text{cons} \\ +\text{back} \\ +\text{low} \\ [+\text{LOW}][+\text{HIGH}] \end{bmatrix}$$

In order to solve this and other representational problems, Goldsmith (1976a, 1976b) proposed a theory of "autosegmental" phonology in which segments and tones appear on separate "tiers," as in (5).

(5) *Autosegmental representations of* H, L, HL, LH

a. *level tones* b. *contour tones*



By so doing, Goldsmith was able to capture the traditional intuitions implicit in the accent and numeral notations and make predictions about what should *vs.* should not be found in tone systems. Armed with the autosegmental framework, enormous strides were made in the analysis of tone as well as in other applications of the framework, e.g. segmental harmonies (Clements 1977, 1981; Hoberman 1988), feature geometry (Clements 1985; Clements and Hume 1995) and prosodic morphology (McCarthy 1981; McCarthy and Prince 1986).

The main question to be addressed in this chapter is the extent to which the key insight of autosegmental phonology, expressed in (6), is still valid:

(6) *Tones are semi-autonomous from their tone-bearing units*

a. tones are on a separate tier, but
b. they are linked to their tone-bearing units by association lines.

The chapter is organized as follows: in §2 we consider some of the predictions of autosegmental tonology in order to see how they have fared since the 1970s. In §3 we consider the issue of underspecification, while §4 addresses the issues of tone features, tonal geometry, and tone-bearing units. §5 evaluates potential limitations of autosegmental representations, while §6 concludes the chapter with a brief consideration of tone in constraint-based phonology. We will see that there is much reason to hold on to the autosegmental insight even as phonological frameworks have evolved.

2 Autosegmental tonology

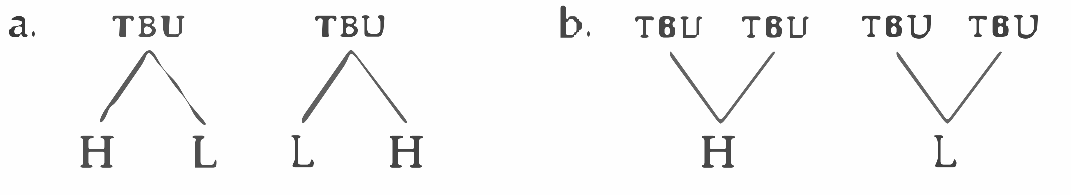
One can distinguish two consequences of autosegmental phonology (see also CHAPTER 14: AUTOSEGMENTS) as applied to tone: those that follow directly from the architecture *vs.* those that involve additional principles or conventions. We take up the first of these here and postpone discussion of the second until §5.

There are at least three direct consequences of the two-tier autosegmental architecture proposed by Goldsmith, as in (7):

(7) a. non-isomorphism between the tiers
  b. zero representation on one *vs.* the other tier
  c. stability effects

2.1 *Non-isomorphism*

Using TBU to represent the tone-bearing unit to which tones link (see §4), the first of these is schematized in (8).

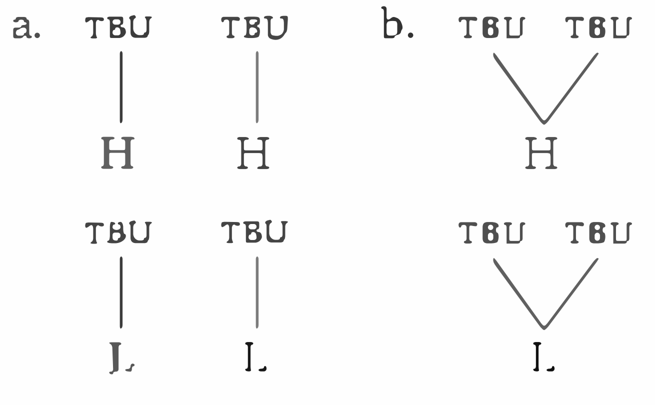
(8)


We have already seen that more than one tone can link to a single TBU, resulting in falling and rising contour tones, as in (8a). Complex contours are also attested, as exemplified in (9).

(9)	a.	falling-rising:	Iau	be^{213}	'tree fern'	(Bateman 1990)
				ba^{243}	'sticking to'	
	b.	rising-falling:	Nzadi	mwǎ̀n	'child'	(personal notes)
				dzǐ̀	'eye'	

In addition, Lomongo is said to have LHLH on one syllable derived from elision (Hulstaert 1961: 164): /èmí là wě bàsàngì/ → [èm_â w_ǎ̌sàngì] 'it's you and I who are related', where _ marks the two places where elision occurs.

The second type of non-isomorphism in (8b) shows that the same tone can link to more than one TBU. What this means is that there is a potential contrast between the representations in (10a) *vs.* those in (10b).

(10)


contrasts with /Ø/ rather than /L/.) Although a rare occurrence, the tautomorphemic version of the OCP is thereby violated.

Similar contrasts and OCP problems are found with respect to /L/ tones in Dioula d'Odienné (Braconnier 1982). In this language there is a rule of the form in (14), where pause (//) may be interpreted with a %L boundary tone:

(14) L → H / {//, L} __ H

As seen in (15), the rule is not iterative, as only one L autosegment can be raised:

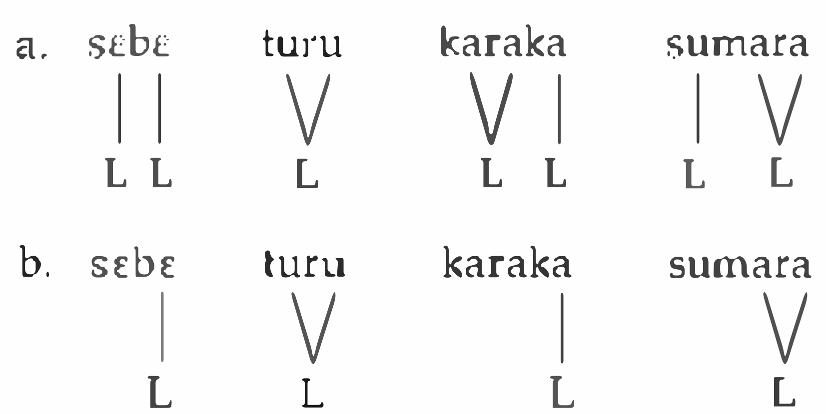
(15)	a.	/i mà dɛ̀ tá/	→	ì mà dɛ̌ tá	'you didn't take any child'
				*ì má dɛ̌ tá	
	b.	/i mà tùrù tá/	→	ì mà túrú tá	'you didn't take any oil'
				*ì má túrú tá	

However, the following examples show that not all monomorphemic L tone nouns have the same properties before H:

(16)	a.	*before pause*	b.	*before* H	
		sɛ̀bɛ̀		sɛ̀bɛ́	'paper'
		tùrù		túrú	'oil'
		kàràkà		kàràká	'bed'
		sùmàrà		sùmárá	'soumbala (a spice)'

One solution is to give the above nouns the representations in (17a), where the OCP is violated similarly to Odden's (1982) tautomorphemic Hs in Shambala.

(17)



Another solution in (17b) might be to posit a distinction between /L/ and toneless TBUs, which later become L by default. While the OCP is violable in the ways just seen, it is important to note that the autosegmental representations provide a straightforward way of encoding the contrastive tonal properties of Kukuya, Kishambaa, and Dioula d'Odienné. Other frameworks would have to resort to ad hoc junctures or diacritics.

2.2 *Zero representation*

The second consequence that directly flows from the autosegmental architecture is the possibility that a tone can exist without a TBU, and vice versa. When a morpheme consists solely of a tone, it is referred to as a tonal morpheme (see also CHAPTER 82: FEATURAL AFFIXES). An oft-cited example is the associative (genitive) marker in Igbo (see Williamson 1986 and references cited therein), which is most transparently observed when preceded and followed by a L–L noun. As seen in (18), the H tonal morpheme is assigned to the preceding TBU in Central Igbo and to the following TBU in the Aboh dialect:

(18) a. *Central Igbo* àgbà + ´ + èŋwè → àgbá èŋwè 'jaw of monkey'
b. *Aboh Igbo* ẹ̀gbà + ´ + èŋwè → ẹ̀gbà éŋwè 'jaw of monkey'

Floating tones can also be part of a lexical morpheme. In Aghem, the nouns [kífú] 'rat' and [kɨ́wɔ́] 'hand' are both pronounced with H–H tones in isolation (Hyman 1979). However, as seen in (19), they have different effects on the following word (the [kí-] prefix drops out when the noun is modified):

(19) a. fú kɨ̂a 'your (SG) rat' b. fú kɨ́n 'this rat'
H L H H
wɔ́ kɨ̀a 'your (SG) hand' wɔ́ ꜜkɨ́n 'this hand'
H L L H L H

In (19a), the H of the stem /-fú/ spreads onto the /L/ of /kɨ̀a/ 'your SG (class 7)' creating a HL contour. As seen, the floating L of /-wɔ́`/ 'hand' blocks the spreading. When followed by the /H/ tone demonstrative /kɨ́n/ 'this (class 7)' in (19b), 'this rat' is realized H–H, while 'this hand' has a downstep conditioned by the same lexical floating L of /-wɔ́`/, which was originally due to a historically lost syllable (cf. Proto-Bantu *-bókò 'hand').

Corresponding to floating tones, which lack a TBU, are TBUs that lack tones. Such toneless morphemes may receive their tonal specification by context or by default tone assignment (see §3). An oft-cited example of the former comes from Mende (Leben 1973, 1978):

(20)

	base noun			+hu 'in'	+ma 'on'
a.	/H/	kɔ́	'war'	kɔ́-hú	kɔ́-má
b.	/L/	bɛ̀lɛ̀	'trousers'	bɛ̀lɛ̀-hù	bɛ̀lɛ̀-mà
c.	/HL/	mbû	'owl'	mbú-hù	mbú-mà
d.	/LH/	mbà	'rice'	mbà-hú	mbà-má
e.	/LHL/	njàhà	'woman'	njàhá-hù	nyàhá-mà

As seen, the tone of the two locative postpositions is the same as the last underlying tone of the nouns to which they attach. The /HL/, /LH/, and /LHL/ "melodies" are linked one-to-one to the noun + postposition constituent.

2.3 *Stability effects*

The third consequence of autosegmental representations, stability, is related to the second: when a TBU is deleted, its tone may still remain, and vice versa. An example of this comes from Tangale, which Kenstowicz and Kidda (1987: 230) cross each other. The representation of L as /Ø/, on the other hand, allows for such long-distance processes without requiring further complications.

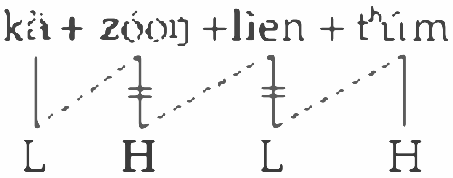
Just as in the case of single *vs.* multiply linked tones, underspecification potentially allows for more distinctions than fully specified representations. Limiting ourselves first to languages that have only two tone heights, (23) summarizes the possible analyses:

(23) a. /H/ *vs.* /L/
  b. /H/ *vs.* /Ø/
  c. /L/ *vs.* /Ø/
  d. /H/ *vs.* /L/ *vs.* /Ø/

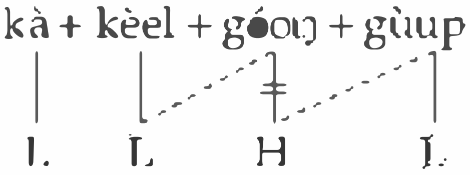
As seen, the tonal contrast may be binary, with /H/ contrasting with /L/; privative, with either /H/ or /L/ contrasting with /Ø/; or both, in the case of where a ternary system of /H, L, Ø/ is required. Thus, the central question concerning any such system is to determine which of the above representations best accounts for the properties of the surface [H] *vs.* [L] contrast.

A /H, L/ system is required when both features are phonologically active (see also CHAPTER 4: MARKEDNESS for more discussion of markedness). This is seen in Kuki-Thaadow, which has both H- and L-tone spreading (Hyman 2010):

(24) a. /kà + zóoŋ + lien + tʰúm/ [kà zòoŋ líen tʰǔm] 'my three big monkeys'



  b. /kà + kèel + góoŋ + gùup/ [kà kèel gòoŋ gûup] 'my six thin goats'



Not only do both /H/ and /L/ spread onto a following /L/ and /H/ syllable, respectively, but the result in final position is a rising (LH) or falling (HL) contour tone. Both spreading and contour tones would be difficult to represent if one of the tones were underspecified. The same would be true of a language which has both floating H and L.

While /H, L/ systems are those in which the phonology refers to both tone values, in a privative tone system only one of the two tones is phonologically active. Many Bantu languages have a /H, Ø/ system, as was seen in the Giryama examples in (22). A few have a /L, Ø/ system, such as Ruund, for which Nash (1992–94) gives the following arguments:

(25) a. Hs are by far more numerous than Ls, hence "unmarked."
  b. Floating L exists, while floating H does not.
  c. Morphological rules assign L tones, not Hs.
  d. Phonological rules manipulate L tones, not Hs.

Athabaskan languages are also known for having H- *vs.* L-marked tone systems (see the various studies in Hargus and Rice 2005). The closely related South American dialects, Bora (Weber and Thiesen 2000) and Miraña (Seifart 2005), have

(1988) propose that Yoruba has the system in (27b), where M tone is not only unmarked, but underspecified. Among the arguments is the fact that Yoruba allows only HL and LH contours, but none involving M tone (*MH, *LM, *HM, *ML). If M = /Ø/, it follows that Ø cannot form a contour with either H or L. Another argument derives from alternations in which H or L overrides M tone. While all of these properties naturally follow from the underspecification of M, the latter can be set up as /M/, as long as other mechanisms are put in place to capture the recessiveness of M tones. Pulleyblank (2004), for instance, proposes an account within Optimality Theory in which, essentially, M tones need not be preserved in outputs to the same degree as H or L. An argument for a /H, Ø, L/ system comes from Peñoles Mixtec (Daly and Hyman 2007), which has an OCP constraint disallowing successive Ls, or Ls which are separated by any number of /Ø/ TBUs. The result is the L-tone deletion rule in (28a).

(28) a. L → Ø / L __
b. ɨ̈^{N} di̱i-ni-kwe-ʃi kada-kwe-ʃi ɨ̈^{N} ɨ̈^{N} ʧi̱u^N (^{N} = nasalization)
L (under di̱i) … L (under ʧi̱u) ↓ Ø
'only one of them will do each of the jobs'

As seen, this rule is responsible for the loss of the second L on /ʧi̱u^N/ 'work' in (28b), which is separated from the first L by twelve toneless TBUs. That the two L tones can 'see' each other over long distances is strong evidence for the underspecification account.

Returning to the systems in (27), (27c) and (27d) seem to be rare or non-occurring. Paster (2003) proposes /H, M, Ø/ for Leggbó, with /Ø/ alternating between H and L in the verb morphology (cf. §4). If correct, the relatively few short-vowel LH, HL, and LM contours would have to be exceptionally marked: [èggǔ] 'catfish (sp.)', [gēppjôn] 'afternoon', [lèssò¯] 'last year'. While I am unaware of any three-height tone system being analyzed as /Ø, M, L/, the system in (27e) seems at first appropriate for Yatzachi Zapotec (Pike 1948). Although this language has only the three surface tones H, M, L (and HM and MH contours on monosyllabic words), there are two kinds of L tones: those that alternate with M *vs.* those that do not. Pike identifies these, respectively, as class A *vs.* class B. For example, a class B L will become M when followed by a M or H within or across words, while a class A L tone will not be affected. This is seen in the minimal pair in (29).

(29)	a.	L_a	[bìa]	'cactus'	bìa gōlī	'old cactus'
	b.	L_b	[bìa]	'animal'	bīa gōlī	'old animal'

A natural interpretation would be to recognize L_a as /L/ and L_b as /Ø/. Where the /Ø/ tone is not realized M, i.e. when not occurring before a /M/ or /H/, it will be realized with a default L tone, hence merging with class A /L/. However, it is hard to evaluate this proposal without considering what the featural representation is of all of the tones in the language (see Hyman 2009). It is likely that Yatzachi Zapotec has two kinds of L tone because it originally contrasted four tone heights. The L_b /Ø/ could therefore have been a tone level between M and or binary and hierarchized (Clements 1983). Anderson (1978) provides a comprehensive appraisal of the different feature proposals up to that date. If assimilations are assumed to spread individual features such as [UPPER] and [RAISED], these latter must occur on independent tiers. This then raises the question of where these tiers link up, e.g. to a laryngeal node or directly to the TBU? There have been numerous proposals in the literature (see the surveys and discussion in Yip 1995, 2002; Bao 1999; Snider 1999; and Lee 2008). The advantage of the first proposal is that tones frequently interact with other laryngeal properties. Halle and Stevens' (1971) system in (35) – an early attempt to capture the relation between tone and obstruent voicing – fails, however, to characterize more than three tone heights:

(35)

	tones			*voiceless obstruents*	*sonorants*	*voiced obstruents*
	H	M	L	p t k f s	m n l w j	b d g v z
stiff	+	–	–	+	–	–
slack	–	–	+	–	–	+

Clearly, one wants to account for the relation between tone and (non-modal) phonations, or the interference of obstruent voicing with tonal assimilations, but not at the expense of losing the generalization that tones are distributed by TBUs. While tones have an autosegmental independence, they ultimately must be realized on something, e.g. a vowel or syllabic consonant. A language may consistently assign tones by mora, such that a CVV syllable receives two tones, or it may assign tones by syllable. Sometimes it is difficult to distinguish between the two. The complexities and corresponding representational possibilities are many. Both Clements *et al.* (2009) and Hyman (2009) have expressed doubts that tones should be analyzed in terms of features at all.

5 Possible limits of autosegmental representation

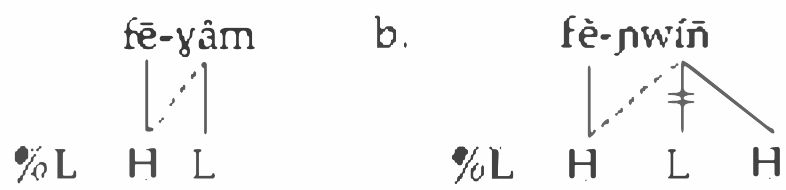
Two general conclusions can be drawn from the preceding sections. First, there is still much merit in the autosegmental insights on tone. Second, there is much more work to do. In this section we first address real or apparent problems for autosegmental tonology. The arguments for autosegmental tonology were enumerated in §2. Some of the evidence concerned the non-isomorphism between the tones and their TBUs: more than one tone can link to a TBU, in which case we get a tonal contour or cluster; conversely, one tone can link to more than one TBU. Contrasting representations such as those in (12a) and (17a) were said to be needed. The question here is whether they ever get in the way: are there cases where it is disadvantageous to represent a tautomorphemic H–H or L–L as a single tone linked to two TBUs?

One such awkward case would seem to arise in Kom, which has an underlying contrast between /H/ and /L/, but a surface contrast between H, M, and L. All M tones are derived by a rule that lowers a single H TBU to M when preceded by a L or initial phrase boundary (which can be represented by a %L boundary tone). This produces outputs such as the following:

(36) a. /fe-ɣam/ → fē-ɣâm 'mat' [M–HL]
H L
b. /fe-ɲwin/ → fē-ɲwíň 'bird' [M–HM]
H LH
c. /fe-buʔ/ → fē-búʔ 'gorilla' [M–H]
HHL
d. /fe-tam/ → fē-tám 'fruit' [M–H]
H H

As in these examples, most nouns in Kom have a /H/-tone prefix followed by a monosyllabic stem that can have any of the four tone patterns exemplified in (36). As schematized in (37), the H of the prefix spreads onto a following L or LH stem:

(37) a. b.

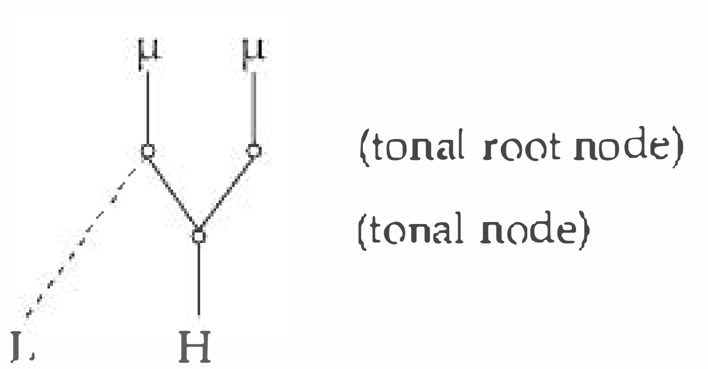


As seen in the transcriptions, the output is M–HL in (37a) and M–HM in (37b). The M tones in question are conditioned by the rule that lowers a H to M after a (linked, floating, or boundary) L. In both forms the prefix is thus lowered to M; in addition, in (37b), the (delinked) stem L lowers the following H to produce the HM contour. Since the prefix lowers to M without affecting the stem, the lowering rule cannot be written as a single-tier rule, as in (38), or we would get the wrong outputs *M–ML and *M–M:

(38) H → M / L __ (cf. the Kukuya rule in (13))

It is clear that the doubly linked H representation is not useful, but it is not fatally contradictory to the autosegmental approach. At least four responses come to mind. First, assuming that the TBU is the mora and that stems are bimoraic, the M can be derived by delinking a H from the first TBU (mora) that immediately follows a L. Second, one can complicate the tonal geometry to include a tonal root node, as in (39), to which the preceding L can link as a register feature, perhaps [–RAISED] (cf. the surveys of similar proposals in Bao 1999 and Snider 1999):

(39)



Third, assuming that floating tones persist into the output, one might argue that the M outputs are derived by phonetic implementation, where the M may be interpolated as part of the aligning of output tones with their segmental supports. A fourth possibility is to spread the L to form a LH TBU, which is phonetically interpreted as M.

(50) a. ne fe-ɣam -fe → nè fè-ɣǎm̄ -fè° 'with a mat' [L–L–ML–L°]
L H L H → L H L H

b. ne fe-ɲwin -fe → nè fè-ɲwīn -fé 'with a bird' [L–L–M–H]
L H L H H → L H L H H

c. ne fe-buʔ -fe → nè fè-búʔ -fē 'with a gorilla' [L–L–H–M]
L H HL H → L H HL H

d. ne fe-tam -fe → nè fè-tám -fé 'with a fruit' [L–L–H–H]
L H H H → L H H H

As seen, Kom has both H- and L-tone spreading. In (50a), L-tone spreading causes the following H to delink (producing a level L° tone on the postposition), while H-tone spreading does not cause the following L to delink. With a single H TBU lowering to M after L (cf. (37)), the noun stem /-ɣàm/ 'mat' surfaces with a ML contour tone. Similarly, the stem /-ɲwìn/ 'bird' surfaces as M, each of the two linked H tones being lowered to M by a preceding L. The important examples are (50c) and (50d). As seen, when the L of /nè/ spreads and delinks the H of the noun prefix /fé-/, the latter's H tone floats. As a result, the floating H shields the roots /-búʔ`/ 'gorilla' and /-tám/ 'fruit' from lowering to M. (The floating L of /-búʔ`/ is responsible for the lowering of the postposition /-fé/ to M in (50c).) Since the H to M lowering process is a late one for which we have even entertained the possibility that it applies in phonetic implementation, it should be clear that the floating Hs must persist into the surface representations. The alternative, that L-tone spreading does not delink a following H, but rather creates LH rising tones that are simplified in phonetic implementation, is perhaps suspect, if not ad hoc. If, on the other hand, the floating tones are allowed to occur in surface representations, the solution is straightforward.

Such a conclusion would also seem to have consequences for other analyses. Particularly within OT, the question arises whether a surviving, delinked floating tone should be able to satisfy MAX(Tone). Data from Kuki-Thaadow suggest maybe not. In this largely monosyllabic Tibeto-Burman language, words are underlyingly /H/, /L/, or /HL/ (Hyman 2010). As seen in (51a), however, a rising tone results from L-tone spreading:

(51) a. /hùon + zóoŋ/ → hùon zǒoŋ 'garden monkey'
L H

b. /hùon + zóoŋ + gùup/ → hùon zòoŋ gûup 'six garden monkeys'
L H L

In (51b) we see that Kuki-Thaadow also has H-tone spreading, and the H of /zóoŋ/ is delinked. This is because rising (LH) and falling (HL) tones are permitted only phrase-finally. Thus, the L of /hùon/ 'garden' is realized on /zóoŋ/ 'monkey', and the H of /zóoŋ/ is realized on /gùup/ 'six'.

Now consider the forms in (52).

```
(52)  a.  /hùon + zóoŋ + thúm/ → hùon zóoŋ thúm    'three garden monkeys'
            |      |       |
            L      H       H

      b.  /hùon + zóoŋ + gîet/  → hùon zóoŋ gîet   'eight garden monkeys'
            |      |       |
            L      H       HL
```

As seen, L-tone spreading does not occur when the targeted H tone is followed in turn by a H or HL tone. The question is: Why not? To answer this we have but to consider what the output would have been if L-tone spreading had applied in (52). The H of /zóoŋ/ would necessarily have had to delink, since a LH rising tone is well formed only in phrase-final position. As a result, the underlying /H/ would, at best, have to float. The question, then, is why this would not be well formed. It cannot be that an input link must remain as such in the output, since the H does delink in (51b). Instead, what seems to allow L-tone spreading to apply in (51a) and (51b) is that the input /H/ of /zóoŋ/ is preserved in the output: It is realized within the LH contour on its own syllable in (51a) and as part of the HL contour on the following syllable in (51b). In other words, Max(H) is satisfied. We must suppose, therefore, that if L-tone spreading applied in (52) and delinked the /H/ of /zóoŋ/, the resulting floating H would not satisfy Max(H). By contrast, H-tone spreading applies whether or not the targeted L remains linked on the surface (see Hyman 2010). We therefore can establish the ranking in (53).

(53) Max(H) >> Spread(H,L) >> Max(L)

In Kuki-Thaadow all input H tones make it to the surface, *vs.* closely related Hakha Lai, where the opposite ranking Max(L) >> Max(H) results in all input L tones being realized on the surface (Hyman and VanBik 2004). It should be noted that the non-application of L-tone spreading in (52) results more straightforwardly from the ranked constraints in (53), rather than from a requirement on recoverability: had L-tone spreading applied in (52a), the output L–L–H sequence would have unambiguously pointed to underlying /L–H–H/, since /L–L–H/ would have been realized as L–L–LH. (The same is not true of (52b), since L-tone spreading does not apply to a following HL tone.) If (53) is correct, OT will have made a unique contribution in providing a constraint, Max(H), which is responsible for the non-application of L-tone spreading in (52). By contrast, a rule-based approach would have to stipulate that L-tone spreading occurs to a L–H sequence when the H is followed either by pause or by a L tone, and would not provide any motivation for why the rule does not apply when the following tone is H or HL. Whether this is an indication that OT is on the right track – and can hence offer more improvements over the pre-OT conceptions of tonology – remains to be seen. In fairness, it must be said that the blocking of an otherwise general L-tone spreading process in (52) is unique to Kuki-Thaadow. As pointed out by Hyman (1973b: 157), we expect L–H–H to be a better target for L-tone spreading than L–H–L. Except for Kuki-Thaadow, this is true whether the H–H sequence is from two /H/ tones or from one /H/ that is doubly linked.

Where does this leave the representation of tone? To summarize the foregoing sections, although the well-formedness and mapping conventions in (44) have been superseded in subsequent work, most of the essential representational insights of autosegmental tonology are still intact. The above discussion has only touched on a small part of the vast world of tone and of the growing constraint-based literature treating tone. Whatever the outcome of ongoing OT interpretations of tone, it is likely that questions of representation will remain central.

REFERENCES

Akinlabi, Akinbiyi. 1985. Tonal underspecification and Yorùbá tone. Ph.D. dissertation, University of Ibadan.

Akinlabi, Akinbiyi & Ngessimo M. Mutaka. 2001. Tone in the infinitive in Kinande: An OT analysis. In Ngessimo M. Mutaka & Sammy B. Chumbow (eds.) *Research mate in African linguistics: Focus on Cameroon*, 333–356. Cologne: Rüdiger Köppe Verlag.

Anderson, Stephen R. 1978. Tone features. In Fromkin (1978), 133–175.

Bao, Zhiming. 1999. *The structure of tone*. New York & Oxford: Oxford University Press.

Bateman, Janet. 1990. Iau segmental and tonal phonology. *Miscellaneous Studies of Indonesian and Other Languages in Indonesia* 10. 29–42.

Biber, Douglas. 1981. The lexical representation of contour tones. *International Journal of American Linguistics* 47. 271–282.

Bogers, Koen, Harry van der Hulst & Maarten Mous (eds.) 1986. *The phonological representation of suprasegmentals*. Dodrecht: Foris.

Braconnier, Cassian. 1982. *Le système du dioula d'Odienné*, vol. 1. Abidjan: University of Abidjan.

Cassimjee, Farida & Charles W. Kisseberth. 1992. The tonology of depressor consonants: Evidence from Mijikenda and Nguni. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 18(2). 26–40.

Cassimjee, Farida & Charles W. Kisseberth. 1998. Optimal Domains Theory and Bantu tonology: A case study from Isixhosa and Shingazidja. In Hyman & Kisseberth (1998), 33–132.

Chao, Yuen-Ren. 1930. A system of tone-letters. *Le maître phonétique* 45. 24–27.

Clements, G. N. 1977. The autosegmental treatment of vowel harmony. In Wolfgang U. Dressler & Oskar E. Pfeiffer (eds.) *Phonologica 1976*, 111–119. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Clements, G. N. 1981. Akan vowel harmony: A non-linear analysis. *Harvard Studies in Phonology* 2. 108–177.

Clements, G. N. 1983. The hierarchical representation of tone features. In Ivan R. Dihoff (ed.) *Current approaches to African linguistics*, vol. 1. 145–176. Dordrecht: Foris.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. & Kevin C. Ford. 1979. Kikuyu tone shift and its synchronic consequences. *Linguistic Inquiry* 10. 179–210.

Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In Goldsmith (1995), 245–306.

Clements, G. N., Alexis Michaud & Cédric Patin. 2009. Do we need tone features? Paper presented at the Symposium on Tones and Features, University of Chicago.
Daly, John P. & Larry M. Hyman. 2007. On the representation of tone in Peñoles Mixtec. *International Journal of American Linguistics* 73. 165–208.
Edmondson, Jerold A. & Kenneth J. Gregerson. 1992. On five-level tone systems. In Shina Ja J. Huang & William R. Merrifield (eds.) *Language in context: Essays for Robert E. Longacre*, 555–576. Arlington: Summer Institute of Linguistics & University of Texas at Arlington.
Fromkin, Victoria A. (ed.) 1978. *Tone: A linguistic survey*. New York: Academic Press.
Goldsmith, John A. 1976a. Autosegmental phonology. Ph.D. dissertation, MIT.
Goldsmith, John A. 1976b. An overview of autosegmental phonology. *Linguistic Analysis* 2. 23–68.
Goldsmith, John A. (ed.) 1995. *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell.
Halle, Morris & Kenneth N. Stevens. 1971. A note on laryngeal features. *MIT Research Laboratory of Electronics Quarterly Progress Report* 101. 198–213.
Hargus, Sharon & Keren Rice (eds.) 2005. *Athabaskan prosody*. Amsterdam & Philadelphia: John Benjamins.
Hoberman, Robert D. 1988. Emphasis harmony in modern Aramaic. *Language* 64. 1–26.
Hulstaert, Gustaaf. 1961. *Grammaire du Lɔmɔngɔ*, vol. 1: *Phonologie*. Tervuren: Musée Royal de l'Afrique Centrale.
Hyman, Larry M. 1973a. The role of consonant types in natural tonal assimilations. In Hyman (1973b), 151–179.
Hyman, Larry M. (ed.) 1973b. *Consonant types and tone*. Los Angeles: University of Southern California.
Hyman, Larry M. 1979. Phonology and noun structure. In Larry M. Hyman (ed.) *Aghem grammatical structure*, 1–72. Los Angeles: University of Southern California.
Hyman, Larry M. 1986. The representation of multiple tone heights. In Bogers *et al.* (1986), 109–152.
Hyman, Larry M. 1987. Prosodic domains in Kukuya. *Natural Language and Linguistic Theory* 5. 311–333.
Hyman, Larry M. 2005. Initial vowel and prefix tone in Kom: Related to the Bantu Augment? In Koen Bostoen & Jacky Maniacky (eds.) *Studies in African comparative linguistics with special focus on Bantu and Mande: Essays in honour of Y. Bastin and C. Grégoire*, 313–341. Cologne: Rüdiger Köppe Verlag.
Hyman, Larry M. 2009. Do tones have features? Paper presented at the Symposium on Tones and Features, University of Chicago.
Hyman, Larry M. 2010. Kuki-Thaadow: An African tone system in Southeast Asia. In Franck Floricic (ed.) *Essais de typologie et de linguistique générale*, 31–51. Lyon: Presses de l'École Normale Supérieure.
Hyman, Larry M. & Charles W. Kisseberth (eds.) 1998. *Theoretical aspects of Bantu tone*. Stanford: CSLI.
Hyman, Larry M. & Daniel J. Magaji. 1970. *Essentials of Gwari grammar*. Ibadan: Institute of African Studies, University of Ibadan.
Hyman, Larry M. & Armindo Ngunga. 1994. On the non-universality of tonal association "conventions": Evidence from Ciyao. *Phonology* 11. 25–68.
Hyman, Larry M. & Maurice Tadadjeu. 1976. Floating tones in Mbam-Nkam. In Larry M. Hyman (ed.) *Studies in Bantu tonology*, 57–111. Los Angeles: Department of Linguistics, University of Southern California.
Hyman, Larry M. & Kenneth VanBik. 2004. Directional rule application and output problems in Hakha Lai tone. *Phonetics and Phonology, Special Issue, Language and Linguistics* 5. 821–861. Taipei: Academia Sinica.
Hyman, Larry M., Francis Katamba & Livingstone Walusimbi. 1987. Luganda and the strict layer hypothesis. *Phonology Yearbook* 4. 87–108.

Hyman, Larry M., Heiko Narrog, Mary Paster & Imelda Udoh. 2002. Leggbó verb inflection: A semantic and phonological particle analysis. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 28, 399–410.

Jenison, Scott D. & Priscilla B. Jenison. 1991. Obokuitai phonology. *Workpapers in Indonesian Languages and Culture* 9. 69–90.

Kamanda-Kola, Roger. 2003. *Phonologie et morphosyntaxe du mono: Langue oubanguienne du Congo R. D.* Munich: Lincom Europa.

Kenstowicz, Michael & Mairo Kidda. 1987. The obligatory contour principle and Tangale phonology. In David Odden (ed.) *Current approaches to African linguistics*, vol. 4, 223–238. Dordrecht: Foris.

Leben, William R. 1973. Suprasegmental phonology. Ph.D. dissertation, MIT.

Leben, William R. 1978. The representation of tone. In Fromkin (1978), 177–219.

Lee, Seunghun Julio. 2008. Consonant–tone interactions in Optimality Theory. Ph.D. dissertation, Rutgers University.

Loving, Richard E. 1966. Awa phonemes, tonemes, and tonally differentiated allomorphs. *Papers in New Guinea Linguistics* 5. 23–32.

Maddieson, Ian. 1971. The inventory of features. In Ian Maddieson (ed.) *Tone in generative phonology*, 3–18. Ibadan: Institute of African Studies, University of Ibadan.

Maddieson, Ian. 1978. Universals of tone. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of human language*, vol. 2: *Phonology*, 335–365. Stanford: Stanford University Press.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In Hyman (1973b), 71–95.

McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12. 373–418.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John J. 2004. Headed spans and autosegmental spreading. Unpublished ms., University of Massachusetts, Amherst (ROA-685).

McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst & Brandeis University.

McCarthy, John J. & Alan Prince. 1994. The emergence of the unmarked: Optimality in prosodic morphology. *Papers from the Annual Meeting of the North East Linguistic Society* 24. 333–379.

Mutaka, Ngessimo. 1994. *The lexical tonology of Kinande*. Munich: Lincom Europa.

Myers, Scott. 1997. OCP effects in Optimality Theory. *Natural Language and Linguistic Theory* 15. 847–892.

Myers, Scott. 1998. Surface underspecification of tone in Chichewa. *Phonology* 15. 367–391.

Nash, Jay A. 1992–94. Underlying low tones in Ruwund. *Studies in African Linguistics* 23. 223–278.

Newman, Paul. 1986. Contour tones as phonemic primes in Grebo. In Bogers *et al.* (1986), 175–193.

Nougayrol, Pierre. 1979. *Le day de bouna (Tschad)*, vol. 1: *Eléments de description linguistique*. Paris: SELAF.

Odden, David. 1982. Tonal phenomena in KiShambaa. *Studies in African Linguistics* 13. 177–208.

Olson, Kenneth S. 2005. *The phonology of Mono*. Dallas: SIL International & University of Texas at Arlington.

Pankratz, Leo & Eunice V. Pike. 1967. Phonology and morphophonemics of Ayutla Mixtec. *International Journal of American Linguistics* 33. 287–299.

Paster, Mary. 2003. Tone specification in Leggbo. In John M. Mugane (ed.) *Linguistic typology and representation of African languages*, 139–150. Trenton, NJ: Africa World Press.

Paulian, Christiane. 1975. *Le kukuya: Langue teke du Congo*. Paris: Societé d'Études Linguistiques et Anthropologiques de France.

Philippson, Gérard. 1998. Tone reduction *vs.* metrical attraction in the evolution of Eastern Bantu tone systems. In Hyman & Kisseberth (1998), 315–329.
Picanço, Gessiane. 2002. Tonal polarity as phonologically conditioned allomorphy in Mundurukú. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 28. 237–248.
Pike, Eunice V. 1948. Problems in Zapotec tone analysis. *International Journal of American Linguistics* 14. 161–170.
Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.
Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology*. Dordrecht: Reidel.
Pulleyblank, Douglas. 1988. Vocalic underspecification in Yoruba. *Linguistic Inquiry* 19. 233–270.
Pulleyblank, Douglas. 2004. A note on tonal markedness in Yoruba. *Phonology* 21. 409–425.
Qingxia, Dai & Lon Diehl. 2003. Jingpho. In Graham Thurgood & Randy J. LaPolla (eds.) *The Sino-Tibetan languages*, 401–408. London & New York: Routledge.
Scott, Graham. 1990. A reanalysis of Fore accent. *La Trobe University Working Papers in Linguistics* 3. 139–150.
Seifart, Frank. 2005. The structure and use of shape-based noun classes in Miraña (North West Amazon). Ph.D. dissertation, Radboud University Nijmegen.
Smith, Neil V. 1968. Tone in Ewe. *MIT Research Laboratory of Electronics Quarterly Progress Report* 88. 290–304.
Snider, Keith L. 1999. *The geometry and features of tone*. Dallas: Summer Institute of Linguistics & University of Texas at Arlington.
Stahlke, Herbert F. W. 1971. The noun prefix in Ewe. *Studies in African Linguistics*. 141–159.
Stahlke, Herbert F. W. 1977. Some problems with binary features for tone. *International Journal of American Linguistics* 43. 1–10.
Wang, William S.-Y. 1967. Phonological features for tone. *International Journal of American Linguistics* 33. 93–105.
Weber, David & Wesley Thiesen. 2000. A synopsis of Bora tone. *Work Papers of the Summer Institute of Linguistics* 45. Available (May 2010) at http://www.und.edu/dept/linguistics/wp/2001.htm.
Welmers, William E. 1962. The phonology of Kpelle. *Journal of African Languages* 1. 69–93.
Williamson, Kay. 1986. The Igbo associative and specific constructions. In Bogers *et al.* (1986), 195–208.
Yip, Moira. 1980. The tonal phonology of Chinese. Ph.D. dissertation, MIT.
Yip, Moira. 1988. Template morphology and the direction of association. *Natural Language and Linguistic Theory* 6. 551–577.
Yip, Moira. 1989. Contour tones. *Phonology* 6. 149–174.
Yip, Moira. 1995. Tone in East Asian languages. In Goldsmith (1995), 476–494.
Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.
Zoll, Cheryl. 2003. Optimal tone mapping. *Linguistic Inquiry* 34. 225–268.

46 Positional Effects in Consonant Clusters

Jongho Jun

1 Introduction

It is commonly observed across languages that phonological processes may apply only in certain ("non-prominent") positions. In contrast, elements in other ("prominent") positions typically resist or trigger these processes. Such prominent *vs.* non-prominent positional distinctions are further applicable to more general patterns of licensing and neutralization of phonological contrasts; namely, featural/segmental contrasts are likely to be licensed in prominent positions, whereas these contrasts are likely to be neutralized in non-prominent positions. Pairs of prominent and non-prominent positions include word-initial *vs.* non-initial, stressed *vs.* unstressed, root *vs.* affix, and prevocalic *vs.* preconsonantal positions. Among the positional effects involving these pairs, this chapter is mainly concerned with those involving the two members of intervocalic consonant clusters. See Beckman (1998) and Barnes (2006) for recent extensive investigations of other positional effects; see also CHAPTER 104: ROOT–AFFIX ASYMMETRIES and CHAPTER 102: CATEGORY-SPECIFIC EFFECTS. In intervocalic C_1C_2 clusters, the preconsonantal C_1 is more likely to undergo phonological processes such as voicing and place assimilation, in contrast with the prevocalic C_2, which is rarely subject to such processes. I will refer to this asymmetric positional effect as the C_2 dominance effect.

This chapter discusses empirical data patterns which display positional effects. Its focus will be on how to explain the C_2 dominance effect. I will begin with a discussion of typical data patterns of the C_2 dominance effect and proceed to less common, somewhat exceptional, patterns which are nonetheless crucial in comparing the previous approaches. Specifically, I will concentrate on the comparison between prosody-based approaches (Itô 1986, 1989; Cho 1990; Goldsmith 1990; Rubach 1990; Lombardi 1995, 1999, 2001b; Beckman 1998; Kabak and Idsardi 2007) and cue-based approaches (Steriade 1993, 1995, 1999, 2001, 2009; Flemming 1995; Jun 1995, 2004; Padgett 1995; Boersma 1998; Hume 1999; Côté 2000; Wilson 2001; Blevins 2003; Seo 2003). It will be shown that current evidence is mixed. Much of the commonly observed data, to be discussed in the following two sections, can be equally well accommodated in the two approaches. However, there exist less common patterns which can be understood under only one of the two approaches. Evidence exclusively supporting the cue-based approach will be discussed in §4, and evidence for the prosody-based approach in §5.

2 C_2 dominance effect

Assimilation occurs in consonant clusters when one of two neighboring consonants takes on some property of the other. I will call the former (i.e. the undergoer of assimilation) the target, and the latter (i.e. the source of the assimilating property) the trigger. With respect to the assimilation in C_1C_2 clusters, it is cross-linguistically true that C_1 and C_2 are the target and the trigger, respectively, and thus the direction of assimilation is regressive. To illustrate this C_2 dominance effect, I will first consider patterns of voicing assimilation and then patterns of place assimilation. Finally, patterns of consonant deletion will be discussed.

As can be seen in (1), in Catalan, Polish and Russian, voiced and voiceless obstruents are separate phonemes, and they may occur unhindered in prevocalic position. But in clusters composed of obstruents, the first constituent of the cluster must agree in voicing with the following constituent. As shown in (1.i), underlyingly voiced obstruents in C_1 become voiceless before a voiceless obstruent in C_2 whereas underlyingly voiceless obstruents in C_1 become voiced before a voiced obstruent in C_2, as in (1.ii). Thus, voicing assimilation occurs in clusters, targeting C_1. This C_2 dominance effect in voicing assimilation can be seen in other languages, including Dutch, Yiddish, Sanskrit, Romanian, Serbo-Croatian, Ukrainian, Hungarian, Egyptian Arabic, and Lithuanian. Steriade (1999) and Beckman (1998) provide in-depth discussion of voicing assimilation, i.e. a type of laryngeal neutralization, in these languages, emphasizing that it is normally regressive, and thus its C_2 dominance effect is quite robust.

(1) *Regressive voicing assimilation*

a. *Catalan* (from Beckman 1998, citing Hualde 1992)

				__ voiceless	
i.	/b/	'ʎoβə	'wolf (FEM)'	'ʎoppə'tit	'small wolf'
	/g/	ə'mɪɣə	'friend (FEM)'	ə'mɪkpə'tɪt	'little friend'
				__ voiced	
ii.	/t/	'gatɔ	'cat (FEM)'	'gaddu'len	'bad cat'
	/k/	'pɔkə	'little (FEM)'	'pɔg'du	'a little hard'

b. *Polish* (from Rubach 1996, Beckman 1998)

				__ voiceless	
i.	/b/	ża[b]a	'frog'	ża[pk]a	'small frog'
	/d/	wo[d]a	'water'	wo[tk]a	'vodka'
				__ voiced	
ii.	/ʧ/	li[ʧ]yć	'count'	li[dzb]a	'numeral'
	/k/	szla[k]-u	'route (GEN SG)'	szla[gb]ojowy	'war route'

c. *Russian* (from Kenstowicz 1994, Kiparsky 1985, Padgett 2002)

				__ voiceless	
i.	/b/	korob-a	'bastbax (GEN SG)'	kirop-ka	'bastbax (DIM)'
	/d/	pod-nesti	'to bring (to)'	pot-pisatʲ	'to sign'

(3) *Casual speech place assimilation*

a. *Korean*[1]

i.	/mit+ko/	[mitk'o] ~ [mikk'o]	'believe (CONN)'
	/ip+ko/	[ipk'o] ~ [ikk'o]	'wear (INF)'
	/cinan pam/	[cinanbam] ~ [cinambam]	'last night'
ii.	/ik+ta/	[ikt'a], *[itt'a], *[ikk'a]	'ripe (INF)'
	/ip+ta/	[ipt'a], *[itt'a], *[ipp'a]	'wear (INF)'
	/paŋ+pota/	[paŋbota], *[pambota], *[paŋgota]	'(more) than room'

b. *English* (based on Bailey 1970)

i.	right poor	righ[p] poor	
	good-bye	goo[b]bye	
ii.	keep track	*kee[t] track	*keep [p]rack
	back track	*ba[t] track	*back [k]rack

This C_2 dominance effect in casual speech assimilation can also be seen in other languages, such as German (Kohler 1990, 1991a, 1991b, 1992), Malay, Thai (Lodge 1986, 1992), Toba Batak (Hayes 1986), Spanish (Harris 1969), and Ponapean (Rehg and Sohl 1981); see also CHAPTER 79: REDUCTION on assimilation as a casual speech phenomenon.

The C_2 dominance effect is not limited to assimilation in consonant clusters, but extends to consonant deletion in clusters. Consonant deletion occurs in clusters when one of two adjacent consonants, i.e. the target, deletes. It has been observed and emphasized in the literature (Côté 2000; Wilson 2001; Jun 2002; McCarthy 2008) that C_1, as opposed to C_2, is always the target in such deletions. For instance, as shown in (4), stops in C_1, not C_2, delete in Diola Fogny, West Greenlandic, and Basque.

(4) *Consonant deletion: C_1 is the target*

a. *Diola Fogny* (Sapir 1965)

/let+ku+jaw/	[le kujaw]	'they won't go'
/kutεb sinaŋas/	[kutε sinaŋas]	'they carried the food'
/εkεt bo/	[εkε bo]	'death there'

b. *West Greenlandic* (Rischel 1974; Fortescue 1980)

/qanik+lerpoq/	[qani lerpoq]	'begins to approach'
/ukijuq+tuqaq/	[ukiju tuqaq]	'old year'
/anguti+kulak/	[angu kulak]	'he goat'

c. *Basque* (Hualde 1987)

/bat+paratu/	[ba-paratu]	'put one'
/bat+kurri/	[ba-kurri]	'run one'
/guk+piztu/	[gu-piztu]	'we light'
/guk+kendu/	[gu-kendu]	'we take away'

[1] In Korean, which has a three-way laryngeal contrast among obstruents, i.e. lenis, aspirated, and tense, lenis obstruents become tense after an obstruent (Post-obstruent tensing), and voiced between sonorants (Inter-sonorant voicing). See Kim-Renaud (1986) and Ahn (1998) for details of these automatic processes.

In addition to those shown above, consonant deletion with a C_1 target can be found in Akan (Lombardi 2001b, citing Schachter and Fromkin 1968), Axininca person prefixes (Lombardi 2001b, citing Payne 1981), Carib (Gildea 1995), and Tunica (Wilson 2001, citing Haas 1946).

Consequently, as summarized below, the cross-linguistic generalization which is common to the three phonological processes (voicing assimilation, place assimilation, and consonant deletion) is that C_1 in intervocalic C_1C_2 clusters is the target, whereas C_2 is the trigger.

(5) *The C_2 dominance effect in voicing assimilation, place assimilation, and consonant deletion*

In intervocalic C_1C_2 clusters, C_1 is a typical target, and C_2 is a typical trigger.

Let us consider how to capture this C_2 dominance effect. It is not difficult to derive patterns with the C_2 dominance effect within the framework of previous theories such as classical generative theory, autosegmental phonology, and underspecification theory. For instance, regressive assimilation can easily be characterized by a rule of the type shown in (6a). However, progressive assimilation can also be formulated with equal complexity, as shown in (6b), and its absence, or at best rarity, in the typology would be a surprise. Representational theories such as autosegmental phonology, feature geometry, and underspecification theory would be no better in this respect than classical generative theory, as there is no plausible reason to differentiate in the complexity of the representation between the two members of a consonant cluster. (See Jun 1995 for the relevant discussion.)

(6) *Rules for consonant place assimilation*

a. $C \rightarrow [\alpha\text{place}]$ / __ $[C, \alpha\text{place}]$ (regressive assimilation)
b. $C \rightarrow [\alpha\text{place}]$ / $[C, \alpha\text{place}]$ __ (progressive assimilation)

Compared to the rule-based theories with a focus on the correct formulation of the language-specific phonological processes, Optimality Theory (McCarthy and Prince 1995; Prince and Smolensky 2004) is more concerned with universal patterns, thus being in a better position to explain positional effects such as the C_2 dominance effect and understand the motivation behind processes showing the effect.

Along with the development of Optimality Theory, there have been two major lines of approach to the analysis of the positional effects, Licensing-by-cue and Licensing-by-prosody (in Steriade's 1999 terminology). The prosody-based approach explains positional asymmetries by reference to prosodic structure. It attributes the C_2 dominance effect to the coda-onset asymmetry since C_1 and C_2 are usually syllabified as a coda and an onset, respectively. The C_1 in the coda is likely to be targeted in the processes since the coda is phonologically non-prominent and marked. In contrast, the C_2 in the onset resists these processes since the onset is phonologically prominent and unmarked. For an analysis of the data of the C_2 dominance effect, either greater faithfulness to the onset and/or dispreference for the coda or (marked) properties in the coda have been called on. Specifically, within the framework of Optimality Theory, positional faithfulness constraints for the onset or/and positional markedness constraints for the coda have been adopted in the literature. (See Casali 1996, 1997 and Beckman 1998 for positional faithfulness analyses, and Zoll 1998 for the comparison between positional faithfulness and positional markedness.)

In contrast, the cue-based approach (Flemming 1995; Steriade 1995, 1999, 2001, 2009; Boersma 1998; Côté 2000; Wilson 2001; Blevins 2003; Jun 2004) explains the C_2 dominance effect by relying on the perceptual factors involved. The C_1 has low perceptibility since it is preconsonantal and thus may lack important perception cues, such as release bursts and C-to-V formant transitions, to laryngeal/place features and segmenthood under overlap with C_2 (Lamontagne 1993; Wright 1996). In contrast, the C_2 is perceptually prominent since it is prevocalic, being able to maintain such perception cues. (See Wright 2004 for a detailed discussion of perception cues.) From the assumption that change in perceptually prominent positions would cause drastic input-output difference, and thus be greatly dispreferred, whereas the comparable change in non-prominent positions would cause less difference, and thus be less dispreferred, it is derived that non-prominent C_1 is more likely to be modified, i.e. targeted in phonological processes, than prominent C_2. Thus, the cue-based approach attributes the C_2 dominance effect to higher perceptibility of C_2 over C_1.

The two approaches under consideration differ in whether the constraints (and rules) adopted to explain positional effects should be expressed as prosody-based or string-based (more precisely, cue-based) statements. However, the empirical data presented thus far will not distinguish the two approaches, since the preconsonantal C_1 which is expected to be the target of the phonological processes in the cue-based approach is normally syllabified as a coda, which the prosody-based approach also expects to be the more likely target. In §4 and §5, I will present the data patterns for which the two approaches make distinct predictions.

3 Non-assimilatory neutralization

Assimilation can be considered as a case of contrast neutralization. As shown in the previous section, assimilation is primarily regressive (i.e. C_1 is the target), and thus potential contrasts of the assimilating feature are neutralized in C_1 position. For instance, in regressive voicing assimilation, consonants in C_1 with distinct voice feature values in their underlying form would have identical phonetic realization with respect to voicing, i.e. they are voiced before a voiced segment and voiceless before a voiceless one. Non-assimilatory neutralization of voicing, as well as other laryngeal features and place of articulation features, targets the C_1 and word-final position. In languages which have voicing assimilation in consonant clusters, the word-final position is the only available target of non-assimilatory neutralization, and in fact languages with voicing assimilation mentioned in the previous section show final devoicing, as in (7) (see also CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION).

(7) *Final devoicing in languages with voicing assimilation*

a. *Catalan* (Beckman 1998, citing Hualde 1992)

		__V		__#	
i.	/t/	['gatə]	'cat (FEM)'	['gat]	'bad (MASC)'
	/k/	['pɔkə]	'little (FEM)'	['pɔk]	'little (MASC)'

(9) *Place assimilation and final place neutralization*

a. *Spanish* (from Harris 1984)
   i. Homorganic NC clusters: ca[mp]o, ma[ns]o, ma[ŋk]o, á[ɱf]ora, ma[n̪t]o
   ii. Final [n] in standard varieties: e[n] Chile, ta[n] frío, u[n] elefante
   iii. Final [ŋ] in non-standard varieties: e[ŋ] Chile, ta[ŋ] frío, u[ŋ] elefante

b. *Japanese* (Vance 1987; Yip 1991; Itô *et al.* 1995; Kager 1999)
   i. *Geminates*

kap.pa	'a legendary being'
kit.te	'stamp'
gak.koo	'school'

   ii. *Homorganic nasal + obstruent*

tom.bo	'dragonfly'
non.do	'tranquil'
kaŋ.gae	'thought'
/jom + te/ → jonde	'reading'
/ʃin + te/ → ʃinde	'dying'

   iii. *Final mora nasal*

hoN	'book'
zeN	'goodness'

There are also languages in which place neutralization occurs only in a non-assimilatory fashion. As discussed in Lombardi (2001b, citing Rice 1989), non-sonorant consonants in Slave (Athabaskan) are realized as /h/ syllable-finally, as shown in (10a). Sonorants are like obstruents in having no place distinctions, although the exact final neutralization patterns are not the same. Syllable-final nasals delete, nasalizing the preceding vowel, and /j/ is the only possible coda among non-nasal sonorants. Another example of non-assimilatory place neutralization is from the Kelantan dialect of Malay (Teoh 1988). Final stops /k t p/ are realized as [ʔ], and final fricatives like /s/ as [h], as shown in (10b).

(10) *Final place neutralization* (from Lombardi 2001b)

a. *Slave*

i.	/ts'ad/	ts'ah	-ts'ade	'hat'
ii.	/xaz/	xah	-ɣaze	'scar'
iii.	/seeɣ/	seeh	-zeeɣe	'saliva'
	/tl'uł/	tl'uh	-tl'ułe	'rope'

b. *Kelantan Malay*

i.	/ikat/	ikaʔ	'tie'
ii.	/səsak/	səsaʔ	'crowded'
iii.	/dakap/	dakaʔ	'embrace'
iv.	/tapis/	tapih	'to filter'

In these languages, contrasts of some features other than place can be maintained finally, for instance obstruents *vs.* sonorants in Slave and stops *vs.* fricatives in Kelantan Malay. But there are also many languages like Burmese in which all consonants are neutralized to [ʔ]. (See Lombardi 2001b and references therein for more details.)

In summary, as stated below, the preconsonantal C_1 is not only the typical target of assimilation, but also the typical target of the non-assimilatory neutralization. The word-final position is an additional typical target of the neutralization.

(11) *The C_2 dominance effect in (assimilatory and non-assimilatory) laryngeal and place neutralization*

Preconsonantal C_1 and word-final positions are common target positions.

It is usually the case that preconsonantal C_1 and word-final positions form a natural class, i.e. coda. Thus it is obvious that the prosody-based approach can provide a unified account of positional neutralizations of C_1 and word-final positions, attributing both cases to the coda-onset asymmetry. The cue-based approach also provides a somewhat unified account for the two common target positions based on the fact that the word-final position lacks C-to-V transition cues, just as preconsonantal C_1 does, and thus it has lower perceptibility compared to the prevocalic C_2 position. Consequently, the relatively common positional neutralization patterns presented thus far do not significantly distinguish the two approaches. In the remainder of this chapter, I will consider the patterns which crucially distinguish them.

4 Evidence for the cue-based approach

4.1 *Neutralization sites ≠ syllable positions*

According to the prosody-based approach, neutralization contexts should be described in prosodic terms: for instance, "codas are the target of laryngeal neutralization." But, as discussed by Steriade (1999), there are cases in which there is no consistent connection between neutralization sites and syllable structure.

First, there are languages in which neutralization targets only C_1, not word-final, positions. Languages in which laryngeal neutralization occurs only in C_1, not at the end of the word, include Yiddish, Romanian, Serbo-Croatian (Lombardi 2001b: 269), French, Hungarian, and Kolami (Steriade 1999). In addition, place neutralization of nasals occurs only in C_1, not at the end of the word, in Diola Fogny (Sapir 1965) and the Souletin dialect of Basque (Hualde 1993). Under the prosody-based approach, it is not clear why word-medial and final codas behave differently, and even less clear why medial codas are more likely to be targeted in the neutralization than word-final codas (see CHAPTER 36: FINAL CONSONANTS for more discussion). In contrast, in the cue-based approach, the asymmetry between preconsonantal C_1 and final positions may be derived from their relative perceptibility difference. C_1 may be considered less perceptible than word-final position, because stops in C_1, which overlap with consonants in C_2, are more likely to be unreleased, thus lacking the release burst and closure duration cues, than those in word-final position.[3]

[3] Blevins (2006: 143) discusses data from Dhaasana, Chadic Arabic, and Maltese in which devoicing occurs exclusively at the end of the word, not in C_1 position. This word-final, but not syllable-final, devoicing can be a problem not only for the prosody-based approach but also for the cue-based approach.

Second, laryngeal neutralization targets C_1 only before obstruents, not before sonorants, regardless of its syllabic assignment in many languages, including Lithuanian, German, Russian, Greek, Sanskrit, Polish, Hungarian, and Kolami. Notice that it is possible that C_1 in intervocalic C_1C_2 clusters is syllabified as a coda when sonorant consonants occupy C_2, and its voicing contrast is then expected to be neutralized in the prosody-based approach. Also, it is usually the case that an obstruent is syllabified as an onset when it occurs as the first constituent of the word-initial clusters composed of obstruents, and then its voicing contrast is expected to be licensed in the prosody-based approach. These two expectations of the prosody-based approach are not satisfied in the languages mentioned above. For instance, in Lithuanian, where consonant clusters are heterosyllabic regardless of composition (e.g. /'auk.le/), voicing of obstruents may be contrastive in the coda when they occur before sonorants, as in (12c). The voicing of an obstruent is neutralized in the onset when it precedes another obstruent, as in (12b). Consequently, it is difficult to provide an adequate description of neutralization contexts in syllabic terms.

(12) *Lithuanian obstruents in clusters* (Steriade 1999)

		voiceless		*voiced*	
a.	licensed onsets	sam'gus	'cheerful'	žmo'gus	'man'
b.	neutralized onsets	spalva	'color'	'lizdas	'nest'
c.	licensed codas	aug.muo	'growth'	ak.muõ	'stone'
d.	neutralized codas	daũ [k]	'much'		

In contrast, to explain the difference in the likelihood of the neutralization between pre-obstruent and pre-sonorant positions, the cue-based approach may still rely on the perceptibility difference of the two positions. Specifically, the pre-obstruent position lacks the main contextual cues (VOT and other release-related cues), and thus is less perceptible than the pre-sonorant position, where the main cues can be maintained.

Finally, there are languages in which neutralization patterns are fixed despite the variable syllabification. As discussed by Steriade (1999), for both Sanskrit and Ancient Greek syllable divisions in obstruent–sonorant clusters were variable, depending on "the dialect, the period, the literary style and the juncture separating the consonants." In contrast, there was no variation in the pattern of laryngeal neutralization: in styles or dialects where $VC_1.C_2V$ divisions were the norm for all clusters, laryngeal neutralization did not take place before heterosyllabic sonorants. This indicates that laryngeal features in these languages are neutralized irrespective of the syllabic affiliation of clusters, and thus the neutralization patterns cannot be adequately described in syllable terms.

The above patterns indicate that syllable positions like codas are neither a sufficient nor a necessary condition for the occurrence of neutralization. Codas are not a sufficient condition in the patterns in which only word-medial, as opposed to final, codas and pre-obstruent, as opposed to pre-sonorant, codas are neutralized. Codas are not a necessary condition in the patterns in which an obstruent onset in word-initial clusters is neutralized. Codas would be totally useless in describing the Sanskrit and Greek patterns with variable syllabification but fixed neutralization patterns. Consequently, all these patterns can be taken as evidence against the prosody-based approach and in favor of the cue-based approach.

C_2 in Murinbata. Miriwung is just like Murinbata, in that the apical contrast is maintained in C_1, but neutralized in C_2. Apical neutralization additionally occurs at the beginning of the word in which only alveolars, not retroflexes, are allowed to occur.

Consequently, the typical targets of apical neutralization may be summarized as below:

(15) *The C_1 dominance effect in (assimilatory and non-assimilatory) apical neutralization*

Postconsonantal C_2 and word-initial positions are common target positions.

This is therefore the complete opposite of the C_2 dominance effect in the neutralization patterns of laryngeality and major C-Places of articulation. Given that C_2 is usually an onset, the prosody-based approach cannot explain the C_1 dominance effect in apical neutralization in the same way as the C_2 dominance effect summarized in (13). In contrast, in the cue-based approach, the C_1 dominance effect may be derived naturally from the perception fact that cues to the apical distinction lie primarily in the V-to-C, not C-to-V, transitions, and thus C_1 is more prominent in the perception of the apical contrast than C_2. As discussed by Steriade (2001), citing Ladefoged and Maddieson (1986), Dave (1976), Stevens and Blumstein (1975), and Bhat (1973), the formant transitions into retroflexes in C_1 show distinctively low F3 and F4 values, compared to those of denti-alveolars, whereas the transitions out of retroflexes in C_2 are not distinct from those of denti-alveolars. This acoustic asymmetry originates from the characteristic articulation of retroflexes, in which the tongue tip moves forward during the closure and releases from the same constriction location as apico-alveolars.

Consequently, both the C_2 dominance effect in major C-Place assimilation and the C_1 dominance effect in apical assimilation may be derived from the main argument of the cue-based approach, i.e. the neutralization targets positions which lack prominent perceptual cues to the contrasts in question. Thus, the apical neutralization typology may form very strong evidence for the cue-based approach by showing a case of contrast-specific neutralization. (See Zhang's 2004 discussion of contour tone typology for an additional case of contrast-specific licensing/neutralization.)

5 Evidence for the prosody-based approach

5.1 *Obstruent–sonorant clusters in Catalan*

The cue-based approach provides a string-based, not prosody-based, account for positional neutralizations. If two sequences are segmentally identical, and thus not significantly different in the perceptibility involved, the cue-based approach expects the two to behave similarly with respect to neutralization even when they have different prosodic structures. Suppose that in obstruent–sonorant C_1C_2 clusters, the C_1 obstruent may be syllabified either as an onset or as a coda, depending on the environment. The cue-based approach expects that the C_1 obstruent will behave invariably with respect to positional neutralization, regardless of whether it is an onset or a coda. If, as predicted by the prosody-based approach, the C_1 is licensed when syllabified as an onset, but neutralized when syllabified

b. dental / __ lateral

voiced		*voiceless*
atleta [d.l]	'athlete'	*none*

Further, comparable active neutralization, which is even more difficult to explain within the cue-based approach, can be observed in the obstruent–sonorant sequences occurring across word boundaries. In Catalan, word-final obstruents assimilate in voicing to the following consonants including liquids. Word-final stops, which are voiceless before an initial vowel of the following word (18.ii), become voiced before an initial sonorant (18.i). Thus the C_1 obstruents here are neutralized with respect to voicing. This final obstruent neutralization would not be expected within the cue-based approach, since the same obstruent-sonorant sequences within a word are not subject to the voicing neutralization, as shown in (16a). So this is a case in which identical sequences behave differently with respect to neutralization, depending on where they occur. According to Wheeler (2005), the only difference between the obstruent-sonorant sequences of (16a) and (18.i) is in syllabic affiliation: the C_1 obstruents in (16a) are onsets, whereas those in (18.i) are codas. Thus, only the prosody-based, not cue-based, approaches can explain the voicing patterns of obstruents in Catalan.

(18) *Neutralized obstruent voicing in Catalan obstruent sonorant clusters*

a. non-sibilant / __ #liquid

i.	*poc logic*	/'pɔk#ə'lɔʒik/	['pɔg.'lɔʒik]	'not very logical'
ii.	*poc amable*		['pɔ.kə'mab.blə]	'not very friendly'

b. non-sibilant / __ #glide

i.	*poc whisky*	/'pɔk#'wiski/	['pɔg.'wiski]	'not much whisky'
ii.	*poc usual*		['pɔ.ku'zwal]	'not very usual'

5.2 *Obstruent–sonorant clusters in Eastern Andalusian Spanish*

Eastern Andalusian Spanish shows an additional case in which sequences with similar perception cues behave differently in contrast distribution and neutralizing processes, thus posing a problem to the cue-based approach. The discussion of this section is mostly based on Gerfen (2001).

As shown in (19), in Standard Peninsular Spanish, /s/ is allowed in preconsonantal and word-final positions. In contrast, in Eastern Andalusian Spanish, /s/ is not allowed to occur in those positions. As shown in (20a), word-final /s/ deletes and aspirates the preceding vowel. Preconsonantal C_1 /s/ also deletes but the deletion is accompanied by the gemination of the following C_2 consonant.

(19) */s/ in Standard Peninsular Spanish* (Gerfen 2001)

a.	word-final coda:	[ga.fas]	'eyeglasses'
b.	pre-C coda:	[gas.ko]	'helmet'

b. *Word-internal obstruent + liquid clusters*
   i. [a.klara] 's/he/it clears up'
   ii. [a.grada] 's/he/it pleases'
   iii. [a.plana] 's/he/it applies'
   iv. [a.trapa] 's/he/it traps'

But it is still unclear why /sl/, an obstruent–liquid sequence, does not behave like stop–liquid sequences. As shown in (20b.ii), /s/ is subject to coda deletion before /l/. Also, as shown below, /tl/, a stop–liquid sequence, behaves like /sl/, not like /kl/ and /gl/. If rich perceptual cues to C_1 before a liquid can guarantee the surface realization of stop–liquid sequences like /kl/, /gr/, /pl/, and /tr/, it is difficult to understand why /tl/ and /sl/ cannot surface as such.

(23) /tl/ *clusters in Eastern Andalusian Spanish* (Gerfen 2001)

/atleta/ [a^{h}lleta] 'athlete'

To summarize, both Catalan and Eastern Andalusian Spanish show asymmetric patterns in which only a subset of obstruent–sonorant sequences is subject to the distributional restrictions and related alternations targeting an obstruent in C_1. For the analysis of these patterns, the prosody-based approach can still attribute the difference among the obstruent–sonorant sequences to the coda-onset asymmetry, but an equally plausible, string-based, solution seems to be unavailable in the cue-based approach. See Kabak and Idsardi (2007) for an additional support for the prosody-based, as opposed to the cue-based, approach. They investigated Korean listeners' perception of non-native sequences, and argue that only syllable-based, not string-based, phonotactic constraints can explain their experimental results.

6 Conclusion

From the literature on phonological typology, we know that it is common for phonological processes not to apply in all positions, and more generally for phonological contrasts not to be licensed in all positions. Such positional effects are characterized by reference to certain pairs of prominent and non-prominent positions such as word-initial *vs.* non-initial, stressed *vs.* unstressed, root *vs.* affix, and prevocalic C_2 *vs.* preconsonantal C_1 positions. Among these, this chapter has been mainly concerned with the C_2 dominance effect in which preconsonantal C_1 in intervocalic C_1C_2 clusters is likely to be targeted for neutralization, whereas prevocalic C_2 is likely to trigger or resist such neutralization. This C_2 dominance effect is quite robust in laryngeal and place neutralization and consonant deletion. I have looked at the relevant data patterns, ranging from well known and common to less known and relatively exceptional, while comparing the cue-based and prosody-based approaches. Common data patterns can be explained equally well by both approaches. In contrast, less common or somewhat exceptional patterns may distinguish the two approaches. However, the evidence so far is mixed. Not only neutralization patterns in which there is no connection between neutralization sites and syllable positions, but also apical neutralization patterns,

Harris, James W. 1984. Autosegmental Phonology, Lexical Phonology, and Spanish nasals. In Aronoff & Oehrle (1984), 67–82.

Hayes, Bruce. 1984. The phonetics and phonology of Russian voicing assimilation. In Aronoff & Oehrle (1984), 318–328.

Hayes, Bruce. 1986. Assimilation as spreading in Toba Batak. *Linguistic Inquiry* 17. 467–499.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.

Hualde, José Ignacio. 1987. On Basque affricates. *Proceedings of the West Coast Conference on Formal Linguistics* 6. 77–89.

Hualde, José Ignacio. 1992. *Catalan*. London & New York: Routledge.

Hualde, José Ignacio. 1993. Topics in Souletin Phonology. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.) *Generative studies in Basque linguistics*, 289–327. Amsterdam & Philadelphia: John Benjamins.

Hume, Elizabeth. 1999. The role of perceptibility in consonant/consonant metathesis. *Proceedings of the West Coast Conference on Formal Linguistics* 17. 293–307.

Itô, Junko. 1986. Syllable theory in prosodic phonology. Ph.D. dissertation, University of Massachusetts, Amherst.

Itô, Junko. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7. 217–259.

Itô, Junko, Armin Mester & Jaye Padgett. 1995. Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26. 571–613.

Jun, Jongho. 1995. Perceptual and articulatory factors in place assimilation: An Optimality Theoretic approach. Ph.D. dissertation, University of California, Los Angeles.

Jun, Jongho. 1996. Place assimilation is not the result of gestural overlap: Evidence from Korean and English. *Phonology* 13. 377–407.

Jun, Jongho. 2002. Positional faithfulness, sympathy and inferred input. Unpublished ms., Yeungnam University, Daegu, Korea. http://ling.snu.ac.kr/jun.

Jun, Jongho. 2004. Place assimilation. In Hayes *et al.* (2004), 58–86.

Kabak, Baris & William J. Idsardi. 2007. Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints. *Language and Speech* 50. 23–52.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, MA & Oxford: Blackwell.

Kim-Renaud, Young-Key. 1974. Korean consonantal phonology. Ph.D. dissertation, University of Hawaii.

Kim-Renaud, Young-Key. 1986. *Studies in Korean linguistics*. Seoul: Hanshin.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Kochetov, Alexei & Marianne Pouplier. 2008. Phonetic variability and grammatical knowledge: An articulatory study of Korean place assimilation. *Phonology* 25. 399–431.

Kohler, Klaus J. 1990. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. Hardcastle & A. Marchal (eds.) *Speech production and speech modelling*, 69–92. Dordrecht: Kluwer.

Kohler, Klaus J. 1991a. The phonetics/phonology issue in the study of articulatory reduction. *Phonetica* 48. 180–192.

Kohler, Klaus J. 1991b. The organization of speech production: Clues from the study of reduction processes. *Proceedings of the 12th International Congress of Phonetic Sciences*, vol. 1, 102–106. Aix-en-Provence: Université de Provence.

Kohler, Klaus J. 1992. Gestural reorganization in connected speech: A functional viewpoint on "articulatory phonology." *Phonetica* 49. 205–211.

Krueger, John R. 1962. *Yakut manual*. Bloomington: Indiana University.

Ladefoged, Peter & Ian Maddieson. 1986. Some of the sounds of the world's languages. *UCLA Working Papers in Phonetics* 64. 1–137.

Lamontagne, Greg. 1993. Syllabification and consonant cooccurrence conditions. Ph.D. dissertation, University of Massachusetts, Amherst.

Lodge, Ken. 1986. Allegro rules in colloquial Thai: Some thoughts on process phonology. *Journal of Linguistics* 22. 331–354.

Lodge, Ken. 1992. Assimilation, deletion paths and underspecification. *Journal of Linguistics* 28. 13–52.

Lombardi, Linda. 1995. Laryngeal neutralization and syllable wellformedness. *Natural Language and Linguistic Theory* 13. 39–74.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17. 267–302.

Lombardi, Linda (ed.) 2001a. *Segmental phonology in Optimality Theory: Constraints and representations*. Cambridge: Cambridge University Press.

Lombardi, Linda. 2001b. Why Place and Voice are different: Constraint-specific alternations in Optimality Theory. In Lombardi (2001a), 13–45.

McCarthy, John J. 2008. The gradual path to cluster simplification. *Phonology* 25. 271–319.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, 249–384. Amherst: GLSA.

Mohanan, K. P. 1993. Fields of attraction in phonology. In John A. Goldsmith (ed.) *The last phonological rule: Reflections on constraints and derivations*, 61–116. Chicago & London: University of Chicago Press.

Padgett, Jaye. 1995. Partial class behavior and nasal place assimilation. In Keiichiro Suzuki & Dirk Elzinga (eds.) *Proceedings of the 1995 Southwestern Workshop on Optimality Theory (SWOT)*, 145–183. Tucson: Department of Linguistics, University of Arizona (ROA-113).

Padgett, Jaye. 2002. Russian voicing assimilation, final devoicing, and the problem of [v]. Unpublished ms., University of California, Santa Cruz (ROA-528).

Payne, David L. 1981. *The phonology and morphology of Axininca Campa*. Austin, TX: Summer Institute of Linguistics.

Prince, Alan & Paul Smolensky 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Rehg, Kenneth L. & Damien G. Sohl. 1981. *Ponapean reference grammar*. Honolulu: University of Hawai'i Press.

Rice, Keren. 1989. *A grammar of Slave*. Berlin & New York: Mouton de Gruyter.

Rischel, J. 1974. *Topics in Greenlandic phonology: Regularities underlying the phonetic appearance of wordforms in a polysynthetic language*. Copenhagen: Akademisk Forlag.

Rubach, Jerzy. 1990. Final devoicing and cyclic syllabification in German. *Linguistic Inquiry* 21. 79–94.

Rubach, Jerzy. 1996. Nonsyllabic analysis of voice assimilation in Polish. *Linguistic Inquiry* 27. 69–110.

Sapir, J. David. 1965. *A grammar of Diola-Fogny*. Cambridge: Cambridge University Press.

Schachter, Paul & Victoria P. Fromkin. 1968. A phonology of Akan: Akuapem, Asante, Fante. *UCLA Working Papers in Phonetics* 9.

Seo, Misun. 2003. A segment contact account of the patterning of sonorants in consonant clusters. Ph.D. dissertation, Ohio State University.

Son, Minjung. 2008. The nature of Korean place assimilation: Gestural overlap and gestural reduction. Ph.D. dissertation, Yale University.

Son, Minjung, Alexei Kochetov & Marianne Pouplier. 2007. The role of gestural overlap in perceptual place assimilation in Korean. In Jennifer Cole & José Ignacio Hualde (eds.) *Laboratory phonology* 9, 507–534. Berlin & New York: Mouton de Gruyter.

Steriade, Donca. 1993. Neutralization and the expression of contrast. Paper presented at the 24th Annual Meeting of the North East Linguistic Society, University of Massachusetts Amherst.

Steriade, Donca. 1995. Positional neutralization. Unpublished ms., University of California, Los Angeles.

Steriade, Donca. 1999. Phonetics in phonology: The case of laryngeal neutralization. *UCLA Working Papers in Linguistics 2: Papers in Phonology* 3. 25–146.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*, 219–250. San Diego: Academic Press.

Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Studies in honor of Paul Kiparsky*, 151–179. Cambridge, MA: MIT Press.

Stevens, Kenneth N. & Sheila Blumstein. 1975. Quantal aspects of consonant production and perception: A study of retroflex consonants. *Journal of Phonetics* 3. 215–233.

Teoh, Boon Seong. 1988. Aspects of Malay phonology revisited: A non-linear approach. Ph.D. dissertation, University of Illinois.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany: State University of New York Press.

Webb, Charlotte. 1982. A constraint on progressive consonantal assimilation. *Linguistics* 20. 309–321.

Wheeler, Max W. 2005. Voicing contrast: Licensed by prosody or licensed by cue? Paper presented at the 13th Manchester Phonology Meeting (ROA-769).

Wilson, Colin. 2001. Consonant cluster neutralisation and targeted constraints. *Phonology* 18. 147–197.

Wright, Richard. 1996. Consonant clusters and cue preservation in Tsou. Ph.D. dissertation, University of California, Los Angeles.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Hayes *et al.* (2004), 34–57.

Yip, Moira. 1991. Coronals, consonant clusters, and the coda condition. In Carole Paradis & Jean-François Prunet (eds.) *The special status of coronals: Internal and external evidence*, 61–78. San Diego: Academic Press.

Zhang, Jie. 2004. The role of contrast-specific and language-specific phonetics in contour tone distribution. In Hayes *et al.* (2004), 157–190.

Zoll, Cheryl. 1998. Positional asymmetries and licensing. Unpublished ms., MIT (ROA-282).

47 Initial Geminates

Astrid Kraehenmann

1 Introduction

There are a number of different ways in which languages make phonemic differences between consonant segments. One, for example, is on the basis of phonological features. The groups of sounds in (1a) differ in terms of voicing (CHAPTER 69: FINAL DEVOICING AND FINAL LARYNGEAL NEUTRALIZATION), the ones in (1b) in terms of continuancy (CHAPTER 13: THE STRICTURE FEATURES), and the ones in (1c) in terms of place of articulation (labial, coronal, dorsal) (CHAPTER 22: CONSONANTAL PLACE OF ARTICULATION), etc.

(1) a. /p t k f s x/ *vs.* /b d g v z ɣ/
b. /p t k b d g/ *vs.* /f s x v z ɣ/
c. /p b f v/ *vs.* /t d s z/ *vs.* /k g x ɣ/

Another way in which consonants may contrast is on the basis of inherent prosodic structure such as sound length and/or weight.[1] This distinction is also called a difference in quantity (see also CHAPTER 37: GEMINATES; CHAPTER 57: QUANTITY-SENSITIVITY). Thus the groups of sounds in (2) differ in terms of quantity, which is generally reflected (a) phonetically as long *vs.* short articulatory duration, (b) phonologically as heavy *vs.* light, and (c) structurally in the way they are associated to syllabic and higher-level prosodic structure.

(2) /pp tt kk ff ss xx/ *vs.* /p t k f s x/

Cross-linguistically, there are some constraints and implicational relationships on the existence of such quantity contrasts. As will be shown in the next section, initial geminates – i.e. geminates that contrast with singletons at the beginning of (lexical) words – are a special case, because they are even rarer than medial geminates. In the subsequent sections we will discuss in turn some issues regarding the phonological representation, prosodic behavior, phonetic properties, perception, and word-edge effects of initial geminates. These issues will provide some interesting insight on why initial geminates are so special.

[1] In this chapter, length/quantity is transcribed by a sequence of identical phonetic symbols.

2 Typological issues

Not many languages are known to have a lexical quantity contrast at word edges as well as word-medially. As an illustration we give examples from Thurgovian Swiss German (Indo-European)[2] in (3), which are representative of the phonological quantity distinction common to the majority of Swiss German dialects.

(3) *Swiss German quantity contrast*

a.	initial	/ppaaʀ/	'pair'	*vs.*	/paaʀ/	'bar'
		/ttaŋkx/	'tank'	*vs.*	/taŋkx/	'thank'
		/kkaaʀ/	'coach'	*vs.*	/kaaʀ/	'cooked'
b.	medial	/vappə/	'crest'	*vs.*	/vapə/	'honeycomb'
		/mattə/	'mat'	*vs.*	/matə/	'maggot'
		/makkə/	'fault'	*vs.*	/makə/	'stomach'
c.	final	/alpp/	'alp'	*vs.*	/xalp/	'calf'
		/vɛʀtt/	'value'	*vs.*	/hɛʀt/	'hearth'
		/maʀkk/	'marrow'	*vs.*	/aʀk/	'bad'

In order to eventually get a better feel for the typological status of initial geminates such as the ones in (3a), it is worth taking a look at the established typological facts of geminates in general. Surprisingly little comprehensive work has been done in this area to date.

If we consult the largest extant phonological database – the UCLA Phonological Segment Inventory Database (UPSID), which is an extended and revised online version of Maddieson (1984), listing the phonemes of 451 languages – only 12 languages are coded for having at least one geminate consonant segment. But not all of them have a quantity contrast to speak of. There are only seven (less than 2 percent of the whole sample) if we discount those languages which – based on the inventories given – have either only a single geminate–singleton pair (!Xu (Khoisan), Iraqw (Afro-Asiatic), and Telugu (Dravidian)) or only geminates that come without a singleton counterpart (Inuit (Eskimo-Aleut) and Trumai (South American isolate)). Those seven are: the North Caucasian Archi, Avar, and Lak; Ocaina (Witotoan), Waray (Austronesian), Wichita (Caddoan), and Wolof (Niger-Congo).

While the general picture that emerges seems to be that quantity contrasts are rather uncommon, it is doubtful that they are quite as uncommon as that. For example, the database also contains languages such as the Indo-European Bengali, Breton, and Norwegian, as well as Finnish (Uralic), for which there is general consensus among linguists that all have quantity contrasts. However, they are not coded as such in UPSID. Thus the typological interpretation of the UPSID data – not only with respect to quantity – must be taken with the necessary pinch of salt (see also Simpson 1999, among others).

One particular piece of information missing in UPSID but of primary interest for our purposes is where within a word the geminates may occur. This information is given, albeit rather implicitly, in the survey of 63 languages with geminates by

2 Language family classification in all cases follow Lewis (2009).

Thurgood (1993). In his sample, Thurgood identifies a number of frequency facts and prerequisite conditions for geminates and proposes some implicational tendencies. For example, he finds preferences for certain places of articulation (alveolar > labial > velar > glottal)[3] and certain prosodic environments (post-tonic > pre-tonic/unstressed; after short vowel > after long vowel). But most importantly, he establishes that geminates are most favored in intervocalic position or, more specifically, if preceded by a short (and stressed) vowel and followed by another short vowel (Thurgood 1993: 129). Although not explicitly stated, these flanking vowels preferably also belong to the same word. This means that if a language has any quantity contrast at all, it will be word-medially. Thus, the existence of initial and/or final geminates implies the existence of medial geminates (see also Muller 1999). While Thurgood (1993) lists the three Austronesian counter-examples Sa'ban, Kelantan Malay, and Pattani Malay, no further explanation is given as to how it could be that they exclusively have initial geminates, which is an interesting issue that we will come back to shortly. Thurgood does not mention word-final geminates separately, let alone whether or not the languages under investigation allow word-final (i.e. syllable-final) consonants in the first place. But, based on the preference factors he identifies, we can conclude that a quantity contrast is cross-linguistically most common medially, somewhat less common finally, and least common initially.

Is there also a specific consonant class that stands out as the most preferred to display a quantity contrast? Based purely on the sonority of consonants (CHAPTER 49: SONORITY), Morén (1999: 110) answers this question negatively. He finds languages with quantity distinctions in stops, continuants, and sonorants alike (Hungarian (Uralic), Brahui (Northern Dravidian), Italian, Baloch, Gajarati (Indo-European)), only in continuants (Tartar (Altaic)), only in sonorants (Hausa (Afro-Asiatic)), and only in obstruents (Chechen, Lak (North Caucasian)), but also languages that allow geminate stops and sonorants to the exclusion of continuants (Kurdish (Indo-European)). He does not list any languages either allowing only stop geminates or excluding only stop geminates, both of which he considers to be accidental gaps. In comparison, Thurgood's (1993) generalizations are a bit more fine-grained for place of articulation and voicing within the major classes of obstruents and sonorants. He observes that geminate stops seem to be a prerequisite for geminate affricates to occur, and presents data that suggest a voiceless alveolar stop or fricative /tt ss/ or an alveolar nasal or liquid /nn ll/ to be good candidates for the prototypical geminate.

To conclude our look at geminates in general, it is very surprising to notice that 46 of the 63 languages (= 73 percent) in Thurgood's (1993) sample are also listed in UPSID, yet obviously with different phoneme analyses from the ones consulted by Thurgood. As regards language diversity, six language families figure very prominently, accounting for more than 50 percent of the whole set: namely, in descending frequency, Afro-Asiatic (10 languages), Indo-European (7), Austronesian (6), Altaic (3), Dravidian (3), and Uralic (3).

[3] This means that Thurgood (1993) finds, on the one hand, that alveolar geminates are the most frequent and glottal geminates the least frequent and, on the other hand, that if a language has, for example, velar geminates it also tends to have labial and alveolar ones. He presents his data as an overview of the phonological systems and does not use them to make any claims about preferences for the *emergence* of geminates.

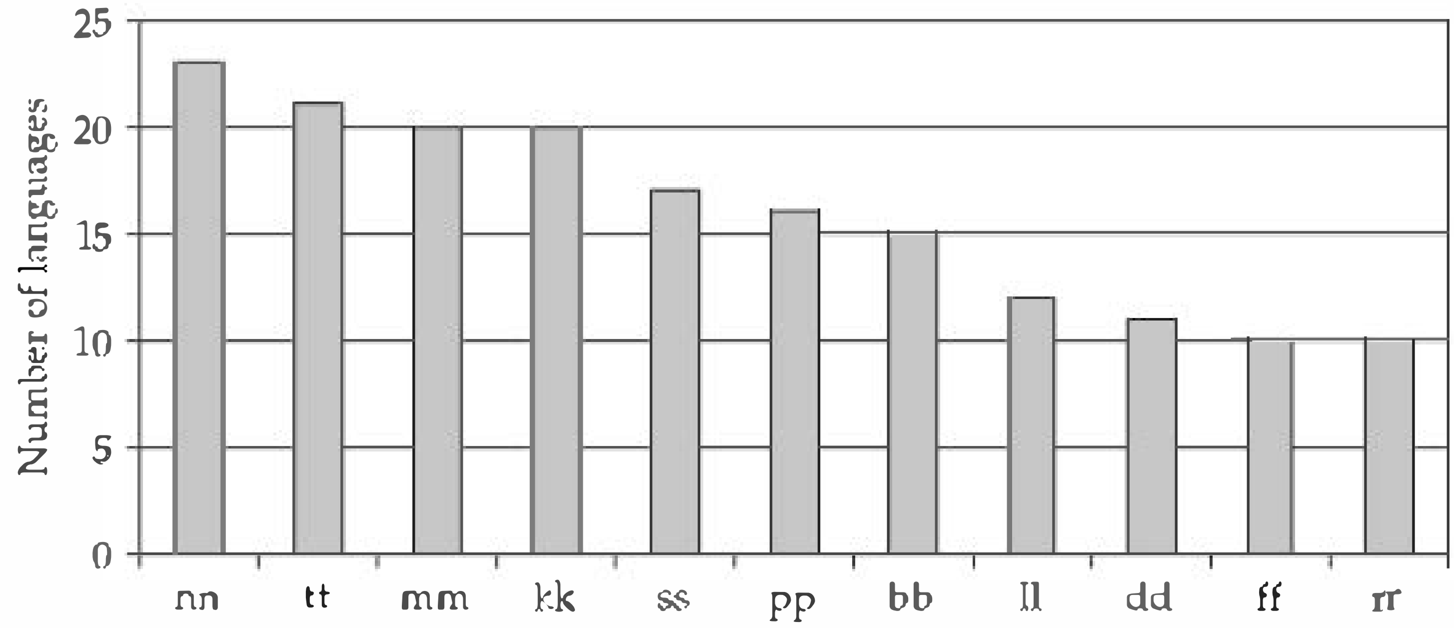


Figure 47.2 The 10 most frequent geminate phonemes in the 29 languages listed in Muller (2001: 204–233)

Looking at the frequency of initial geminate phonemes, there are some interesting observations to be made with respect to major consonant class, manner, place of articulation, and voicing.[7] But before looking at these factors, let us first consider the 10 most frequent geminates in these languages, as illustrated in Figure 47.2.

The nasals /nn mm/ and the voiceless stops /tt kk pp/ are among the most commonly occurring geminates. The voiceless fricatives /ss ff/, the voiced stops /bb dd/, and the liquids /ll rr/ are next on the top-10 frequency list. Thus, within the sonorants /nn/ is most universally present, and within the obstruents it is /tt/ for the stops and /ss/ for the fricatives. Very significantly, all these sounds follow the universal preference patterns (cf. Ladefoged and Maddieson 1996; Ladefoged 2001): they are coronal as well as voiced if sonorant and voiceless if obstruent, which is the less marked voicing characteristic within each respective class.

Considering the entire set of initial geminates, stops and fricatives together make up almost two-thirds, with nasals taking the lion's share in the last third (see Figure 47.3). Geminate fricatives imply geminate stops, except for Hatoma, which has /ff/ and /ss/, but does not allow any geminate stops or affricates. Overall, geminate affricates are least frequent and only occur in conjunction with

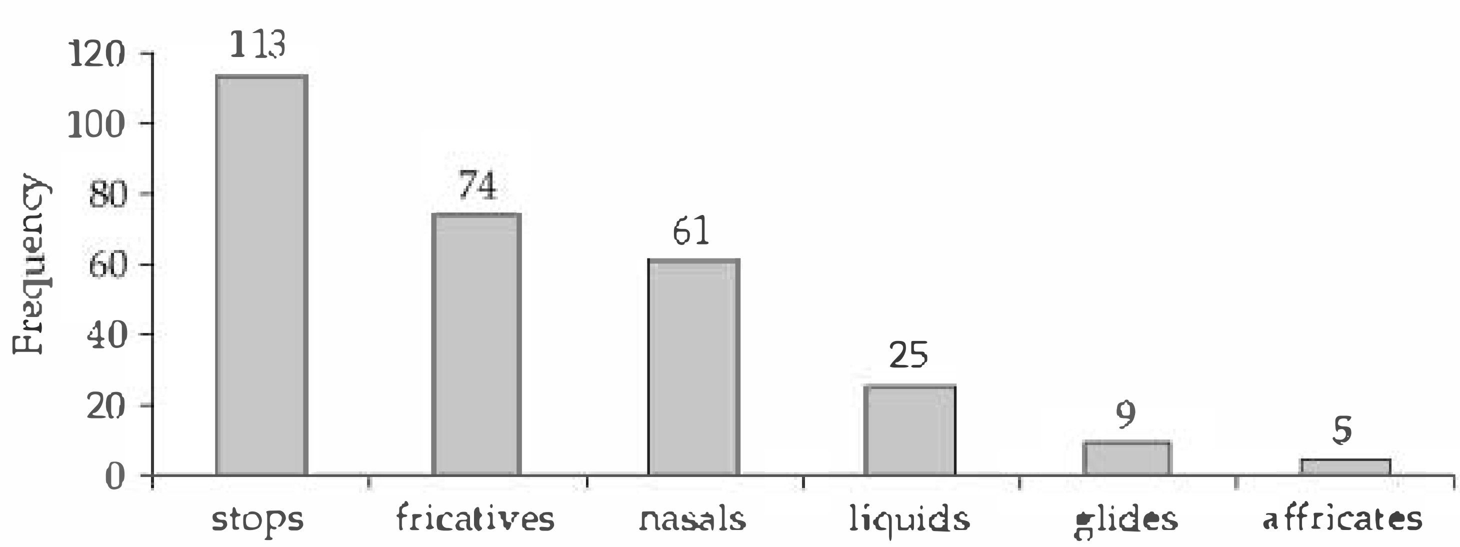


Figure 47.3 Initial geminates by manner (data from Muller 2001: 204–233)

[7] In this chapter, I refer only to *type* frequency of phonemes and members of natural classes. I have no information about *token* frequency of the sounds in the respective languages.

both stops and fricatives. As for the sonorants, nasals are a prerequisite for any other type of sonorant geminate (CHAPTER 8: SONORANTS). The only exception is Yapese, which has only a coronal lateral /ll/ and no nasal or other sonorant geminate, although it has labial, coronal, and dorsal nasal singletons.

In the complete set of languages, there are 22 that have both obstruent and sonorant geminates. However, Circassian, Lak, Bernese, and Thurgovian Swiss German allow only obstruents, while Kiribati, Piro, and Ponapean allow only sonorants. In the latter group, all three languages have geminate nasals, most often a coronal /nn/.

Figure 47.4 illustrates the frequencies of the different places of articulation[8] in the sample. As expected, the predominance of coronals over labials and dorsals mirrors the general attested cross-linguistic preferences. Finally, regarding voicing where it is contrastive, namely within the obstruents, voiceless geminates outnumber voiced ones by a little over 100 percent (see Figure 47.5). There are seven languages without voicing distinction, for which only "voiceless" geminates are listed: Atapec Zapotec, Trukese, Hatoma, Lak, Leti, Puluwat, and Thurgovian Swiss German. In three languages, there seem to be "voiced" geminates without any "voiceless" counterpart: Lugbara has /bb dd/ but no */pp tt/, Ngada has /bb/ but no */pp/, and Yapese has /gg/ but no */kk/. However, in all three cases, these phonemes are the only obstruent geminates in the respective systems. Dobel is a slightly different case, in that it lacks */pp/, with /bb/ being part of a bigger set, which includes voiceless /tt kk ʔʔ ss/.

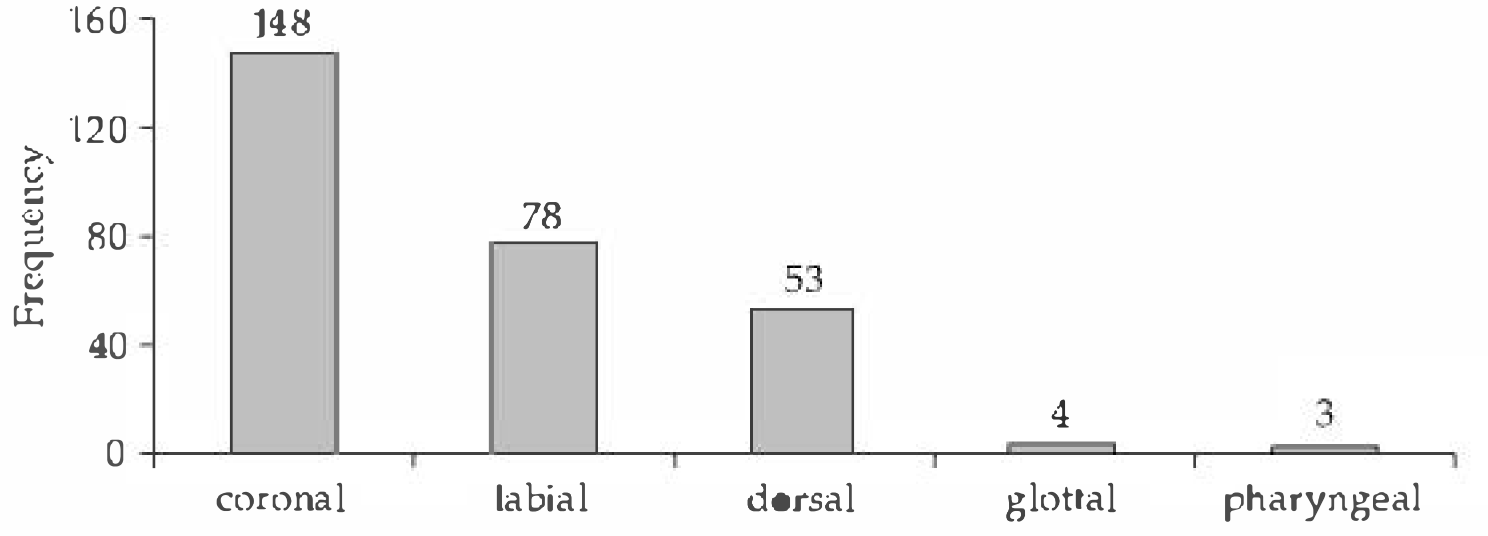


Figure 47.4 Initial geminates by place of articulation (data from Muller 2001: 204–233)

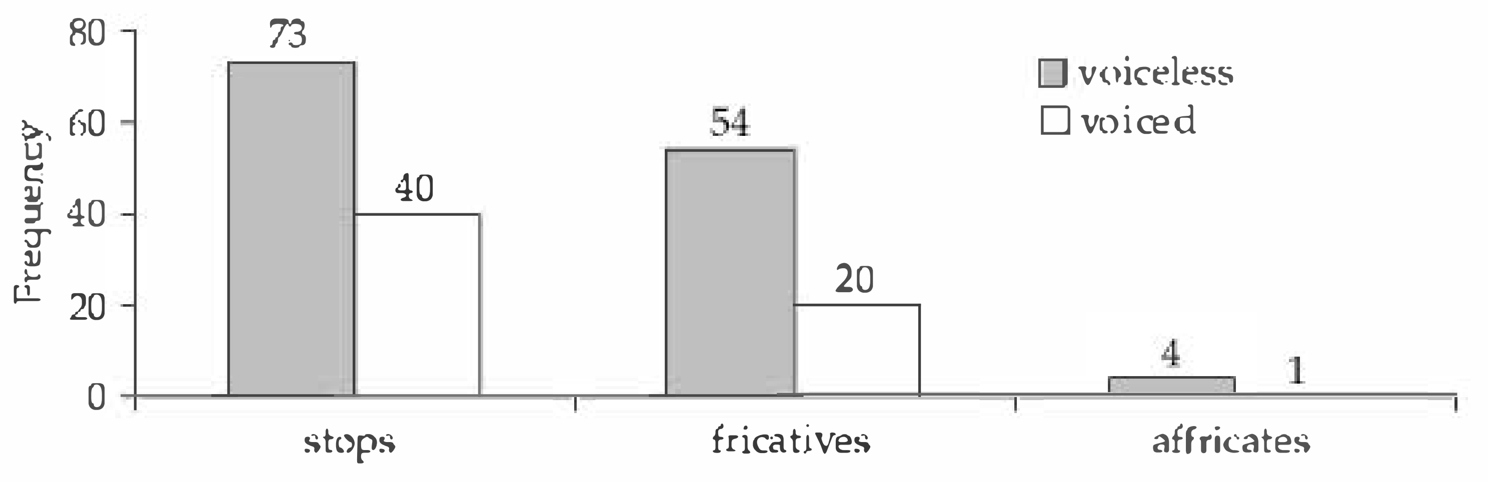


Figure 47.5 Initial obstruent geminates by manner and voicing (data from Muller 2001: 204–233)

[8] In Muller's (2001) terminology, labials comprise bilabials and labio-dentals, coronals refer to (inter)dentals, alveolars, and palatals, and dorsals are velars.

a preceding syllable and is the onset of a following syllable; see the structures in (5). As a rule, "intervocalic" is interpreted to mean "within one and the same word."

(5) *Syllabification of geminates and singletons*[9]

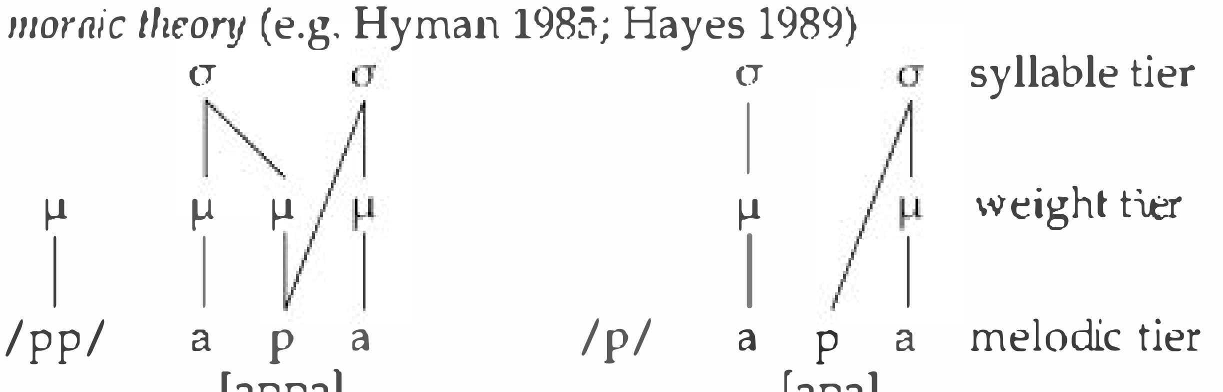
a. *ambisyllabic link: geminate* b. *tautosyllabic link: singleton*

While there is considerable agreement as to the ambisyllabic nature of geminates (but see Topintzi 2008 for a different view), theories vary in what unit they posit to link up with these syllable constituents. These assumptions have far-reaching consequences for how geminates are phonologically represented, especially so if the goal is to establish a single representation that is to capture not only language-specific but also cross-linguistic patterns of geminates (e.g. Selkirk 1990; Tranel 1991; Broselow 1995; Broselow *et al.* 1997; Ham 1998; Davis 1999a, 1999b, 2003; Muller 2001; Curtis 2003; Kraehenmann 2003; Topintzi 2008; Ringen and Vago 2010; among others).

For convenience, we briefly sketch the main tenets of the two competing theories at this point.[10] In moraic theory, a geminate is associated with a unit of weight, the mora – see (6a) – and should therefore participate in weight-related phenomena such as stress patterns, compensatory lengthening, and minimal word conditions. In skeletal theory (CHAPTER 54: THE SKELETON), though, a geminate is doubly linked to the timing tier – see (6b) – and is prosodically long without pre-association to any weight tier. As a consequence, weightless geminates pose a challenge for the former theory, moraic geminates for the latter.

(6) *Underlying representation and syllabification*

a. *moraic theory* (e.g. Hyman 1985; Hayes 1989)



[9] The segmental labels linking up to syllable nodes are shorthand for the root nodes representing these segments.

[10] See Curtis (2003) for a concise survey of the different models of geminate representation. CHAPTER 37: GEMINATES also discusses in detail the basic controversy between the weight account and the timing account, as well as the Composite Model (e.g. Hume *et al.* 1997; Muller 2001; Curtis 2003; Kraehenmann 2003; Baker 2008). A full exploration of this theory goes beyond the scope of this chapter. Suffice it to say, however, that it is a combination of the other two models: geminates are associated to two timing positions and weight, if it plays a role at all, is a consequence of syllabification, mainly one of Weight-by-Position (Hayes 1989). One worry with this view is that it runs the risk of overgeneration, since it allows for all types of combinations between length and weight, possibly even unattested ones.

et al. (1997: 393), Kraehenmann (2003: 26), and Topintzi (2008: 180), the mora cannot count for weight purposes by any measure of analysis when not associated to higher level prosodic structure. Finally, Topintzi's (2008) proposal for Trukese and Pattani Malay – and moraic onsets in general – is the one illustrated in (8d), in which the geminate mora is directly associated with the syllable node to which the following vowel belongs.

This tautosyllabic representation is striking for at least two reasons. First, it treats the geminate like a regular onset consonant by making it entirely the initial constituent within the first syllable. There is no ambisyllabic double linking, which is characteristic of geminates in all other proposals. Second, the geminate carries a mora equal in all respects to the mora of the syllable nucleus. The implications of such structures are, on the one hand, that moraic onsets are expected to occur not only word-initially but also word-medially and, on the other hand, that, to an even greater extent, moras play the double role of a unit of weight *and* length.

In support of the first implication, Topintzi (2008: 170–172) presents some intriguing data from Marshallese stress patterns and Trique compensatory lengthening processes which suggest an analysis of medial geminates with tautosyllabification (see also CHAPTER 55: ONSETS; Topintzi 2010). In the majority of cases, though, medial geminates must be analyzed ambisyllabically as illustrated in (8), cases like Marshallese being the exception rather than the rule. Regarding the length issue, Topintzi replaces the basic tenet of traditional moraic theory that onset consonants are never moraic with another, based on findings by Hubbard (1994) and Ham (1998), namely "that moras are allocated a minimum target duration" (2008: 174), the controversial proposal being that surface phonetic duration is in *direct* relation to underlying prosodic length.

To summarize this section, there are a number of different challenges that initial geminates pose for the existing theoretical models of representation. Like medial geminates, initial geminates do not present a uniform behavior across languages. In some cases it is clear that they fully participate in weight-sensitive processes (e.g. Trukese), in other cases weight is not involved at all (e.g. Leti). If, in the interest of a universal geminate representation, geminates are *not* underlyingly moraic, how do we explain the Trukese cases? And vice versa, if geminates *are* underlyingly moraic, how do we explain the Leti cases? Moreover, how can such Leti-type geminates occupy the syllable onset, *the* non-weight position? It is evident that strict adherence to the traditional concepts of the two theories is not very fruitful. A combination of the two approaches seems rather more promising but the proper balance still needs to be established (see also note 10).

4 Phonetic properties and perception

In terms of phonetic properties, initial geminates are a model case for illustrating the interdependence between articulation, acoustics, and perception because not only segmental features, such as place of articulation, manner, etc., are involved, but prosodic features representing duration also play a role.

Temporal characteristics have received most attention in phonetic studies of geminates to date. Although there is still some discussion on whether geminates are to be considered long or tense consonants, it is generally accepted that the manifestations of the tense articulation of geminates. One example is the finding that the normalized Root Mean Square (RMS) amplitude of the stop release is slightly higher for geminates than for singletons for all but one of the five speakers, who shows the reverse pattern. Another example would be the fact that geminates tend to devoice for aerodynamic reasons (cf. Ohala 1983), depending in degree on factors such as speaker, place of articulation, and position within the word. Also, voiced stop singletons have a weakish propensity to lenite, i.e. become fricatives, while geminates never do. As for durational correlates apart from the primary cue, Ridouane (2007: 129) finds that the duration of the stop release – also called "voice onset time" (VOT) or "after closure time" by other authors – shows no difference in voiceless stops but is longer for voiced geminates than for singletons.[16] None of these additional correlates, however, has yet been tested for its significance in perception. The jury is still out on whether Tashlhiyt Berber native listeners can and do rely on these cues when the primary duration cue is not available.

While VOT plays no role for voiceless stops of Tashlhiyt Berber, it is a temporal measure correlating with the quantity distinction of these sounds in Cypriot Greek. Parallel to Tserdanelis and Arvaniti's (2001) findings on medial geminates, Muller (2001, 2003) establishes that VOT is also significantly longer for geminates than for singletons in word-initial position. Like native listeners of Pattani Malay, native listeners of Cypriot Greek can distinguish voiceless stop geminates from singletons at the beginning of an utterance. Muller hypothesizes that the subjects of her very basic perception experiment thus utilize VOT as one of possibly many secondary cues to the phonological quantity distinction.

Finally, as well as Pattani Malay, Tashlhiyt Berber, and Cypriot Greek, Swiss German initial geminates have also received some attention in the literature, in particular in the Thurgovian dialect of Swiss German (Kraehenmann 2001, 2003, 2009; Kraehenmann and Lahiri 2008). Kraehenmann (2001, 2003) confirms earlier findings on other dialects, for example by Enstrom and Spörri-Bütler (1981) and Fulop (1994), that VOT does not participate in the quantity opposition: there is no difference in the duration of the stop release for geminates and singletons, be they initial, medial, or final. Since voiceless stops are the only consonants occurring in initial length opposition, the question is again whether the phonological difference is enhanced by cues other than the primary one. While lacking any corroborating perception evidence, Fulop (1994) claims to have found potential secondary non-temporary cues, namely increased intensity, movement, and clarity of post-release sonorant formants above F2 for geminates in comparison to singletons. However, anecdotal evidence in Moulton (1979) and the results of a pilot perception study reported in Kraehenmann (2003) call into question whether listeners can make use of these acoustic differences. Their listeners seemed unable to recover the quantity contrast in utterance-initial context. A crucial fact to verify especially in this case is whether the difference is actually produced by the speakers or whether contrast neutralization occurs. Using electropalatography (EPG), Kraehenmann and Lahiri (2008) set out to do just that in an articulatory study. They find that geminates are indeed articulated significantly longer (by about 85 msecs) than singletons even in absolute utterance/phrase-initial position (see Figure 47.6).

16 Exactly the same finding is reported in Mikuteit and Reetz (2007) for voiceless and voiced stops in East Bengali.

201) rightly speculates that the segmental inventory of the language in question may affect the "availability" of such acoustic cues. But the complexity of syllable structures and the degree of the functional load of the contrast must also play a significant role.

5 Edge effects

In the field of the phonology–phonetics interface, one area of investigation focuses on how the articulation of speech sounds is influenced by the position within prosodic structure they occur in. A number of effects have been established that involve the edges of prosodic domains such as the syllable (CHAPTER 33: SYLLABLE-INTERNAL STRUCTURE), the phonological word (CHAPTER 51: THE PHONOLOGICAL WORD), the phonological phrase (CHAPTER 50: TONAL ALIGNMENT), etc. What is most interesting and relevant with respect to the topic of this chapter is the fact that these domain edges – beginnings in particular – have a significant effect on sound duration. More specifically, the length of segments or particular properties of segments is greater the higher up in the prosodic hierarchy their edges are (e.g. Fujimura 1990; Byrd *et al.* 2005).

This has become known in the literature as *initial articulatory strengthening* or *prosodic strengthening*. It is important to note at this point, however, that most studies so far have been on sounds that are not in a quantity opposition. For example, in their seminal EPG study, Fougeron and Keating (1997) tested the English dummy syllable *no* in trisyllabic words with initial, medial, and final stress embedded in a carrier phrase. They found that the amount of linguopalatal contact as well as the acoustic duration of the syllable-initial nasal increased with each increasing prosodic level. Fougeron and Keating conclude that these "more extreme articulations" (1997: 3738) perform an important function for perception. On the one hand, they facilitate segmentation into higher prosodic domains, but particularly into words. On the other hand, they enhance the acoustic cues for identifying sounds and thus assist lexical access. Although the correlation between the amount of articulator contact and acoustic duration is reported as being only weak, it is there nonetheless. Because EPG studies are still few and far between, we will also use this acoustic finding as the main basis for our remaining discussion.

In a phonetic study on real words, Cho and McQueen (2005) investigated word-initial Dutch alveolar stops /t d/ in different phrasal and word-stress contexts. As expected, the closure durations of both voiced and voiceless stops were significantly longer in stronger prosodic positions. In addition, however, the VOTs of voiceless stops were cumulatively shorter in the same prosodic positions, i.e. the opposite of the findings for English /t/. Cho and McQueen argue in essence that the difference is due to the difference in how the phonological encoding of the voicing contrast is phonetically enhanced in the two languages. That is, Dutch /t/, as the phonologically unmarked member of the pair (voiceless *vs.* voiced), has the phonetic (default) specification [–spread glottis], which is enhanced by shorter aspiration; in contrast, English /t/, as the phonologically marked member of the pair (aspirated *vs.* non-aspirated), has the phonetic specification [+spread glottis], which is enhanced by longer aspiration. One of their main claims is thus that prosodic strengthening acoustically amplifies the difference between sounds in

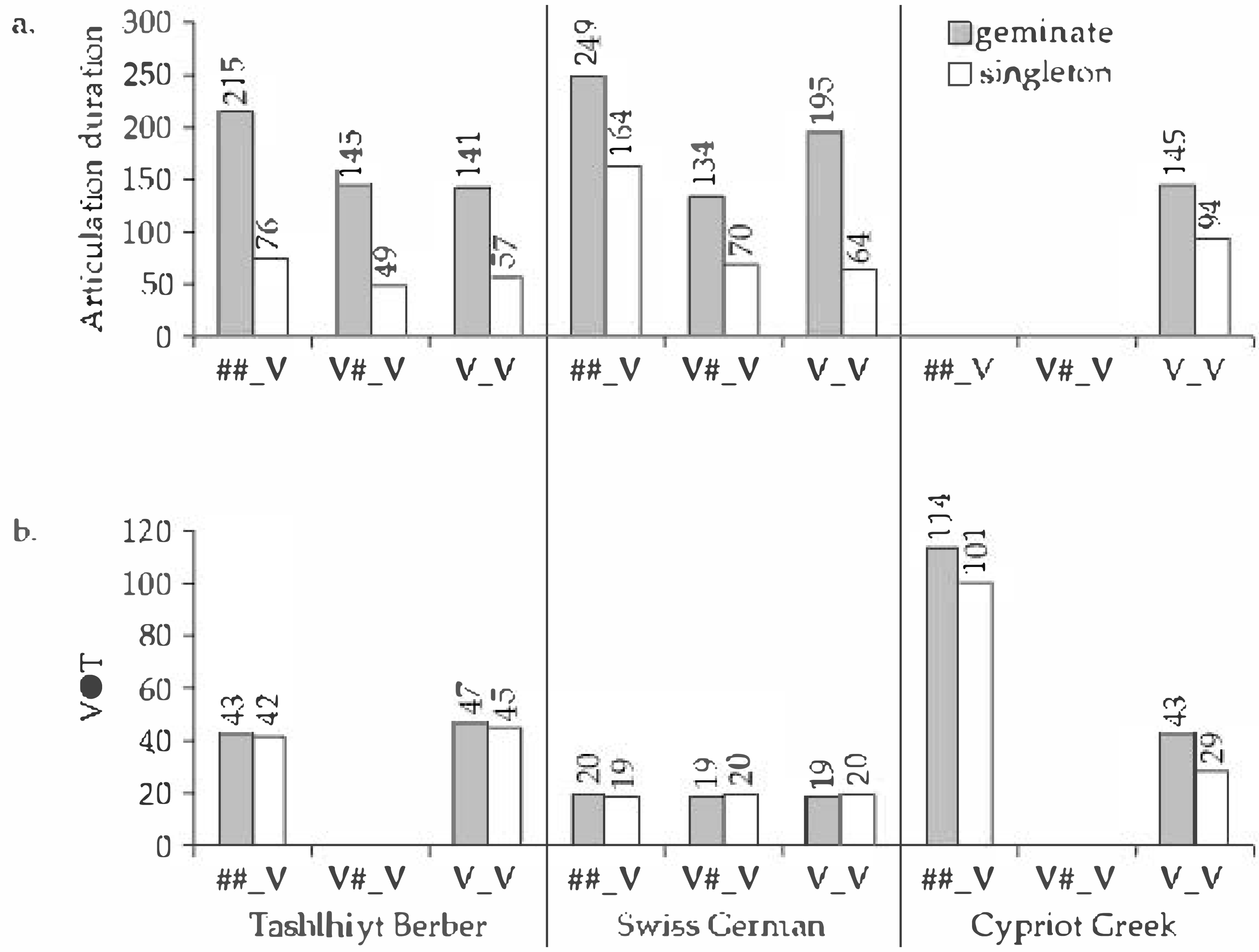


Figure 47.7 Articulation/closure duration[a] and VOT in msec for initial geminates and singletons in phrase-initial (## __ V) and phrase-medial post-vocalic (V# __ V) context and medial geminates and singletons in intervocalic (V __ V) context in three languages: Tashlhiyt Berber (Ridouane 2007: 128–129), Swiss German (Kraehenmann and Lahiri 2008: 4450–4451), and Cypriot Greek (Muller 2001: 29, 32)

[a] The measures of the medial geminates and singletons in Tashlhiyt Berber and Cypriot Greek are acoustic (closure duration); all others are articulatory (linguo-palatal contact)

phonological contrast. With respect to VOT effects, this claim seems to be supported by studies on Korean word-initial voiceless stops (Jun 1993; Cho and Keating 2001; Keating *et al.* 2003), which found that VOT measures were longest phrase initially, somewhat shorter word-initially within a phrase, and shortest word-medially within a phrase.

The question we want to ask now is whether articulatory strengthening also applies to word-initial geminates and singletons and, if so, whether it shows different characteristics. There are no studies comparable to Fougeron and Keating (1997) and their later work that directly address this issue in any depth. However, some generalizations can be drawn from the studies on voiceless stops in Cypriot Greek (Muller 2001), Tashlhiyt Berber (Ridouane 2007), and Swiss German (Kraehenmann and Lahiri 2008), by comparing medial *vs.* initial geminates and – in the latter two cases – initial geminates in different phrasal contexts.

The measures we are therefore most interested in are articulatory/acoustic closure duration and VOT. To start with the latter, we would not expect VOT to become a differentiating characteristic for initial geminates and singletons if it does not play any role for the medial contrast. This is indeed what we find. As can be seen in Figure 47.7b, Tashlhiyt Berber VOTs are only slightly but not significantly shorter when syllable, word, phrase, and utterance boundaries coincide (## __ V), compared to word-medially (V __ V). For the Swiss German data, the non-effect is even more pronounced, with basically identical measures in all three contexts. If anything had changed, longer measures would have been expected in the vocalic context (V# __ V) for initial stops and the longest ones at the phrase edge (## __ V), at least for geminates, if not for singletons. In contrast to these two languages, there are drastically longer measures in the Cypriot Greek data, where VOT is a secondary cue: geminate VOTs are longer by 71 msecs (265 percent), singleton VOTs by 72 msecs (348 percent). Although the 13 msecs difference in the phrase-initial context is still statistically significant, the differential has substantially decreased from 1.5 : 1 to 1.1 : 1, which goes against Cho and McQueen's (2005) claim of contrast amplification at increasing prosodic domain edges. Unfortunately, Muller (2001) does not provide any data for the contact/closure duration in utterance-initial or vocalic context. Thus we do not know how the primary cue is affected by prosodic strengthening.

This, however, is different for Tashlhiyt Berber and Swiss German (see Figure 47.7a). In Tashlhiyt Berber the trend seems to go in the expected direction in two of the three contexts. The geminate to singleton proportion is 2.5 : 1 word-medially and increases to 3 : 1 for word-initial stops after a vowel-final word. But at the utterance boundary, the highest prosodic domain, it decreases again to 2.8 : 1. In Swiss German we find exactly the opposite of what prosodic strengthening would predict, namely that the contrast magnitude decreases as the prosodic level increases. The difference between geminates and singletons is biggest in word-medial intervocalic context (syllable level, 3 : 1), somewhat smaller in word-initial intervocalic context (word level, 2 : 1), and smallest in utterance-initial context (phrase/utterance level, 1.5 : 1).

To conclude, there are some strengthening effects in languages with word-initial quantity contrast but only to the extent that the primary correlates lengthen substantially at the highest prosodic level, the utterance. However, this lengthening affects both geminates and singletons, resulting in contrast diminishment rather than augmentation. The segmental context, rather than the prosodic level, might therefore be the more reliable predictor of how the quantity contrast is realized, because intervocalic geminates – be they word-medial or phrase-medial word-initial – are actually not *at* a domain boundary, but contain one, because they straddle two syllables; see (6). The only time initial geminates occur at a domain edge is utterance-initially, which, as we have seen, is a very special context in many respects.

6 Conclusion

The field of word-initial quantity contrasts shows many faces and still awaits unveiling and scrutinizing investigations of a wealth of issues. A start has been made. Yet the existing disputes about phonological representation (length *vs.* weight), acoustic and articulatory properties, and perception are far from settled and promise to provide discussion material and theoretical arguments for years and decades to come. The historical question has only been hinted at in this chapter but constitutes another fertile ground and worthwhile area of research to be followed.

REFERENCES

Abramson, Arthur S. 1986a. Distinctive length in initial consonants: Pattani Malay. *Journal of the Acoustical Society of America* 79. 527–527.

Abramson, Arthur S. 1986b. The perception of word-initial consonant length: Pattani Malay. *Journal of the International Phonetic Association* 16. 8–16.

Abramson, Arthur S. 1987. Word-initial consonant length in Pattani Malay. In Thomas V. Gamkrelidze (ed.) *Proceedings of the 11th International Congress of Phonetic Sciences*, vol. 6, 68–70. Tallinn: Academy of Sciences of the Estonian SSR.

Abramson, Arthur S. 1991. Amplitude as cue to word-initial consonant length: Pattani Malay. *Proceedings of the 12th International Congress of Phonetic Sciences*, vol. 3, 98–101. Aix-en-Provence: Université de Provence.

Abramson, Arthur S. 1999. Fundamental frequency as cue to word-initial consonant length: Pattani Malay. In Ohala *et al.* (1999), 591–594.

Abramson, Arthur S. 2003. Acoustic cues to word-initial stop length in Pattani Malay. In Solé *et al.* (2003), 387–390.

Baker, Brett. 2008. *Word structure in Ngalakgan.* Stanford: CSLI.

Bender, Byron W., Ward H. Goodenough, Frederick H. Jackson, Jeffrey C. Marck, Kenneth L. Rehg, Ho-min Sohn, Stephen Trussel & Judith W. Wang. 2003. Proto-Micronesian reconstructions I. *Oceanic Linguistics* 42. 1–110.

Broselow, Ellen. 1995. Skeletal positions and moras. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 175–205. Cambridge, MA & Oxford: Blackwell.

Broselow, Ellen, Su-I Chen & Marie Huffman. 1997. Syllable weight: Convergence of phonology and phonetics. *Phonology* 14. 47–82.

Byrd, Dani, Sungbok Lee, Daylen Riggs & Jason Adams. 2005. Interacting effects of syllable and phrase position on consonant articulation. *Journal of the Acoustical Society of America* 118. 3860–3873.

Cho, Taehong & Patricia Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29. 155–190.

Cho, Taehong & James M. McQueen. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* 33. 121–157.

Clements, G. N. 1986. Compensatory lengthening and consonant gemination in LuGanda. In W. Leo Wetzels & Engin Sezer (eds.) *Studies in compensatory lengthening*, 37–77. Dordrecht: Foris.

Clements, G. N. & Samuel J. Keyser. 1983. *CV phonology: A generative theory of the syllable.* Cambridge, MA: MIT Press.

Curtis, Emily. 2003. Geminate weight: Case studies and formal models. Ph.D. dissertation, University of Washington.

Davis, Stuart. 1999a. On the representation of initial geminates. *Phonology* 16. 93–104.

Davis, Stuart. 1999b. On the moraic representation of underlying geminates: Evidence from prosodic morphology. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*, 39–61. Cambridge: Cambridge University Press.

Davis, Stuart. 2003. The controversy over geminates and syllable weight. In Féry & van de Vijver (2003), 77–98.

Davis, Stuart. Forthcoming. Quantity. In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.) *The handbook of phonological theory*. 2nd edn. Malden, MA & Oxford: Wiley-Blackwell.

Dmitrieva, Olga. 2009. Geminate typology and perception of consonant length: Experimental evidence from Russian. Paper presented at the 83rd Annual Meeting of the Linguistic Society of America, San Francisco.

Enstrom, Daly & Sonja Spörri-Bütler. 1981. A voice onset time analysis of initial Swiss-German stops. *Folia Phoniatrica* 33. 137–150.

Féry, Caroline & Ruben van de Vijver (eds.) 2003. *The syllable in Optimality Theory.* Cambridge: Cambridge University Press.

Fougeron, Cécile & Patricia Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101. 3728–3740.

Fujimura, Osamu. 1990. Methods and goals of speech production research. *Language and Speech* 33. 195–258.

Fulop, Sean A. 1994. Acoustic correlates of the fortis/lenis contrast in Swiss German plosives. *Calgary Working Papers in Linguistics* 16. 55–63.

Gordon, Matthew. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistics* 23. 595–653.

Ham, William. 1998. Phonetic and phonological aspects of geminate timing. Ph.D. dissertation, Cornell University.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce. 1999. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In Michael Darnell, Edith Moravcsik, Frederick Newmeyer, Michael Noonan & Kathleen Wheatley (eds.) *Functionalism and formalism in linguistics,* vol. 1: *General papers,* 243–285. Amsterdam & Philadelphia: John Benjamins.

Hubbard, Kathleen. 1994. Duration in moraic theory. Ph.D. dissertation, University of California, Berkeley.

Hume, Elizabeth, Jennifer Muller & Aone van Engelenhoven. 1997. Non-moraic geminates in Leti. *Phonology* 14. 371–402.

Hyman, Larry M. 1985. *A theory of phonological weight.* Dordrecht: Foris.

Jackson, Frederick H. 1984. The internal and external relationships of the Trukic languages of Micronesia. Ph.D. dissertation, University of Hawai'i.

Jensen, John T. 1977. *Yapese reference grammar.* Honolulu: University Press of Hawaii.

Jun, Sun-Ah. 1993. The phonetics and phonology of Korean prosody. Ph.D. dissertation, Ohio State University.

Keating, Patricia, Taehong Cho, Cécile Fougeron & Chai-Shune Hsu. 2003. Domain initial articulatory strengthening in four languages. In John Local, Richard Ogden & Rosalind Temple (eds.) *Phonetic interpretation: Papers in laboratory phonology VI,* 145–163. Cambridge: Cambridge University Press.

Kennedy, Robert. 2005. Reflexes of initial gemination in Western Micronesian languages. Paper presented at the 12th Meeting of the Austronesian Formal Linguistics Association, University of California, Los Angeles.

Keyser, Samuel J. & Kenneth N. Stevens. 2006. Enhancement and overlap in the speech chain. *Language* 82. 33–63.

Khattab, Ghada. 2007. A phonetic study of gemination in Lebanese Arabic. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences,* 153–158. Saarbrücken: Saarland University.

Kiparsky, Paul. 2003. Syllables and moras in Arabic. In Féry & van de Vijver (2003), 147–182.

Kraehenmann, Astrid. 2001. Swiss German stops: Geminates all over the word. *Phonology* 18. 109–145.

Kraehenmann, Astrid. 2003. *Quantity and prosodic asymmetries in Alemannic: Synchronic and diachronic perspectives.* Berlin & New York: Mouton de Gruyter.

Kraehenmann, Astrid. 2009. The perception of the word-initial quantity contrast in voiceless Swiss German stops. Paper presented at the 17th Manchester Phonology Meeting.

Kraehenmann, Astrid & Aditi Lahiri. 2008. Duration differences in the articulation and acoustics of Swiss German word-initial geminate and singleton stops. *Journal of the Acoustical Society of America* 123. 4446–4455.

Ladefoged, Peter. 2001. *Vowels and consonants: An introduction to the sounds of languages.* Malden, MA & Oxford: Blackwell.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages.* Oxford & Malden, MA: Blackwell.
Lahiri, Aditi & Jorge Hankamer. 1988. The timing of geminate consonants. *Journal of Phonetics* 16. 327–338.
Levin, Juliette. 1985. A metrical theory of syllabicity. Ph.D. dissertation, MIT.
Lewis, M. Paul (ed.) 2009. *Ethnologue: Languages of the world.* 16th edn. Dallas: SIL International. Available (August 2010) at www.ethnologue.com.
Maddieson, Ian. 1984. *Patterns of sounds.* Cambridge: Cambridge University Press.
Marti, Werner. 1985. *Berndeutsch-Grammatik.* Bern: Francke.
McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.
Mikuteit, Simone & Henning Reetz. 2007. Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language and Speech* 50. 247–277.
Morén, Bruce. 1999. Distinctiveness, coercion and sonority: A unified theory of weight. Ph.D. dissertation, University of Maryland at College Park.
Moulton, William G. 1979. Notker's "Anlautgesetz." In Irmengard Rauch & Gerald F. Carr (eds.) *Linguistic method: Essays in honor of Herbert Penzl*, 241–251. The Hague: Mouton.
Muller, Jennifer. 1999. Geminate markedness: Evidence from Chuukese. Paper presented at the 6th Meeting of the Austronesian Formal Linguistics Association, University of Toronto.
Muller, Jennifer. 2001. The phonology and phonetics of word-initial geminates. Ph.D. dissertation, Ohio State University.
Muller, Jennifer. 2003. The production and perception of word initial geminates in Cypriot Greek. In Solé *et al.* (2003), 1867–1870.
Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.) *The production of speech*, 189–216. New York: Springer.
Ohala, John J., Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.) 1999. *Proceedings of the 14th International Congress of Phonetic Sciences.* Berkeley: University of California.
Payne, Elinor & Eftychia Eftychiou. 2006. Prosodic shaping of consonant gemination in Cypriot Greek. *Phonetica* 63. 175–198.
Ridouane, Rachid. 2007. Gemination in Tashlhiyt Berber: An acoustic and articulatory study. *Journal of the International Phonetic Association* 37. 119–142.
Ringen, Catherine & Robert M. Vago. 2010. Geminates: Heavy or long? In Charles Cairns & Eric Raimy (eds.) *Handbook of the syllable.* Leiden: Brill.
Selkirk, Elisabeth. 1990. A two root theory of length. *University of Massachusetts Occasional Papers* 14. 123–171.
Sievers, Eduard. 1901. *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen.* 5th edn. Leipzig: Breitkopf & Härtel.
Simpson, Adrian P. 1999. Fundamental problems in comparative phonetics and phonology: Does UPSID help to solve them? In Ohala *et al.* (1999), 349–352.
Solé, M. J., D. Recasens & J. Romero (eds.) 2003. *Proceedings of the 15th International Congress of Phonetic Sciences.* Barcelona: Causal Productions.
Stevens, Kenneth N. & Samuel J. Keyser. 1989. Primary features and their enhancement in consonants. *Language* 65. 81–106.
Stevens, Kenneth N., Samuel J. Keyser & Haruko Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In Joseph S. Perkell & Dennis H. Klatt (eds.) *Invariance and variability in speech processes*, 426–449. Hillsdale, NJ: Lawrence Erlbaum.
Swadesh, Morris. 1937. The phonemic interpretation of long consonants. *Language* 13. 1–10.
Thurgood, Graham. 1993. Geminates: A cross-linguistic examination. In Joel Ashmore Nevis, Gerald McMenamin & Graham Thurgood (eds.) *Papers in honor of Frederick H. Brengelman on the occasion of the twenty-fifth anniversary of the Department of Linguistics, CSU Fresno*, 129–139. Fresno: Department of Linguistics, California State University, Fresno.

Topintzi, Nina. 2008. On the existence of moraic onsets. *Natural Language and Linguistic Theory* 26. 147–184.

Topintzi, Nina. 2010. *Onsets: Suprasegmental and prosodic behaviour*. Cambridge: Cambridge University Press.

Tranel, Bernard. 1991. CVC light syllables, geminates and moraic theory. *Phonology* 8. 291–302.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christiane A. M. Baltaxe as *Principles of phonology*. Berkeley & Los Angeles: University of California Press.

Tserdanelis, Georgios & Amalia Arvaniti. 2001. The acoustic characteristics of geminate consonants in Cypriot Greek. *Proceedings of the 4th International Conference on Greek Linguistics*, 29–36. Thessaloniki: University Studio Press.

Yupho, Nawanit. 1989. Consonant clusters and stress rules in Pattani Malay. *Mon-Khmer Studies* 15. 125–137.

48 Stress-timed *vs.* Syllable-timed Languages

Marina Nespor
Mohinish Shukla
Jacques Mehler

1 Introduction

Rhythm characterizes most natural phenomena: heartbeats have a rhythmic organization, and so do the waves of the sea, the alternation of day and night, and bird songs. Language is yet another natural phenomenon that is characterized by rhythm. What is rhythm? Is it possible to give a general enough definition of rhythm to include all the phenomena we just mentioned? The origin of the word *rhythm* is the Greek word ρυθμός, derived from the verb ρεί, which means 'to flow'. We could say that rhythm determines the flow of different phenomena.

Plato (*The Laws*, book II: 93) gave a very general – and in our opinion the most beautiful – definition of rhythm: "rhythm is order in movement." In order to understand how rhythm is instantiated in different natural phenomena, including language, it is necessary to discover the elements responsible for it in each single case. Thus the question we address is: which elements establish order in linguistic rhythm, i.e. in the flow of speech?

2 The rhythmic hierarchy: Rhythm as alternation

Rhythm is hierarchical in nature in language, as it is in music. According to the metrical grid theory, i.e. the representation of linguistic rhythm within Generative Grammar (cf., amongst others, Liberman and Prince 1977; Prince 1983; Nespor and Vogel 1989; chapter 41: the representation of word stress), the element that "establishes order" in the flow of speech is stress: universally, stressed and unstressed positions alternate at different levels of the hierarchy (see chapter 39: stress: phonotactic and phonetic evidence).

Two examples of stress alternation are given in (1) and (2), on the basis of Italian and English, respectively. The first level of the grid assigns a star (*) to each syllable, and is meant to represent an abstract notion of time; on the second, third, and fourth level, a star is assigned to every syllable bearing secondary word stress, primary word stress, and phonological phrase stress, respectively.

(1)

```
                      *                      *                            *                   *
          *           *            *         *                    *       *                   *
          *           *      *     *         *       *            *       *             *     *
     *    *  *    *   * *    *  *  *   *     *   *   *   *    *   *   *   *     *   *   *  *  * *
```

Domani mattina partiremo presto con il barcone nuovo di Federico
'Tomorrow morning we will leave early with the new boat of Federico'

(2)

```
      *                     *            *                          *                              *
      *                     *            *              *           *                              *
      *      *              *            *              *           *            *         *       *
      *  *   *      *   *   *      *     *  *  *    *   *   *       *    *       *    *    *   *   *   *
```

Guinevere will arrive with Oliver tomorrow morning with a transatlantic

Indeed, these examples clearly show that in the two languages there is a similar alternation of stresses ranging from secondary word stress to primary word stress to phonological phrase stress.

The level that is problematic in the metrical grid is the basic level, i.e. the level corresponding to the syllable. This representation does not show any alternation, or any element that establishes order in movement: if we restrict our attention to this level, all syllables are represented with equal prominence. It is clear, however, that grids that are identical at all levels, as in the two following Italian and English sentences, may represent very different rhythms. In particular, the first level – which represents an abstract notion of time for syllables – does not represent important differences between languages, precisely because it is abstract: simple syllables and very complex ones receive identical representations.

(3) a.

```
          *         *                  *
          *         *                  *
          *         *        *         *
    *     *   *     *   *    *     *   *
```

Domani Luca tornerà
'Tomorrow Luca will return'

b.

```
          *              *                       *
          *              *                       *
          *              *            *          *
    *     *    *         *   *        *      *   *
```

Tomorrow Albert will return

There are thus empirical differences in rhythm between languages that are not represented in a metrical grid. Long before the metrical grid theory was proposed, phoneticians (e.g. Pike 1945) had proposed the existence of *rhythmic classes* to account for the rhythmic differences between languages like English or German, on the one hand, and languages like Spanish or Italian, on the other.

3 Linguistic rhythm as isochrony

The idea that languages have different rhythms was first advanced by Lloyd James (1940), who observed that the rhythm of Spanish recalls that of a machine gun languages, adjacent primary word stresses constitute a stress clash and are eliminated in much the same way (Liberman and Prince 1977; Nespor and Vogel 1979, 1989).

That languages vary in their rhythm is a fact. However, from these studies it can be concluded that it is not different rhythms that trigger different phonological phenomena. Rather, different rhythms arise as a consequence of a series of independent phonological properties (cf. also Dasher and Bolinger 1982).

4 Infants' sensitivity to rhythmic classes

Linguists were not alone in investigating rhythmic classes. The discovery in developmental psychology that newborns are capable of discriminating a switch from one language to another (Mehler *et al.* 1987; Mehler *et al.* 1988) triggered further experiments to explore which cues were responsible for this early human ability. In particular, the grouping of languages into different rhythmic classes attracted the attention of cognitive scientists interested in understanding how language develops in the infant's brain. Mehler *et al.* (1996) relied on the classification of languages into syllable-timed, stress-timed, and mora-timed to advance a proposal as to how infants may access the phonological system of the language they are exposed to. In particular, they proposed that the rhythmic class of the language of exposure determines the unit exploited in the segmentation of connected speech: infants exposed to stress-timed languages would use the stress foot (that is, the interstress interval), those exposed to syllable-timed language the syllable and those exposed to a mora-timed language the mora (Cutler *et al.* 1986; Otake *et al.* 1993; Mehler *et al.* 1996).

Most convincing are a number of experiments carried out with French newborns, which show that they are able to discriminate English from Japanese, but not English from Dutch, in low-pass filtered sentences, that is, in sentences whose segmental information is reduced, while prosodic information is largely preserved (Nazzi *et al.* 1998). In order to show that rhythm – rather than any other property of the test languages – is responsible for this discrimination ability, Nazzi *et al.* also tested newborns on a set of randomly intermixed English and Dutch sentences, and showed that they discriminate these from a set of randomly intermixed Spanish and Italian sentences. However, the discrimination ability disappeared when the newborns were tested on a set of English and Spanish sentences *vs.* a set of Italian and Dutch sentences. Thus the intuitions that many phoneticians shared about different rhythms in English and Italian, for example, are confirmed by newborns' sensitivity to this distinction. It is thus clear that some physical property must be present in the signal to account for this difference, but until recently it has not been clear what this property was.

5 Rhythm as alternation at all levels

If isochrony is not responsible for the machine-gun and Morse-code effects, we should ask which characteristics in the signal are responsible for it. That is, what is there in the signal that would account for the clear rhythmic differences of languages belonging to different classes? Or what is the element that establishes order at this level? Ramus *et al.* (1999) answered this question starting from the hypothesis that newborns hear speech as a sequence of vowels interrupted by unanalyzed noise, i.e. consonants; this hypothesis is known as the *Time-Intensity Grid Representation* (TIGRE; Mehler *et al.* 1996). Ramus *et al.* (1999) proposed that, at the basic level, the perception of different rhythms is created by the way in which vowels alternate with consonants. It is thus the regularity with which vowels recur that establishes alternation at this level: vowels alternate with consonants. Starting from the observation that as we go from stress-timed to syllable-timed and then to mora-timed languages the syllabic structure tends to get simpler – and the observation that simple syllables imply the presence of proportionately greater vocalic spaces – vowels would occupy less time in the flow of speech in stress-timed languages than syllable-timed languages. Likewise, in syllable-timed languages vowels would occupy less time than in mora-timed languages, which have the largest amount of time per utterance occupied by vowels. This difference is clear from the rough division into Vs and Cs in the three sentences in (5)–(7). Notice that, in agreement with Ramus *et al.* (1999), glides are treated as Cs if prevocalic and as Vs if post-vocalic.

(5) *English*
The next local elections will take place during the winter
CVCVCCCCVVCVCVCVCCCVCCCVCCVVCCVVCCVCVCCV

(6) *Italian*
Le prossime elezioni locali avranno luogo in inverno
CVCCVCCVCVVCVCCVCVCVCVCVVCCVCCVCCVCVVCVCCVCCV

(7) *Japanese*
Tsugi no chiho senkyo wa haruni okonawareru daro
CVCVCVCVCVCVCCVVCVCVCVCVVCVCVCVCVCVCV

Ramus *et al.* (1999) tested this idea on a corpus of eight languages: English, Polish, and Dutch, representatives of the stress-timed category, French, Italian, Spanish, and Catalan, representatives of the syllable-timed category (Abercrombie 1967), and Japanese, representative of the mora-timed languages (Ladefoged 1975). They observed that languages from the same rhythmic class had similar values for %V – i.e. a similar amount of time occupied by vowels in the speech stream – as compared to languages from different rhythmic classes. The computation of %V was carried out on the basis of a careful segmentation, on the basis of both auditory and visual cues from the spectrogram (cf. Ramus *et al.* 1999). Given the assumption that newborns do not retain the difference between individual Cs and individual Vs, for each sentence only the vocalic and consonantal intervals were measured. Adjacent vowels and adjacent consonants are thus treated as vocalic and consonantal chunks, respectively. A second measure that clusters the languages into three groups is the standard deviation of the duration of consonantal intervals (ΔC), i.e. a broad measure of the regularity with which vowels recur (see Figure 48.1). Both measures are related to syllable structure. A high %V implies that the repertoire of the possible syllable types is restricted, thus also that the consonantal intervals do not vary a great deal, given that there are no languages in which all

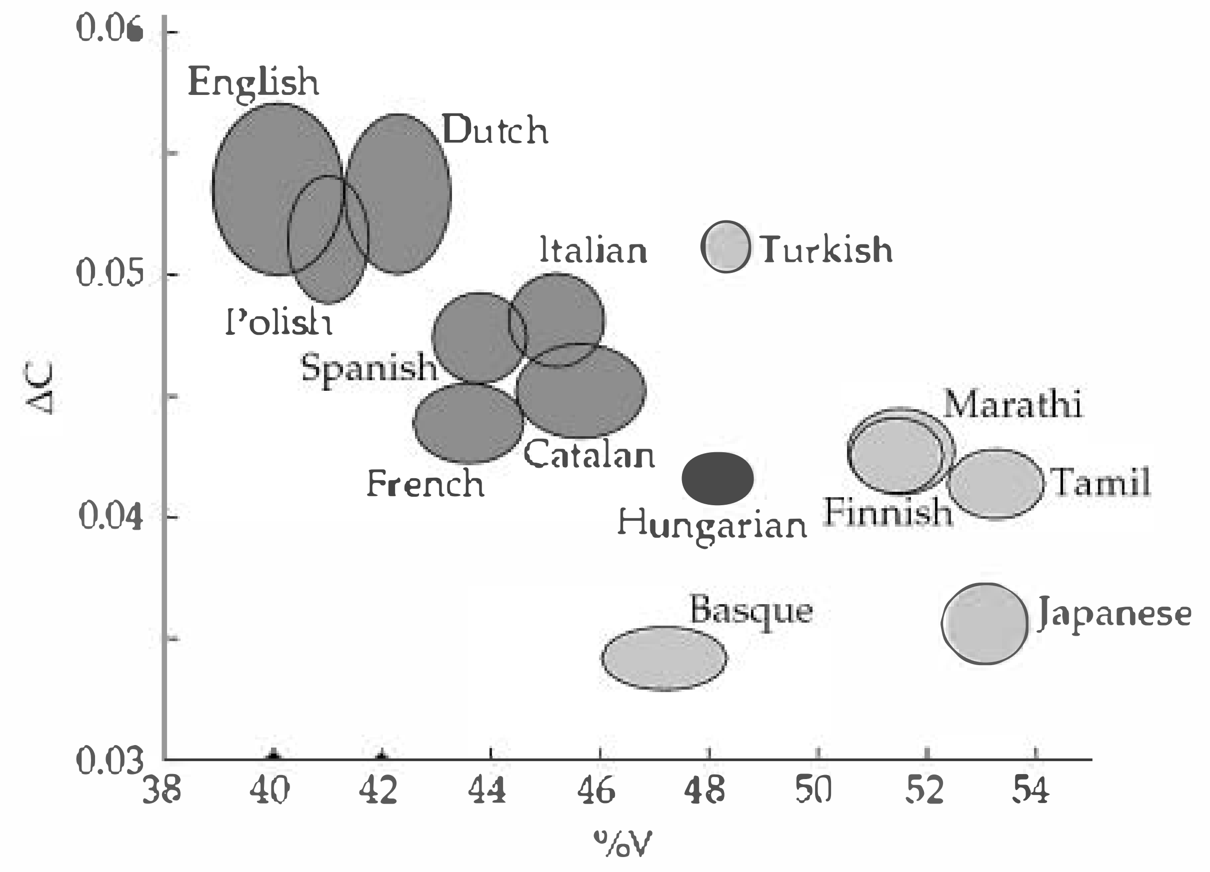


Figure 48.1 ΔC, the standard deviation of the consonantal intervals, *vs.* %V, the amount of time per utterance spent in vowels, for 14 languages. The widths of the ellipses along the two axes represent standard errors of the mean along the axes. Dark ellipses represent head-initial languages, and light ellipses head-final languages. Turkish, Hungarian, Basque, Finnish, Marathi, and Tamil are from unpublished results. Data for the remaining languages are from Ramus *et al.* (1999)

syllables are complex. Rather, even in languages with the greatest variety of syllable types, the basic syllable type – CV – is the most unmarked (Blevins 1995; Rice 2007).

Thus, according to this proposal, rhythm is alternation at all levels: of consonants and vowels at the basic level and of stressed and unstressed syllables, feet, and words at subsequent levels. This order in the flow of speech is always established by the alternation of more and less audible elements.

6 Other proposals

The analysis proposed by Ramus *et al.* (1999) is not the only one to rely on a purely acoustic-phonetic description of the speech stream in trying to understand the basis of linguistic rhythm. In Ramus *et al.* the %V and ΔC variables do not consider the relative ordering of long and short intervals inside an utterance. That is, sequences like CCVː.CCVː.CV.CV and CCVː.CV.CV.CCVː (where Vː is a long vowel, as opposed to VV, which denotes two different adjacent vowels) will yield identical values for their two variables.

Grabe and colleagues therefore chose to examine the *pairwise variability indices* (PVI) in the vocalic and intervocalic intervals in speech (e.g. Low *et al.* 2000; Grabe and Low 2002). This measure is meant to capture a little more of the (local) temporal patterns in speech by considering the variability of all *pairs* of vocalic or intervocalic intervals.

classes differ in their syllabic structure: going from a low %V to a high %V, languages go from having more complex to having simpler syllabic structures. Typologists have in fact observed that various morphosyntactic properties are correlated with the complexity of syllables in a language (Gil 1986; Fenk-Oczlon and Fenk 2004) and, in addition, with its rhythmic patterns (Donegan and Stampe 1983). The computation of %V might therefore offer cues to very different properties of the language of exposure.

Shukla *et al.* (in progress) hypothesize that the correlates of linguistic rhythm, %V and ΔC, have consequences for acquiring correlated morphosyntactic properties like agglutination and word order. These researchers extend the results from Ramus *et al.* to a larger and more varied set of languages. The results indicate that there is a tendency for languages with a low %V to differ from languages with a high %V in head direction, degree of agglutination, richness of the case system, and flexibility of word order (see Figure 48.1). Thus, it is proposed that a simple syllabic structure is correlated with agglutination: if many suffixes can be attached to a word, complex syllabic structure would make these words excessively long and possibly hard to parse.

The question remains why agglutination is found almost exclusively in head-final languages. Two different reasons, both syntactic in nature, have been given. In van Riemsdijk (1998), the explanation for the correlation between head-finality and agglutinative morphology is based on head adjunction, the syntactic device that assembles independent, phonetically realized morphemes in complex words. A principle states that head adjunction can take place only between linearly adjacent heads; since heads are adjacent in head-final languages, while they are separated by intervening specifiers in head-initial languages, head adjunction – and thus agglutination – is expected to take place in OV languages only.

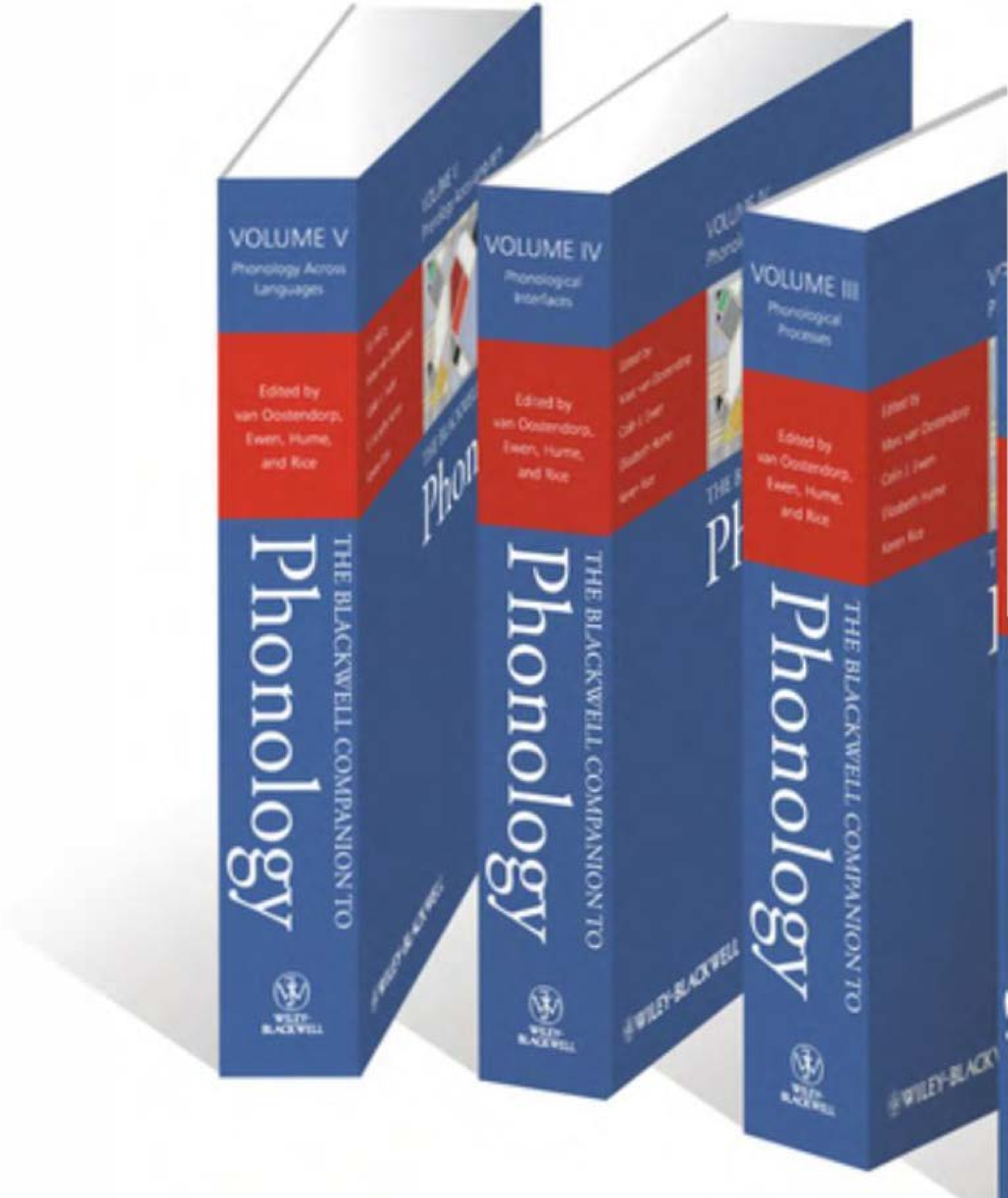
More recently, Cecchetto (forthcoming) assumes that morphological conflation, responsible for fusional morphology, requires that a direct syntactic dependency be established between a selecting head and a selected one. However, in head-final languages this dependency would go backwards, since the selecting head linearly follows the selected one, and backward dependencies are disfavored, for processing reasons (e.g. Fodor 1978). As a consequence, in head-final languages affixes cannot be fused, and result in agglutination instead.

If there is indeed a syntactic explanation for the correlation between head direction and agglutination, the identification of the rhythmic class of the language of exposure would be one of the mechanisms that would assist the infant in the bootstrapping of both the favored morphological operations and word order in the language to which they are exposed.

8 Concluding remarks

In conclusion, linguists' intuitive notion of stress-timed and syllable-timed rhythm is most likely a consequence of the phonological organization of different languages, which can be captured by two relatively simple acoustic-phonetic cues, such as %V and ΔC. Languages appear to be grouped into three rhythmic classes: one corresponding to so-called stress-timed languages, one to so-called syllable-timed languages, and one to so-called mora-timed languages. The rhythmic class to which a language belongs appears also to determine the segmentation unit used by its

O'Connor, J. D. 1965. *The perception of time intervals.* London: University College London.
O'Dell, Michael L. & Tommi Nieminen. 1999. Coupled oscillator model of speech rhythm. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.) *Proceedings of the 14th International Congress of Phonetic Sciences*, 1075–1078. Berkeley: Department of Linguistics, University of California, Berkeley.
Os, Els den. 1988. Rhythm and tempo in Dutch and Italian: A contrastive study. Ph.D. dissertation, University of Utrecht.
Otake, Takashi, Giyoo Hatano, Anne Cutler & Jacques Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32. 258–278.
Pike, Kenneth L. 1945. *The intonation of American English.* Ann Arbor: University of Michigan Press.
Plato. *The Laws*, book II, 93.
Price, P. J., M. Ostendorf, S. Shattuck-Hufnagel & C. Fong. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 90. 2956–2970.
Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14. 19–100.
Ramus, Franck, Marina Nespor & Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73. 265–292.
Rice, Keren. 2007. Markedness in phonology. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*, 79–97. Cambridge: Cambridge University Press.
Riemsdijk, Henk van. 1998. Head movement and adjacency. *Natural Language and Linguistic Theory* 16. 633–678.
Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.
Shen, Yao & Giles G. Peterson. 1962. Isochronism in English. *Occasional Papers, University of Buffalo Studies in Linguistics* 9. 1–36.
Shukla, Mohinish, Marina Nespor & Jacques Mehler. In progress. Grammar on a language map.
Singhvi, Abhinav & David M. Gomez. In progress. Automatic recognition of language rhythm.
Whaley, Lindsay J. 1997. *Introduction to typology: The unity and diversity of language.* Thousand Oaks, CA: Sage.


VOLUME V
Phonology Across Languages
Edited by van Oostendorp, Ewen, Hume, and Rice
THE BLACKWELL COMPANION TO
Phonology
WILEY-BLACKWELL
VOLUME IV
Phonological Interfaces
Edited by van Oostendorp, Ewen, Hume, and Rice
THE BLACKWELL COMPANION TO
Phonology
WILEY-BLACKWELL
VOLUME III
Phonological Processes
Edited by van Oostendorp, Ewen, Hume, and Rice
THE BLACKWELL COMPANION TO
Phonology
WILEY-BLACKWELL